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†

January 15, 2019

Abstract

In 2016 the city of Bogotá doubled police patrols and intensified city services on high-crime streets. They did so based on a policy and criminological consensus that such place-based programs not only decrease crime, but also have positive spillovers to nearby streets. To test this, we worked with Bogotá to experiment on an unprecedented scale. They randomly assigned 1,919 streets to either 8 months of doubled police patrols, greater municipal services, both, or neither. Such scale brings econometric 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. In contrast to the consensus, we find intensifying state presence in Bogotá had modest but imprecise direct effects and that such crime displaced nearby, especially property crimes. Confidence intervals suggest we can rule out total reductions in crime of more than 2–3% from the two policies. More promising, however, is suggestive evidence that more state presence led to an 5% fall in homicides and rape citywide. One interpretation is that state presence may more easily deter crimes of passion than calculation, and place-based interventions could be targeted against these incredibly costly and violent crimes.

JEL codes: K42, O17, E26, J48, C93

Keywords: crime, violence, police, public services, state building, spillovers, field experiment

*Acknowledgments: 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. 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, Zach Tausanovitch 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.

†Blattman (corresponding author): University of Chicago, blattman@uchicago.edu; Green: Columbia University, dpg2110@columbia.edu; Ortega: Development Bank of Latin America (CAF), and IESA dortega@caf.com; Tobón: University of Chicago and IPA, tobon@uchicago.edu.

1 Introduction

Police and city services are the most elementary tools of crime control. When crime rises, cities intensify police patrols, improve lighting, and increase other kinds of state service 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 the people responsible.[1]

The current policy and criminological consensus is that place-based policies reduce total crime. A range of studies have shown that increasing the intensity or quality of policing on high-crime hot spots reduces crime on those corners, streets, or neighborhoods, as does tackling social disorder.[2] Two systematic reviews of the US literature argue that place-based policing not only deter crimes, but that these crimes do not displace nearby. Rather, because there are more instances of positive spillovers to nearby streets than negative ones, the reviews conclude that the benefits of place-based policies diffuse nearby.[3]

Why would benefits diffuse? One hypothesis is that crime is closely coupled to particular places. Another is that many crimes are based on a short-lived motive or circumstances. If so, state presence at the right time and place might deter the crime altogether.

Whether place-based interventions deter or simply displace crime is important for two reasons. First, and most obviously, it is essential to evaluate the policy's overall effectiveness. Second, the answer has implications for our understanding of criminal incentives and 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 should increase criminals' perceived risk of detection and capture in that place, and reduce the likelihood they commit a crime there.[4] If

[1] On policing see Weisburd and Telep (2016); Police Executive Research Forum, (2008). On disorder see Braga et al. (2015). On place-based theory 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).

[4] Police presence disrupts this crime or raises the risk of capture, while city services light dark areas or

that criminal does not commit the crime elsewhere (that is, if place-based policies reduce the total number of 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; that the supply of crime is elastic to the actual or perceived risk of apprehension in a small number of areas; or that most crimes do not have a sustained motive, and are the product of momentary emotions or opportunities.[5]

Based on the policy consensus and these theories, 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.

There are a few reasons for caution, however. One is the question of statistical power to detect spillovers. The median hot spots policing study has fewer than 30 hot spots per treatment arm. Collectively, the existing studies have greater power, and the balance of evidence indeed points to spillovers being beneficial. Even so, the evidence on the direction of spillovers is mixed, with studies pointing both ways.[7] Moreover, some of these studies may not capture the full range of spillovers, as some examine relatively small catchment areas of 1–2 blocks, or assess spillovers after the intervention is completed (rather than during the intensification of state presence). As we discuss in the conclusion, large adverse spillovers are almost certainly within the confidence interval of the collective studies so far. Thus the literature needs more statistical power in more cases.

With this 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

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]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 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.

subtle spillovers over flexible ranges, and to test whether the interventions have different effects on different types of crime.

Scale and spillovers bring econometric challenges to program evaluation. In particular, we show how spillovers in dense networks bias treatment effects and complicate variance estimation through "fuzzy clustering." We show how study design and randomization inference can be used to flexibly estimate spillovers in dense networks, and to 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. Besides the unusual scale, Bogotá has street-level, geo-referenced crime data on all 136,984 streets in the city, plus locations of police patrols every 30 seconds. Hence we can use the universe of city streets to examine spillovers.[8]

We find no evidence that a major reallocation of state presence to higher crime places reduced crime overall. If anything, crime may have increased somewhat, particularly property crime, and our confidence intervals appear to rule out declines of more than 2–3% in all crime. This is striking given the intensity of the intervention — a doubling of normal police time — and the policy consensus that led the Mayor to make these interventions the first plank of his election platform.

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. Direct effects on treated streets tend to be modest and imprecise. This crime, especially property crime, seems to displace nearby. 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 this still seems to be more than outweighed by property crime that spills over nearby.

The remainder of this introduction summarizes the design and results in more detail. 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. This is a large increase in police attention — police are already physically patrolling control segments 6% of each day. Treatment increases this to 12% of the day. The city also targeted 201 segments for municipal services. We randomized assignment to intensive

[8]As we show below, we focus on streets with a positive probability of exposure to spillovers, so ultimately the sample does not include the universe of city streets.

policing, more municipal services, both, or neither. We also monitor spillovers onto the 77,848 non-experimental segments within 250 meters.

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 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.

These two interventions are broadly similar to a set of smaller US policing interventions and evaluations where police time increased roughly 1-3 hours per day (discussed below). These and Bogotá's interventions differ greatly, however, from the interventions in a set of limiting cases that study (at one extreme) 24-hour police presence or (at the other extreme) incremental increases of just 10 minutes (e.g. Di Tella and Schargrodsky, 2004; Draca et al., 2011; Blanes I Vidal and Mastrobuoni, 2017).

The challenge with experimenting at Bogotá's 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 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. We draw similar conclusions if we model spillovers with more structure, such as 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 is not correlated with measurement error, increasing our trust in official crime data.

Intensifying state presence only modestly and imprecisely reduced crime on treated streets. Crime fell roughly 0.13 to 0.15 standard deviations from intensive policing and municipal services. After accounting for spillovers, however, neither decrease is statistically significant.

In aggregate, the direct effects of both treatments are also modest. We can approximate total impacts by aggregating total direct effects and spatial spillovers over the eight months. If we consider crimes reported to the police alone, our main estimate is that over 8 months a total of just 100 crimes were prevented on directly treated street as a consequence of the interventions. This aggregate decrease is not statistically significant.

Meanwhile, we see some 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 800, or 2%. A 90% confidence interval includes zero, and so this should not be taken as strong evidence of adverse effects. Altogether, the results appear to rule out reductions in city-wide crime of more than 2–3%.

If we restrict our attention to the highest-crime streets (true "hot spots"), we can roughly simulate a smaller and more targeted intervention. The same patterns hold. The effect of each intervention is roughly proportional to the levels of crime. At the 90th percentile of baseline crime, for example, the treatment effect of both interventions is a decrease of roughly 25% in reported crimes relative to control experimental streets with similar baseline crime levels (neither effect is statistically significant).

Note that, when a street received both interventions, reported crimes fell by 57%, statistically significant at the 1% level. Hence it is possible that there are increasing returns to both police patrols and municipal services. But this estimate is based on a small number of both-treated streets (just 75) and the estimates are not robust to other estimation strategies, in part because (by chance) streets treated with both interventions had higher levels of initial crime. Hence we take these interaction results with caution.

Importantly, in all of the above 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 990 over the eight months. By the same method, violent crimes fall 177, including a 1% decline in homicides and sexual assaults (60 over eight months), though neither change is statistically significant. The difference between aggregate effects on property and violent crime is statistically significant at conventional levels.

This is an important distinction. 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.[9]

The evidence from US cities has tended not to find many adverse spillovers, at least within a 1–2 block radius. But several large, recent, and non-US studies tell similar stories to Bogotá. 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 hard to deter. An experiment with drunk driving checkpoints in India shows that drunks just take other 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.

[9] 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).

2 Setting and interventions

2.1 Bogotá

Bogotá is a middle income city of roughly 8 million people, where income per capita is about a third of the US. The nature of Bogotá's crime varies, from pickpocketing and cell phone theft in busy commercial areas, to burglary of businesses and homes, to drug sales and any resulting violence. Most violent crimes appear to be crimes of passion, as the Mayor's office estimates that 81% of homicides in 2015 resulted from fights, whereas 12% were contract killings and 5% were violent robberies. Homicide levels are comparable to a large US city, at 15.6 per 100,000 people in 2016. 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 concentrated. From 2012–15 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 widely dispersed. 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. We review crime data in more detail below.

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, compared to 350 in Colombia nationally, 413 in New York, 444 in Chicago, 611 in Washington, or 257 in Los Angeles.[10]

Police patrols are generally well-regarded. We discreetly observed police patrols and qualitatively interviewed residents on 100 of the treated streets, as described below. The force is not without problems, but patrol officers are generally regarded as competent and non-corrupt. If anything, residents complained that officers were not present often enough.

2.2 Interventions

In January 2016 a new mayor came to power. The first item of his election platform was to identify the 750 highest-crime streets in the city and, within the first 100 days, target them with a greater city services, especially police patrols and municipal services. We can view both interventions as an intensification of normal services. Since no new funds or personnel

[10] 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.

were added, this was a randomized reallocation of services.[11] 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

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.

Quadrants have six assigned 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 suspicious people, and will seize illegal weapons (usually knives) and contraband. Patrols tend to focus on young men.

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 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.

Prior to the intervention, from 2012–15, we estimate that 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.

Intensive policing began in February 2016 and ended in October.[12] It generally meant

[11] When it comes to the police, the Mayor's office can influence tactics, force allocations, and equipment, but has little say in total force size. City police forces in Colombia are a branch of the National Police and report up to the Minister of Defense. 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.

[12] 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

almost doubling normal police patrol time on treated streets. As we will see below, during the intervention control streets received roughly 92 minutes of patrol time on average (as expected), with treated streets receiving an additional 77 minutes—an 84% increase.[13]

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 additional times per day for at least 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.[14]

Municipal services

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

the research team enough time to fund and conduct a survey of citizens.

[13] 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 A.1.

[14] 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.

Comparison to other policing studies

The Bogotá intervention is broadly similar in style and intensity to several US hot spots policing interventions. The most comparable interventions 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); and a Sacramento study that added 15-minute police patrols (Telep et al., 2014).[15] This is also broadly comparable to an unpublished Medellín hot spots policing program, where 384 hot spots were treated with 50–70 more minutes patrols over six months in 2015 (Collazos et al., 2018).[16]

The Bogotá intervention is distinct in targeting and duration from several "limiting cases" of extremely high or low changes in police attention. Two natural experiments evaluated round-the-clock police in strategic areas of London and Buenos Aires (Draca et al., 2011; Di Tella and Schargrodsky, 2004). Blanes I Vidal and Mastrobuoni (2017), in another natural experiment, evaluate the effect of an added 10 minutes of patrolling per day in the 200 meter areas around the site of the prior week's burglaries. Otherwise, policing studies tend to examine a change in policing style rather than intensity, and so are not comparable.[17]

3 Experimental sample and design

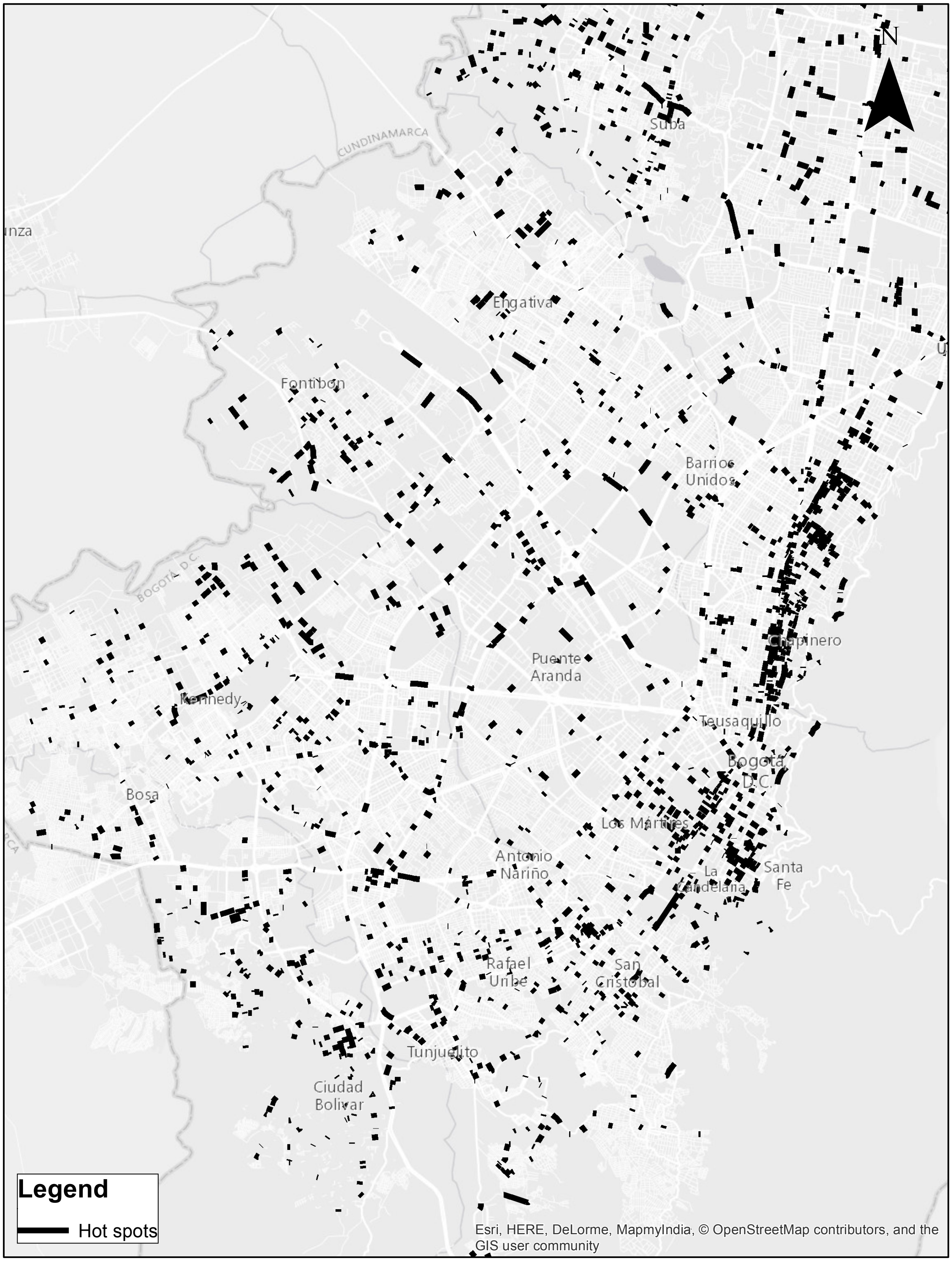
Figure 1 maps Bogotá's 136,984 street segments and indicates the 1,919 segments in our experimental sample. To create this sample, the city started with the 2% highest-crime segments, using an index of reported crimes from geo-coded official statistics, between January

[15] One difference is that our study 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.

[16] 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.

[17] Some hot spots 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; and still others place license plate readers on street corners, or crack down on drug corners and houses. There are also studies of the quality of police response, such as i Vidal and Kirchmaier (2018).

Figure 1: Map of experimental sample



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

2012 and September 2015.[18] The city then asked each station's commanders and staff to use their knowledge (such as crime calls, or observed street disorder) to verify the segments, because (i) most petty crime is unreported, and (ii) crimes could be geo-located to the wrong street. 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.[19] As we discuss below, this led to an experimental sample with varying levels of crime, from modest to acute.

At this scale, we were ex-ante powered to detect direct effects of 0.15 standard deviations. and are powered to detect spillovers as small as 0.02 standard deviations (with 80% power, see Appendix B).

3.1 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.[20]

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 in a single city. The same is true of any intervention in a spatial or social network.

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 of about 1–2 blocks or about 100–150m.[21] We felt this radius could be too narrow, however,

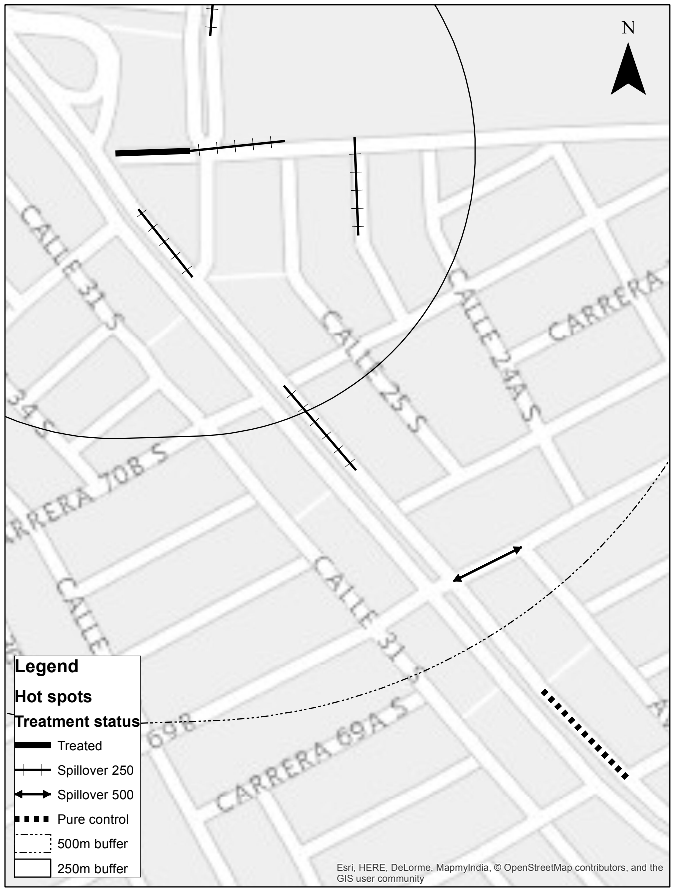
[18] We constructed a geo-fence of 40m around each segment and assigned a reported crime to that segment whenever it fell within its geo-fence. Appendix A.1 reports further details. A calculation error meant that 608 segments outside the top 2% were included in this initial sample. These were generally 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.

[19] 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.

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

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

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



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. Then, within the 500m radius, we need a pre-specified rule for deciding whether to use the 250m or 500m radius for spillovers. We discuss this below.

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 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. This approach ignores the possibility of spillovers beyond 500m, however, as well as non-spatial spillovers. Some crime is undoubtedly displaced in non-Euclidean ways (e.g., to distant segments where crime benefits are high and detection risk is low).[22] We will fail to estimate such spillovers.

3.2 Experimental design and randomization procedures

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.[23]

In March 2016, we selected streets for municipal services. We sent enumerators to take five photographs and rate segments for the presence of disorder.[24] 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.[25]

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.[26]

[22]Ferraz et al. (2016) find evidence of non-spatial spillovers in Rio de Janeiro's favela pacification. Such non-spatial spillovers would lead us to overstate direct effects and understate spillovers.

[23]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.

[24]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.

[25]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.

[26]Technically there are 3×4 "ineligible" conditions, since streets that were diagnosed as having no need

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.

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.

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.[27]

3.3 Estimation strategy

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 municipal services could be <250m, 250–500m, or >500m from either treatment.

[27]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.

for possible confounders, but the estimated treatment coefficients still have the interpretation of mean differences. Following a pre-analysis plan, our regression specification is:

$$Y_{sqp} = \beta_1 P_{sqp} + \beta_2 M_{sqp} + \lambda_1 S^P_{sqp} + \lambda_2 S^M_{sqp} + \gamma_p + \Theta X_{sqp} + \epsilon_{sqp} \tag{1}$$

where Y is the outcome in segment s, quadrant q and police station p; P is an indicator for assignment to intensive policing; M is an indicator for assignment to municipal services; S^P and S^M are indicators for the relevant spillover region (either <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.[28] Weights are the inverse probability weights (IPWs) of assignment to each experimental condition.

Our primary analysis does not estimate the interaction between the two treatments, in part because of the small number of overlapping streets (just 75). However our pre-analysis plan did specify an interest in the interaction, which we estimate by including $\beta_3(P \times M)_{sqp} + \lambda_3(S^P \times S^M)_{sqp}$ in equation 1. We report results with and without the interaction.

Finally, whereas equation 1 estimates spillovers only within the experimental sample of 1,919, we can improve statistical power by taking advantage of the 77,848 nonexperimental segments lie within 250m of one of the 1,919 streets in the experimental sample to estimate spillovers. To estimate spillovers on the full range of streets, we pool the experimental and nonexperimental samples, adding an indicator variable E_{sqp} to equation 1, which takes the value of one for experimental street segments.[29] We can only do this for outcomes using administrative data, of course.

Each of these regressions preserve the comparison of means across treatment conditions. In equation 1, 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

[28] This regression departs slightly from the pre-analysis plan. The plan indicated that we would 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, ignoring municipal services treatment and doing a pairwise comparison of intensive policing treatment and control streets produces biased results, since assignment to municipal services is slightly imbalanced across intensive policing experimental conditions (see Table 2). Hence we 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.

[29] In effect, 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. Also, if 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.

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 (when included) β_3 estimates the marginal effect of receiving both. A negative sign on β_3 implies positive interactions and, when accounting for the interaction term, the effect of both interventions is the sum, $\beta_1 + \beta_2 + \beta_3$.[30]

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.[31] 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.

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:

$$Y_{sqp} = \breve{\beta}_1 P_{sqp} + \breve{\beta}_2 M_{sqp} + \breve{\lambda}_1 \sum_{t \in T_P} f(d_{sqp,t}) + \breve{\lambda}_2 \sum_{t \in T_M} f(d_{sqp,t}) + \breve{\gamma}_p + \breve{\Theta} X_{sqp} + \epsilon_{sqp} \tag{2}$$

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 indicate segments treated with intensive policing and municipal services. 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

[30] 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.

[31] 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 A.2 describes and maps IPWs in our sample.

effects. We consider an exponential decay function, $f(d_{sqp,t}) = 1/(e^{d_{sqp,t}})$, as well as an inverse linear decay. We calculate statistical significance using randomization inference.[32]

3.4 Summary statistics and randomization balance

Table 2 reports summary statistics and balance tests for the experimental sample. Our primary focus is crime, and we have several measures. 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.[33] Experimental segments had between up to 82 crimes reported in the previous four years using the original data, and up to 461 with the updated data.In both cases, however, the average is much more modest: about 5 reported crimes per segment, or just one per year.[34] Thus our sample has a variety of moderate to high crime streets. We discuss this variation in more detail in section 4.2 below.

Random assignment produced the expected degree of balance along covariates in our main and primary test, without the treatment interaction between policing and municipal services. Perhaps because of the small number of streets receiving both treatments, however, we see an imbalance in assignment to the "both" category that is due to chance. As we will see, our results are not that sensitive to this baseline imbalance, but nonetheless it is a reason to take any results with treatment interactions with caution (beyond the small sample size of both-treated segments).

Table 2 reports the weighted means for a selection of baseline covariates, by experimental assignment, for experimental and non-experimental segments. Recall that we randomized the interventions sequentially. For the most part, background attributes appear balanced across experimental conditions when we ignore treatment interactions in columns 4–7. There are some minor differences between treatment and control segments (for instance, treated segments are slightly less likely to be in industrial zones or middle income status, and treated quadrants have slightly fewer experimental segments), but overall the imbalance is consistent with chance and is robust to alternative balance tests.[35] Columns 8–13 report balance

[32] We cannot use 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. 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.

[33] 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.

[34] 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.

[35] 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 experi-

Table 2: Descriptive statistics for the experimental sample and tests of balance (N=1,919)

	Summary statistics			Without interaction (primary)				With interaction					
				Intensive policing		Municipal services		Intensive policing		Municipal services		Both	
	Mean	Std. Dev.	Max.	Coeff	p-val	Coeff	p-val	Coeff.	p-val	Coeff.	p-val	Coeff.	p-val
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
# of reported crimes on street, 2012-15 (original)	4.53	5.72	82	-0.18	0.68	-0.14	0.89	-0.45	0.28	-0.46	0.38	0.96	0.39
# of violent crimes	1.88	2.94	56	-0.20	0.37	-0.06	0.89	-0.44	**0.05**	-0.34	0.20	0.70	0.13
# of property crimes	2.66	3.97	50	0.03	0.96	-0.08	0.91	0.00	0.95	-0.11	0.82	0.27	0.89
# of reported crimes on street, 2012-15 (updated)	5.18	18.24	461	-0.35	0.86	-0.18	0.91	-2.86	**0.07**	-3.11	**0.05**	6.95	**0.01**
# of violent crimes	1.40	5.38	78	0.33	0.28	0.33	0.23	-0.55	0.14	-0.71	**0.08**	2.35	**0.00**
# of property crimes	3.78	14.09	407	-0.69	0.61	-0.50	0.81	-2.31	**0.07**	-2.40	**0.07**	4.59	**0.03**
Average # of reported crimes per segment in quadrant, 2012-15	3.56	5.13	61	-0.36	0.30	0.14	0.31	-0.03	0.88	0.54	**0.02**	-0.65	**0.10**
Daily mean patrol time (11/2015-01/2016), mins	38.03	70.27	1029	-2.61	0.69	3.02	0.56	-1.45	0.89	4.38	0.38	-2.77	0.73
Rating of baseline disorder (0–5)	1.18	0.74	5	-0.02	0.48	0.08	0.11	0.02	0.69	0.13	**0.01**	-0.11	0.14
Meters from police station or CAI	551.37	351.46	2805	-25.70	0.24	-1.80	0.79	-22.42	0.30	2.04	0.68	-10.28	0.84
Zoned for industry/commerce	0.38	0.49	1	-0.10	**0.01**	0.04	0.23	-0.05	0.20	0.09	**0.01**	-0.11	**0.06**
Zoned for service sector	0.13	0.34	1	0.02	0.31	0.03	0.24	0.02	0.44	0.02	0.36	0.01	0.72
High income street segment	0.07	0.25	1	0.00	0.91	0.01	0.49	0.00	0.91	0.01	0.49	0.00	1.00
Medium income street segment	0.55	0.50	1	-0.05	0.11	0.01	0.83	-0.02	0.46	0.05	0.13	-0.07	0.11
# of segments in quadrant	127.21	86.99	672	2.59	0.72	-3.95	0.47	-0.05	0.92	-7.04	0.18	6.64	0.47
# of experimental segments in quadrant	3.67	2.68	14	-0.05	0.39	-0.24	0.24	0.22	0.95	0.08	0.46	-0.25	0.17
# segments treated with policing in quadrant	1.15	0.95	3			-0.09	0.16			-0.69	**0.00**		
# segments treated with services in quadrant	0.66	0.69	3	-0.08	**0.04**			-0.35	**0.00**				

Notes: Columns 1–3 display summary statistics for our experimental sample of 1,919 segments, weighted by the probability of being in the observed experimental condition. Columns 4–13, report balance tests for treated vs all control units using weighted least squares, with and without the interaction between the two treatments.

tests with the interaction term. We can see that higher-crime streets are more likely to have received both treatments. In principle, this should bias our treatment effects to zero. When analyzing program results, we will tend to focus on the non-interacted regressions and interpret the interaction coefficient with caution, but also we control for these baseline conditions to alleviate any potential imbalance, and test whether results are sensitive to their inclusion as robustness tests. In brief the imbalance does not affect our conclusions.

3.5 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.

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.

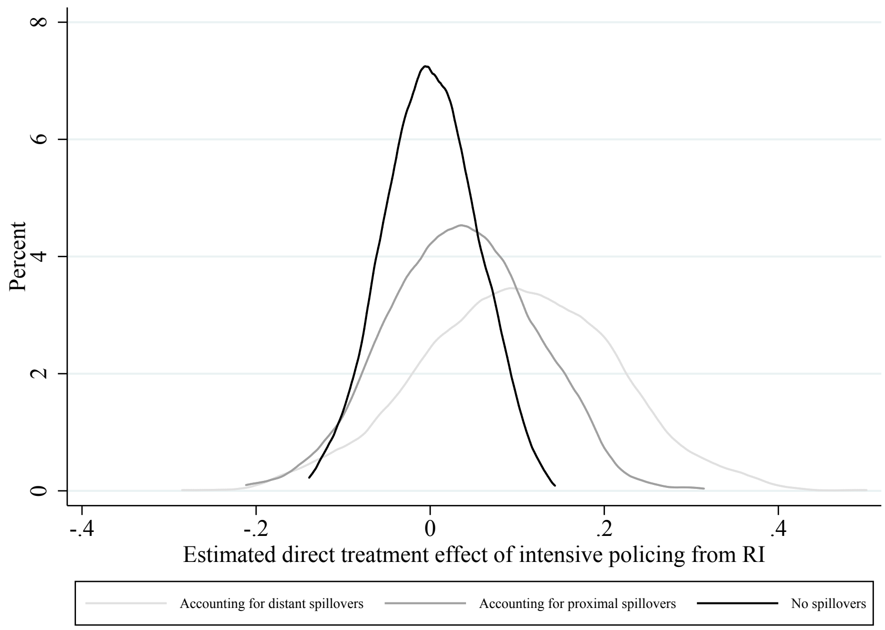
Most importantly, the distribution widens when accounting for spillovers. The no-spillovers case has the narrowest distribution. The distribution widens as we account for wider spillover regions. That is, we are more likely to get large treatment effects by chance.

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 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 probabil-

mental 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.

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



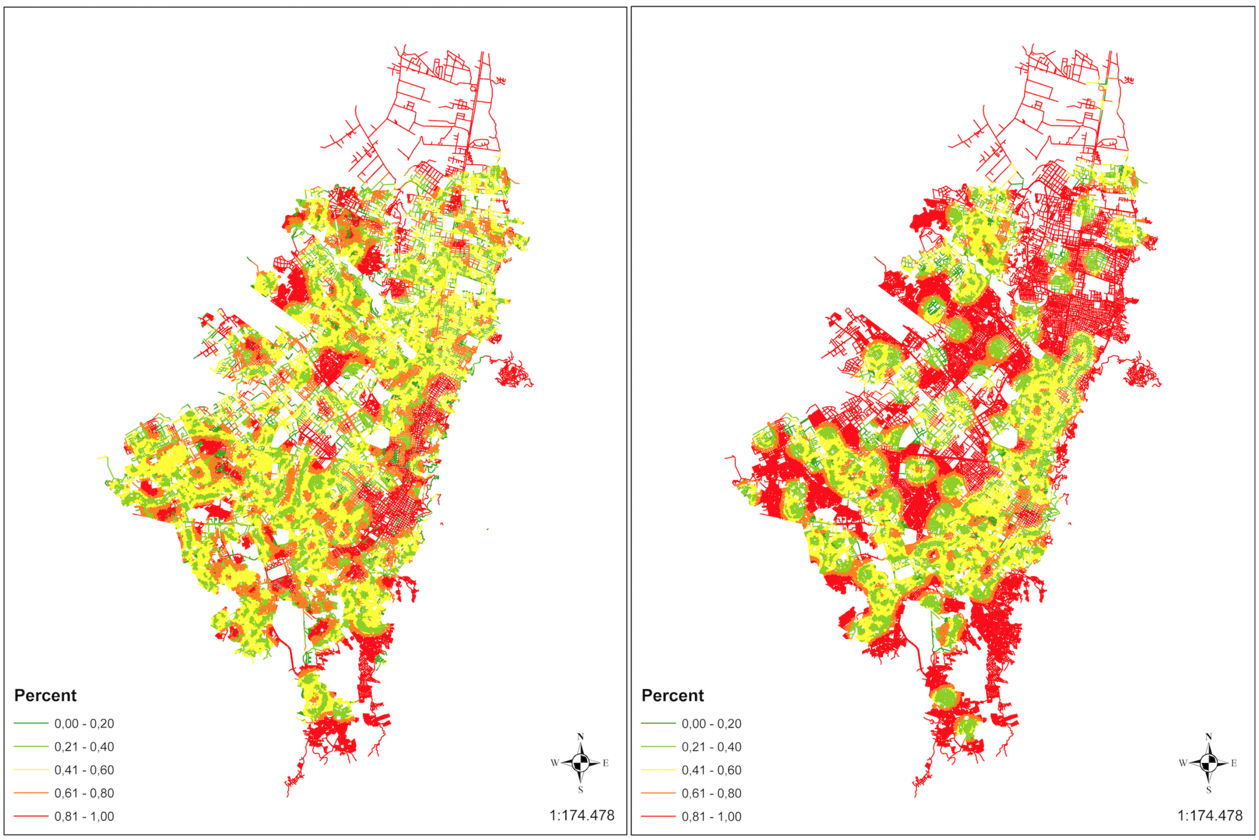
Notes: The figure displays the empirical distribution of treatment effects on the insecurity index for intensive policing. We simulate the randomization procedure 1,000 times and estimate treatment effects for each randomization using post-treatment data under the sharp null of no treatment effect for 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.

ity 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 our boundaries. The figures imply that, instead of having thousands of independent segments, we have dozens of clusters with no fixed area.

Finally, the simulations in Figure 3 show that the distributions of simulated treatment effects with spillovers are not centered at zero. Equation (1) 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 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

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).

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 (reported in Appendix A.2). All of our tables report bias-corrected treatment effects.

4 Data

We draw on five main sources of data.

1. *Administrative data on police, municipal services, and streets.* The police shared GPS patrol locations for all 136,984 streets, 2015–17.[36] For streets assigned to municipal services, agencies shared their diagnosis and compliance. The city also shared administrative data on the baseline variables reported in Table 2.

2. *Officially reported crimes and calls for service.* Police shared data on reported crimes and operations 2012–17, geolocated to 136,984 streets.[37]

3. *Crime survey of 24,000 residents.* We conducted a survey for three reasons. First, as we will see, a majority of crime and nuisances go unreported. Second, we wanted to test whether treatment increased crime reports, inflating our treatment effects. Third, we wanted to measure secondary outcomes such as citizen trust in police. In October 2016 we surveyed a convenience sample of 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. The survey collected: 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,

[36]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.

[37]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).

garbage, and boarded-up buildings on a 0–5 scale.[38]

5. *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.

4.1 Outcomes

Our primary outcome is criminal insecurity, as specified in our pre-analysis plan. When we assess aggregate and spillover effects on all streets, our sole measure of insecurity is officially-reported crimes on the segment. However, to measure direct effects alone, we prespecified two insecurity indexes (one of which incorporates officially reported crimes):

1. *Perceived risk of crime and violence on the segment.* Our survey asked respondents to rate perceived risk on that segment on a 4-point scale from "very unsafe" to "very safe" in 5 situations: general risk during the day; a young woman walking alone after dark; a young man walking alone after dark; talking on a smartphone. We construct an index of perceived risk that takes the average across all respondents in the segment.

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;[39] 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.

[38] We visited 1,534 of a total of 1,919 scheduled streets in March (three months before the municipal services intervention began) in order to narrow down the number of eligible 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 3.2, 1,459 were eligible for the municipal services interventions and 414 of them were assigned to treatment. Those streets were split in two batches of 201 and 213 streets respectively in order to randomize timing, but only the first batch was effectively treated. Then, in order to assess the levels of compliance, we sent enumerators to the 414 streets in the first and second batches in June (one to two weeks after municipal services started to be delivered) and December (two months after the end of the intervention). Again, because of security concerns of the enumerators, we visited 409 in June and 410 in December.

[39] 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.

We use standardized summary indices in order to reduce the number of hypotheses tested and avoid the need for multiple comparison adjustments (following Kling et al. 2007). When calculating aggregate and spillover effects we consider officially reported crimes only, and for direct effects we focus on an index of perceived risk and crime incidence.[40]

4.2 Sample variation in crime

The scale of the experiment and our street selection mechanism means that the experimental sample includes a range of segments, from moderate to very high levels of crime. There is also considerable variation in the non-experimental sample of potential spillover streets. To see this, Appendix A.3 graphs and compares the samples. In brief, we draw four conclusions from this descriptive analysis.

First, only some of the experimental streets correspond to what the U.S. literature calls a "hot spot", in the sense of extreme levels of crime. As a result, we should compare effectiveness to other hot spot interventions with some caveats. An advantage of this 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. Below we will "simulate" a hot spots intervention by looking at impacts on the subsample of highest crime streets.

Second, "hot spot" or not, experimental segments are relatively high-crime. On average, segments in the top 2% have about 5 times as many reported crimes as those in the non-experimental sample.[41]

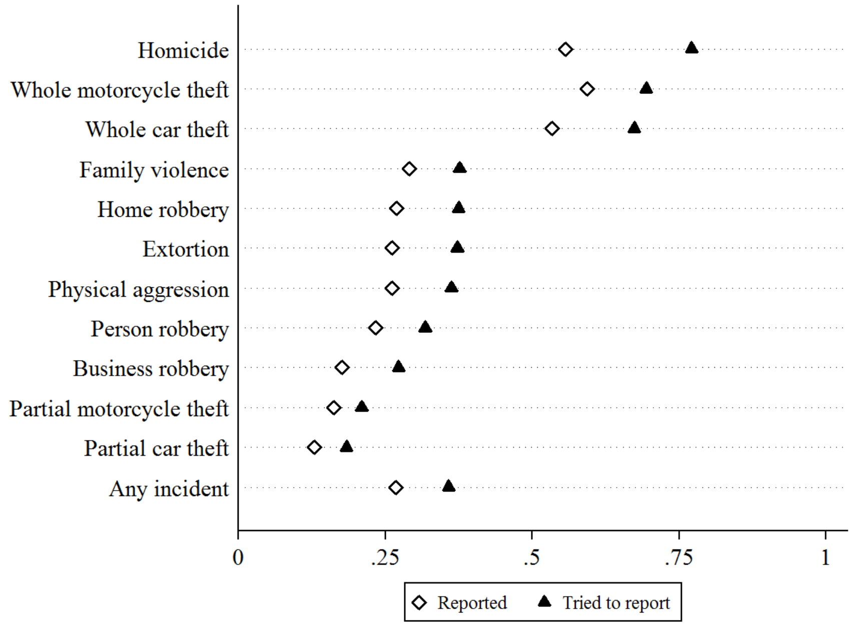
Third, involving the police in street selection meant that our experimental sample includes some segments with low levels of reported crime, but moderate to high unreported crime and nuisances. According to the survey responses, 3 in 10 of the people stopped on each of the experimental segments reported 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.[42] Local police

[40] We discuss secondary outcomes, particularly the perceived legitimacy of the police and local government, in Appendix C.1.

[41] This is in spite of the fact that our procedure ensures there are also some high-crime non-experimental segments. These are simple to explain: the police limited the number of treated segments to two per quadrant. In high-crime quadrants, this means many relatively high-crime segments are in the non-experimental sample.

[42] Some top 2% streets also have a small number of officially-reported crimes in the pre-intervention period. Also, a small number of the non-2% streets have a sizable number of crimes. Why is this so? To understand the lower-crime "top 2%" streets, recall 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. 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. Furthermore, none of the experimental streets should be considered "low-crime", simply lower reported crime.

Figure 5: 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.

reviewed every segment. Those with low levels of reported crime nonetheless have high levels of unreported crime and nuisances.

Fourth, a majority of unreported crimes are pettier crimes. Figure 5 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 most crimes are never reported.

4.3 Is treatment correlated with measurement error in crime?

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.

5 Results

We estimate equation (1) above, including the sample of non-experimental streets to estimate spillovers when analyzing administrative data. Unless otherwise noted, our spillover condition is limited to streets within 250m of treated segments only. This follows from our pre-specified rule, whereby 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.

5.1 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 3 reports the effects of assignment to each program on various first-stage outcomes.

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.[43]

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 geolocators 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.[44]

[43]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.

[44]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 3: "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 4: 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.

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 4 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 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 3). 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 3. 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.

5.2 Direct program effects on insecurity

Doubling police presence or improving municipal service delivery has a modest but imprecise effect on perceived and actual security on directly treated streets. Table 5 reports the direct

Table 5: Program impacts on security in the experimental sample (N=1,916), accounting for spillovers within 250m, with p-values from randomization inference, with interaction between treatments

Dependent variable	Control mean	Effects without interaction (primary): Any intensive policing	Effects without interaction (primary): Any municipal services	Effects with interaction: Any intensive policing	Effects with interaction: Any municipal services	Effects with interaction: Both interventions	Effects with interaction: Sum of (2), (3), and (4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Insecurity index, z-score (+ more insecure)	-0.003	-0.129	-0.153	-0.098	-0.105	-0.125	-0.329
		0.274	*0.172*	*0.386*	*0.399*	*0.445*	***0.057***
Perceived risk index, z-score (+ riskier)	0.049	-0.139	-0.127	-0.116	-0.093	-0.090	-0.299
		0.168	*0.184*	*0.244*	*0.381*	*0.625*	***0.060***
Crime index, z-score (+ more crime)	-0.054	-0.076	-0.128	-0.048	-0.083	-0.119	-0.249
		0.520	*0.274*	*0.714*	*0.591*	*0.422*	*0.163*
Perceived & actual incidence of crime, z-score (survey)	0.059	-0.056	-0.132	-0.072	-0.157	0.067	-0.163
		0.750	*0.204*	*0.519*	*0.121*	*0.409*	*0.504*
# crimes reported to police on surveyed segment (admin)	0.743	-0.094	-0.088	0.011	0.080	-0.442	-0.351
		0.488	*0.687*	*0.816*	*0.406*	***0.038***	***0.085***

effects of the intervention in the 1,916 experimental streets only. (We assess and report spillovers in the following section, on the full sample of city streets.) In addition to our two main insecurity measures, we report 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. Columns 2–3 report our primary specification without the interaction, and columns 4–7 report results with the interaction (including the sum of the three coefficients).

Ignoring the treatment interaction, each intervention is associated with a 0.13 to 0.15 standard deviation security improvement on directly treated streets, not statistically significant. The coefficients on the component indexes — perceived risk and actual crime — are similar.[45] Thus the program impacts are in the expected direction but imprecise.

We only see large and statistically significant impacts of state presence in the 75 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 7). Baseline crime levels in these both-treated segments were somewhat higher, and the regression controls for that imbalance. Omitting baseline covariates dramatically reduces the precision of the estimate of both treatments (Appendix C.3). Hence we regard it with caution. Nonetheless, it is suggestive evidence 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.

5.3 Direct, spillover, and aggregate program effects on officially-reported crime

The previous section looked at reported crimes in the directly-treated streets only. To see spillover and aggregate effects as well, Table 6 reports estimates of direct and spillover effects on the full sample of segments, experimental and non-experimental.[46] Again, we report impacts with and without the the interaction term between intensive policing and municipal services.

[45] After completion of the experiment, we also received calls-for-service data from police. We did not pre-specify 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).

[46] 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. Appendix C.7 estimates the "unpooled" results on the experimental and non-experimental samples separately.

These administrative data are useful for estimating spillover and aggregate impacts because we have data on 100-fold more streets. We do not have survey data on all streets, but these officially reported crimes are probably a good proxy for two reasons. One is that, in the previous section, direct treatment effects for the administrative and survey data were very similar. Second, as noted in Section 4.3, we see no evidence of measurement correlated with treatment in these administrative data.

Direct treatment effects As before, the interventions appear to have had a relatively modest and imprecise direct effect on officially-reported crime. This is true on a per segment basis, but more importantly the total amount of crime directly deterred seems to be very small.

We see only weak evidence that increases in police patrols and municipal services reduce crime. Without 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 -0.098, a 13% improvement.[47] 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 again see a large and statistically significant impacts in the segments that were assigned to both interventions. The sum of the three coefficients is -0.423 with a p-value of 0.008 (not reported in the table), and corresponds to a 57% decrease in reported crimes on the 75 streets that received both interventions. The coefficients on policing and municipal services alone actually switch signs to point to a slight increase in crime, although both are imprecise.[48] Again, however, we need to regard this estimate with caution because it was not a primary specification, there was baseline imbalance in crime levels, and because when we omit baseline covariates the reduction in crime on these 75 segments is no longer statistically significant.

Importantly, 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.

[47]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 patrols is a decrease of about 0.1 crimes. This is similar to the average effect of -0.098, as the average treatment street received 76 minutes of extra patrolling time. Appendix C.4 has these IV estimates.

[48]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 (primary)				Interaction between treatments			
				Estimated total impact =				Estimated total impact =
	Coeff.	RI p-value	# segments	$(1) \times (3)$	Coeff.	RI p-value	# segments	$(5) \times (7)$
Impacts of treatment	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Direct treatment effect								
Intensive policing	-0.098	*0.386*	756	-74.431	0.035	*0.665*	756	26.796
Municipal services	-0.133	*0.185*	201	-26.825	0.073	*0.598*	201	14.587
Both					-0.530	***0.010***	75	-39.776
Subtotal				-101.256				41.383
B. Spillover effect								
Intensive policing	0.017	*0.112*	52095	871.760	0.019	*0.108*	52095	1005.599
Municipal services	0.002	*0.645*	21286	42.365	0.006	*0.967*	21286	119.259
Both					-0.007	*0.556*	15772	-111.442
Subtotal				914.124				1124.858
Net increase (decrease) in crime				812.868				1015.023
			95% CI	(-914, 1926)			*95% CI*	(-1065 ,2252)
			90% CI	(-583 , 1720)			*90% CI*	(-824 , 1958)

Notes: 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.

They did not deter any crime when we account for the interaction term.

Spillover effects Panel B of Table 6 estimates spillover effects on the experimental and non-experimental spillover streets, pooled. The evidence suggests that any crime deterred may have been displaced to nearby segments. For intensive policing, all spillover coefficients 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—914 when we do not allow for the interaction and 1,125 when we do.

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.

This back-of-the-envelope calculation helps us rule out large decreases in crime from the reallocation of police. Note that this is not a true general equilibrium estimate. Rather it is simply a means place some policy-relevant magnitudes on total crime impacts.

These crude aggregate estimates suggest the treatments increased crimes by about 813 city-wide, or 2% 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.

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 2–3%. 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 (table not shown). 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.[49] 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 (table not shown).

[49] 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: 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	(-914 , 1926)	(-583 ,1720)	1015.0	(-1065 , 2252)	(-824 ,1958)
Property crime	989.7	(-316 ,1941)	(-167 , 1769)	1384.0	(-394 , 2413)	(-136 , 2193)
Violent crime	-176.8	(-894 , 347)	(-786 , 250)	-369.0	(-1126 , 234)	(-1012 , 115)
Homicides and sexual assaults	-59.6	(-176 , 56)	(-159 ,43)	-86.0	(-229 , 43)	(-200 , 24)
Difference between property and violent crime	1166.5			1752.9		
p-value	0.071			0.017		

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.

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.5. 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).

5.4 Heterogeneity by type of crime

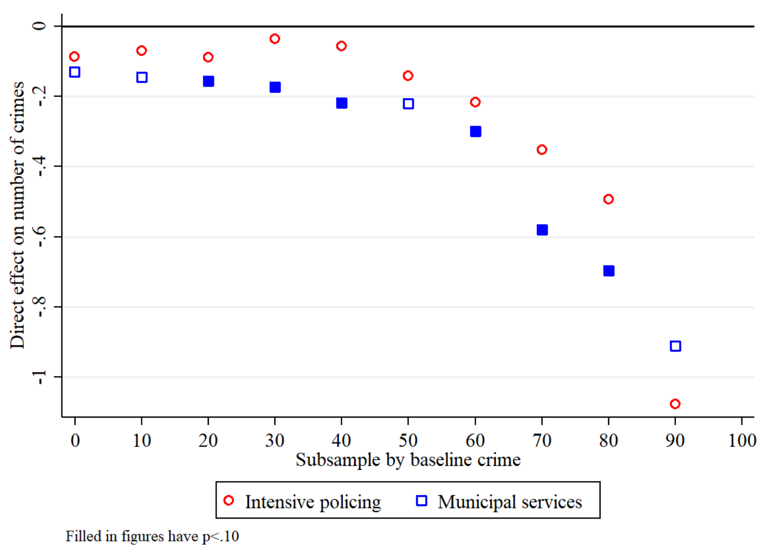
Police tend to prioritize violent crimes such as assault, rape and murder over property crimes such as burglary or theft. Table 7 takes the aggregate impacts on officially-reported crime from Table 6 and disaggregates these total effects into violent and property crimes. (Appendix C.6 reports the full tables).

The interventions have opposing effects on property and violent crime. Our best estimate is that aggregate violent crimes fell by 177 crimes in total (1% relative to the total number of violent crimes) when we do not allow for the interaction between treatments. The two most socially costly crimes, homicides and sexual assaults, fall by 60. This represents a large proportion of very serious crimes—5% relative to the total number of homicides and sexual assaults reported 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 990 in aggregate (4% relative to the total number of property crimes). This increase is not statistically significant. Importantly, the difference between aggregate effects in property and violent crimes is statistically significant at the 10% level when we do not allow for the interaction, and at the <1% level with the interaction between treatments.

5.5 Heterogeneity by level of initial crime

Figure 6: Program impacts in nth percentile highest-crime street segments



Notes: We estimate equation 1 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 right side of the figure represents the highest crime "hot spots").

We pre-specified one major form of heterogeneity analysis, by baseline crime. This helps us compare our results to the U.S. hot spot policing literature. Figure 6 reports estimated direct treatment effects on the insecurity index for the n% highest-crime hot spots. Specifically, we estimate equation 1 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. Broadly speaking, the results are consistent with direct effects being roughly proportional to the levels of crime. For instance, at the 90th percentile of baseline crime, there are just over 4 crimes reported per segment. The treatment effect of roughly -1 crime is a decline equal to 25%% of the the average crime totals at control experimental streets with similar baseline crime levels.

5.6 Robustness

Robustness to estimation strategy Panels A and B of Table 8 compare estimates with and without baseline covariate adjustment. Panel A replicates our main results from above, for comparison. The estimates in Panel B omit all baseline covariates. The estimates are statistically indistinguishable. Appendix C.3 reports estimates without covariates and an interaction between treatments. As discussed above, the decline in crime on both-treated segments is sensitive to the inclusion of baseline covariates.

To see the effect of other design and estimation choices, in Appendix C.8 we estimate "naïve" treatment effects ignoring IPWs and randomization inference, but including covariates. Direct treatment effects are slightly smaller than in Table 5, but the patterns remain similar. The estimated spillover effects in this "naïve" case, however, are much larger and highly statistically significant compared to Table 5. Hence failing to account for interference between units and clustering of treatment conditions would have led us to severely exaggerate the degree to which the interventions push crime elsewhere.

Robustness to alternative spillover functions Table C.9.1 reports the results of alternative methods of spillover estimation. All adjust for baseline covariates. In Panel C, 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 D, we estimate equation 2 using an exponential rate of decay rather than our fixed radii. Finally Panel E 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, for direct treatment effects we draw similar conclusions regardless of method (see Table 6 above): both intensive policing and municipal services have a negative but not statistically significant effect on crime. The decline in violent crimes resulting from intensive policing is not robust to the change in spillover adjustment, however, as panels C through E show non significant results.

Table 8: Estimated direct and spillover effects using alternative covariate adjustments and methods of spillover estimation, with RI p-values

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)	(5)	(6)
A. Main specification (covariate adjustment, spillover indicator)					
# of total crimes	0.743	-0.098	-0.133	0.017	0.002
		0.386	*0.182*	*0.112*	*0.642*
# of violent crimes	0.250	-0.066	-0.050	0.001	-0.009
		0.049	*0.224*	*0.606*	***0.074***
# of property crimes	0.494	-0.033	-0.084	0.015	0.011
		0.813	*0.319*	***0.085***	*0.871*
B. Main specification, without covariate adj.					
# of total crimes	0.743	-0.112	-0.117	0.015	0.021
		0.337	*0.326*	*0.200*	*0.905*
# of violent crimes	0.250	-0.064	-0.057	0.004	-0.008
		0.065	*0.170*	*0.311*	*0.173*
# of property crimes	0.494	-0.049	-0.060	0.011	0.029
		0.608	*0.546*	*0.235*	*0.712*
C. Spillover count measure, with covariate adj.					
# of total crimes	0.743	-0.019	-0.063	-0.014	-0.004
		0.532	*0.505*	*0.962*	*0.957*
# of violent crimes	0.250	-0.026	-0.034	-0.003	0.000
		0.289	*0.343*	*0.974*	*0.990*
# of property crimes	0.494	0.007	-0.029	-0.010	-0.004
		0.748	*0.704*	*0.954*	*0.959*
D. Exponential decay spillover function, with covariate adj.					
# of total crimes	0.743	-0.027	-0.094	0.055	-0.048
		0.482	*0.300*	*0.246*	*0.107*
# of violent crimes	0.250	-0.028	-0.038	0.006	0.001
		0.236	*0.253*	*0.733*	*0.958*
# of property crimes	0.494	0.001	-0.056	0.049	-0.049
		0.695	*0.474*	*0.257*	*0.056*
E. Linear decay spillover function, with covariate adj.					
# of total crimes	0.743	-0.026	-0.086	0.017	-0.021
		0.498	*0.343*	*0.467*	***0.096***
# of violent crimes	0.250	-0.027	-0.038	0.003	0.000
		0.250	*0.262*	*0.696*	*0.932*
# of property crimes	0.494	0.001	-0.048	0.014	-0.022
		0.722	*0.531*	*0.502*	***0.047***

Notes: Randomization inference p-values are in italics. This table estimates the coefficients on spillovers, $\check{\lambda}$, using equation 1 for panels A and B, and equation 2 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.

Turning to spillovers, we generally observe adverse spillovers resulting from intensive policing. The only exception is Panel C, where we estimate the effect of an additional treated street in the surroundings. In this case, the margin of uncertainty does not allow to conclude on the final direction of the spillovers (all p-values are larger than 0.95). On the other hand, we do not observe a clear pattern on the spillover effects resulting from municipal services. Panels A and B suggest there are adverse spillovers on total and property crime, and beneficial spillovers on violent crime. But panels C through E suggest the exact opposite. This may suggest the functional form of the spillover effects resulting from municipal services are complex and generally unknown to us.

6 Discussion

Two recent meta-analyses of place-based policing demonstrate that, on average, instances of positive spillovers outweigh negative ones (Braga et al., 2012; Weisburd and Telep, 2016). This has contributed to a policy and criminological consensus that place-based policies not only stop crime on targeted streets, but the benefits also diffuse to nearby streets (Abt and Winship, 2016). The consensus has shaped crime policy worldwide. If true, it also challenges some common notions in the economic analysis of crime. Positive spillovers run against an economic intuition that criminals with sustained motives simply offend nearby.

Our view is that, even before considering the Bogotá results, policymakers should view the direction of spillovers as fundamentally uncertain. Rather than question the average effect across studies, consider the confidence interval. To use one meta-analysis as an example, of the 13 individual spillover estimates documented in Braga et al. (2014, Figure 2), eight estimates have a p-value smaller than 0.001. Yet these studies are fairly small, with a median number of treated units below 30. Even in a study the size of Bogotá's, levels of precision were never as high as $p<0.001$. The true confidence interval probably includes very large positive and negative spillovers.

The challenge facing any meta-analysis is that the individual papers seldom report sufficient or comparable information. And only recently has it become common to move beyond simple t-tests of means. Few studies account for small sample distributions or adjust standard errors for clustering of treatment assignment. Thus any meta-analysis involves a degree of guesswork and needing to take improbably precise estimates at face value.

Bogotá offers a new data point, one with the benefits of scale. Based on the consensus, the city dramatically reorganized personnel, doubling police time and intensifying normal city service on hundreds of high-crime streets. We find that crime was only moderately responsive on the directly targeted streets, and that on balance there is little evidence of positive or adverse spillovers within 250m (although there are some indications of adverse spillovers for property crime). Indeed our confidence intervals on total effects of the reorganization of police and city services seems to rule out more than a 1–2% decrease in city-wide crimes. When we confine our analysis to the higher-crime hot spots, we draw similar conclusions.

Together with our commentary of the meta-analyses, and recent studies in Medellin and Essex, our Bogotá results suggest a revision of the current policy consensus, to regard the direction of spillovers as uncertain, and perhaps adverse on average and across all types of crime. Future meta-analysis will need to look more closely at the hyper-precise estimates from small samples more critically.

Yet the Bogotá evidence also holds out some hope for place-based interventions. It 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). These patterns are consistent with theories of crime deterrence that emphasize the importance of a sustained criminal motive.

Policy conclusions and cost-effectiveness

Cost-effectiveness in this case is in the eye of the beholder. Bogotá's government viewed 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 4% is worth what could 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 personnel had 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. Intensive policing was a major reform, and like any bureaucracy, the police can only undertake so many reforms in a year.

Broader lessons for place-based security interventions

The Bogotá 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 could be increasing returns to state presence. This combination deserves to be tested at scale, in more contexts.[50]

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.

It is worth noting that the broader policing literature has 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 a 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).

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 two ways. First, 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. Second, we show how multi-level randomization reduces the differential probabilities of assignment to spillover and control conditions that are so problematic for estimation.

Besides illustrating the uses of design, this paper is also a rare example of the practical uses of randomization inference. Bogotá offers a textbook case: units of varying size, with widely different probabilities of assignment to experimental conditions, with spillovers that 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.

[50] 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.

i Vidal, J. B. and T. Kirchmaier (2018). The effect of police response time on crime clearance rates. *Review of Economic Studies.*

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 Additional data and design details

A.1 Patrolling time

Figure A.1.1 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.

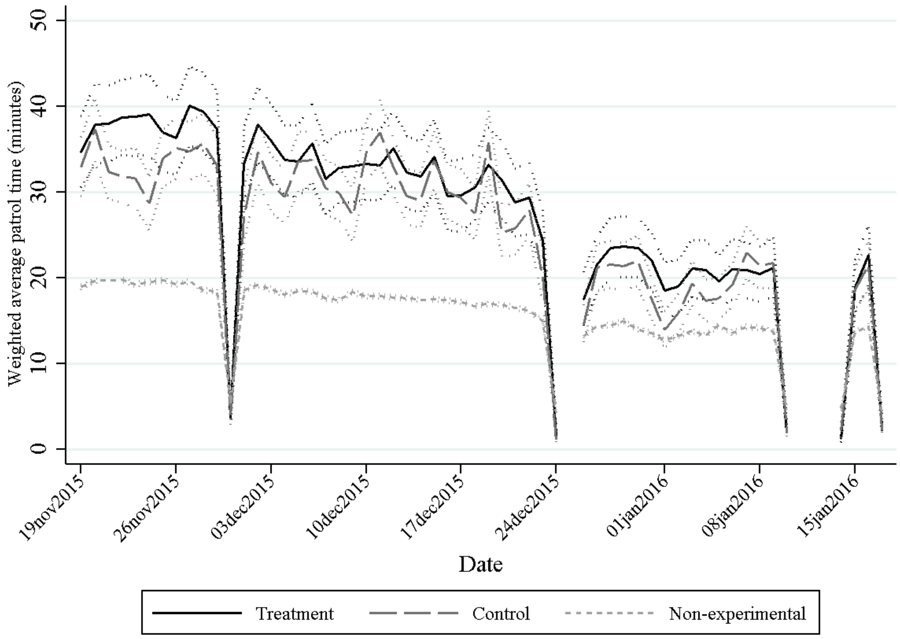
Our estimates of average daily patrolling times are lower in the pre-treatment period because of data quality. During the pre-treatment period not all police patrols had GPS devices and some were working irregularly as the equipment was being piloted. During the treatment period there were also windows of intermittence. These malfunctioning periods, however, affected all streets equally.[51] 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.[52]

[51]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.

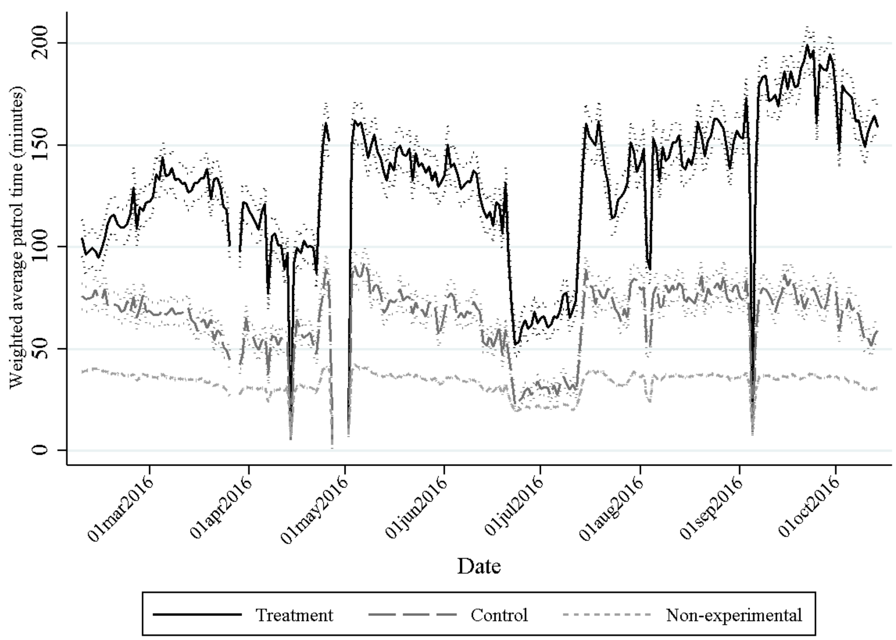
[52]For our estimates, we follow each GPS device chronologically, thus we track the moment at which the device enters a street and when does it leave. We made two assumptions to estimate patrolling time: (i) If we see only one GPS ping in a street and then the device moves to other streets, we impute 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 A.1.1: 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.

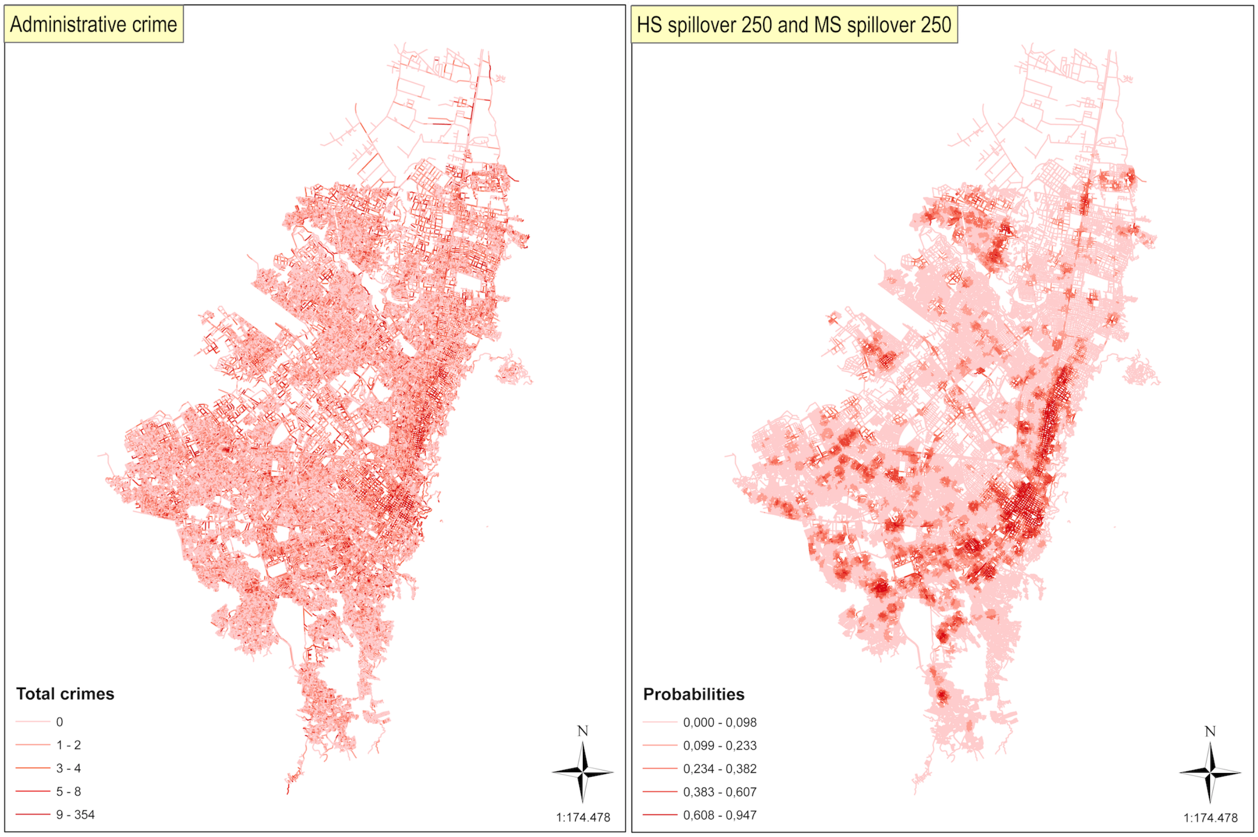
A.2 Inverse probability weighting

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 A.2.1 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 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 A.2.1 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.

Figure A.2.1: 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.

Table A.2.1: IPW bias

		Experimental sample							
		Treatment effect				Spillover effect			
	Interaction included?	Intensive policing	Municipal services	Interaction effect	Both (2+3+4)	Intensive policing	Municipal services	Interaction effect	Both (6+7+8)
Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Insecurity index, z-score (+ more insecure)	No	0.040	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?	Intensive policing	Municipal services	Interaction effect	Both (6+7+8)
	(1)	(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.

A.3 Distribution of officially-reported crime in the sample

Table 1 in this section reports means and mean differences between the three samples. Figure 1 in this section 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. Our experimental sample includes a range of segments from moderate to very high levels of crime. Table 1 reports means and mean differences between the three samples, and Appendix A.3 graphs distributions. 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.

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.[53]

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.

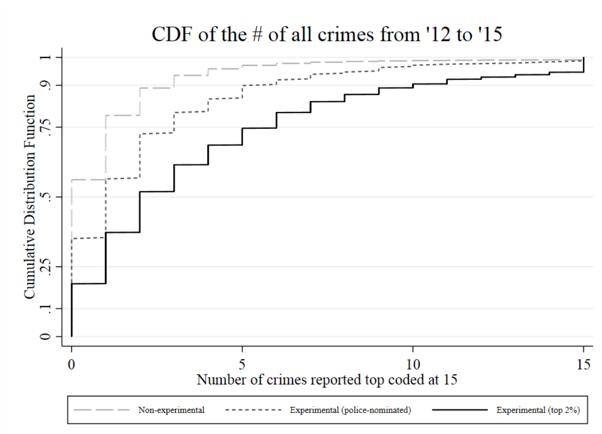
[53] Other reasons include the fact that less serious crimes were given less weight in the sample selection, so that a street with one murder was more likely to enter the experimental sample than one with several muggings. Finally, a small miscalculation in the sample selection admitted a small number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

Table 1: Summary statistics of crime in the experimental and non-experimental samples (administrative and survey measures)

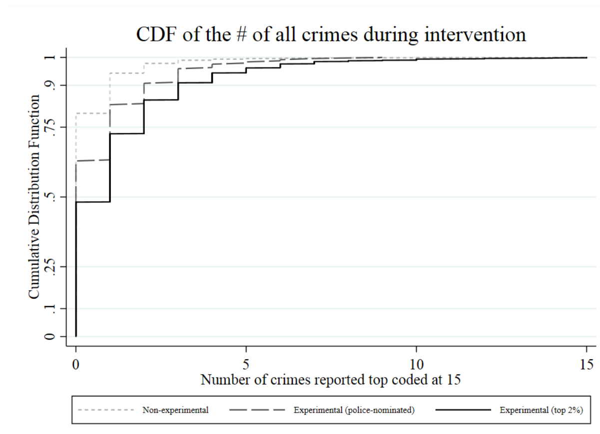
			Experimental sample, by source			t-test of difference in means		
Dependent variable		Non-exp	Police	"Top 2%"		Non-exp vs Police	Non-exp vs Top 2%	Police vs Top 2%
# of officially reported crimes pre- and post-treatment (2012–16)	Mean	1.685	4.763	7.621	Diff.	3.078	5.935	2.857
	SD	2.827	17.380	24.821	SE	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

Figure 1: Cumulative distribution functions (CDFs) of crime measures, by experimental and non-experimental samples

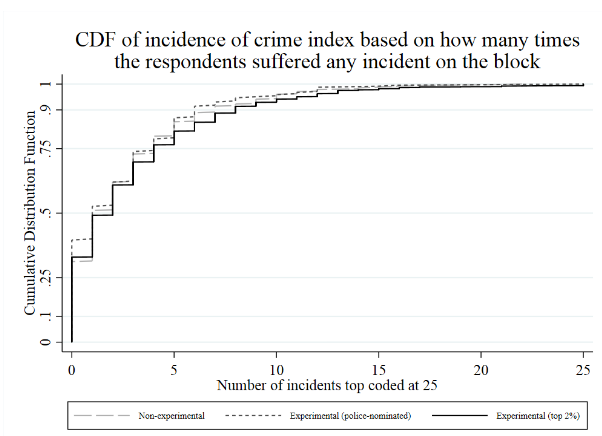
(a) CDFs for the pre-intervention period



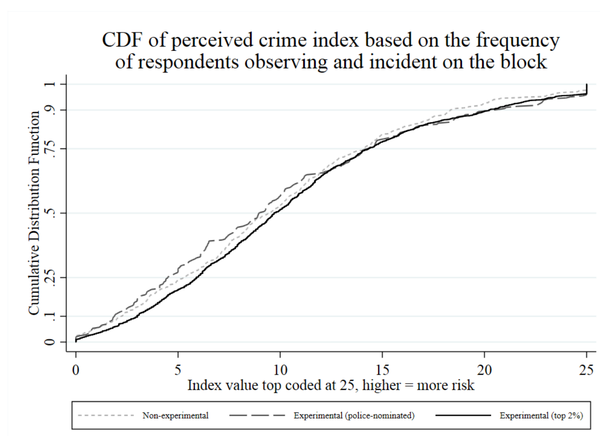
(b) CDFs for the intervention period



(c) CDFs of incidence of crime index



(d) CDFs of perceived crime index



B Statistical power analysis

Figure B.1 takes studies from recent systematic reviews and plot sample size and effect sizes for both direct and spillover effects.[54] 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.[55] Note that even the largest studies do not exceed 50 or 100 treated hot spots, with a similarly modest number of spillover segments.[56] The average effect 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 B.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.

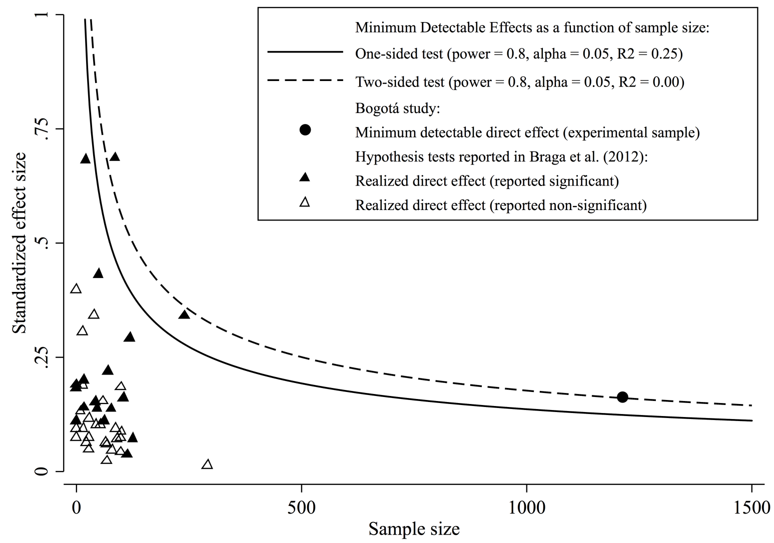
[54]This is one reason why most studies were designed to address direct treatment effects, and spillovers are a secondary outcome. One exception is Weisburd et al. (2006), who study drug and prostitution hot spots. Their findings suggest the benefits from the intervention diffuse to nearby areas.

[55]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.

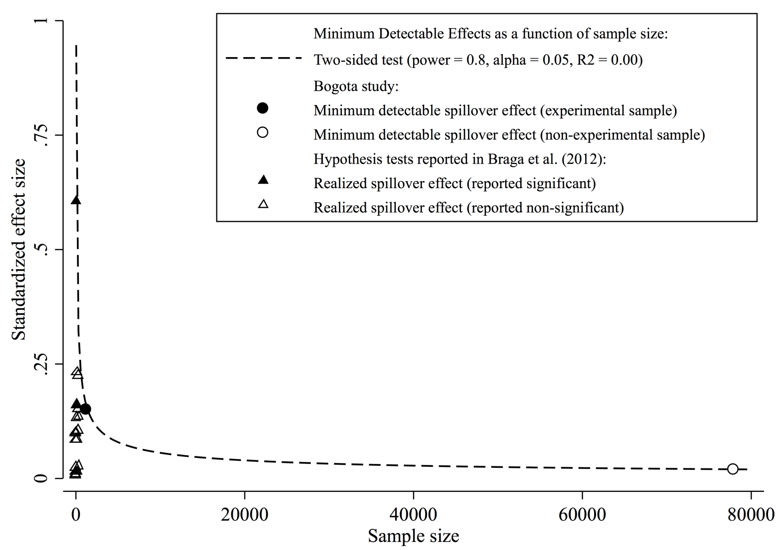
[56]Randomized controlled trials of intensive policing have sample sizes of 110 hot spots (55 treated) in Minneapolis (Sherman and Weisburd, 1995), 56 hot spots (28 treated) in Jersey City Weisburd and Green (1995), 24 hot spots (12 treated) in a different intervention in Jersey City (Braga et al., 1999), 207 hot spots (104 treated) in Kansas City (Sherman and Rogan, 1995), 100 hot spots (50 treated) in Oakland (Mazerolle et al., 2000), 34 hot spots (17 treated) in Lowell (Braga and Bond, 2008), 83 hot spots (21 treated with police patrols and 22 with problem oriented policing) in Jacksonville (Taylor et al., 2011), 120 hot spots (60 treated) in Philadelphia (Ratcliffe et al., 2011), and 42 hot spots (21 treated) in Sacramento (Telep et al., 2014). Interestingly, the first hot spots study was conducted in Minneapolis in 1989 and had a larger sample size with 250 residential addresses of which 125 were assigned to treatment and 250 commercial addressees of which also 125 were assigned to treatment Sherman et al. (1989). One of the only other large studies, by a subset of this paper's author's, is in the Colombian city of Medellín, with 384 of 967 hot spots treated Collazos et al. (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.

Figure B.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.

C Additional results and robustness analysis

C.1 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.[57]

Overall, we see little evidence that the interventions increased trust in or legitimacy of the state. Table C.1.1 reports ITT effects using equation 1. 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.

[57] 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 C.1.1: Impacts on state legitimacy allowing spillovers within 250m, with RI p-values

		ITT of assignment to:				Impact of spillovers <250m:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both interven-tions	Sum of (2), (3), and (4)	Any intensive policing	Any municipal services	Both interven-tions	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 1, estimating the direct effects of the two interventions (Columns 2 to 4) and the spillover effects (Columns 6 to 8) via a WLS regression of each outcome on treatment indicators, spillover indicators, police station (block) fixed effects, and baseline covariates. Columns 5 and 9 report the sum of the three preceding coefficients. The three measures come from our citizen survey.

C.2 Tests of spillovers

Table C.2.1 reports the p-values from general tests 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.2.1 are weak, there is a reasonable argument for calculating treatment effects ignoring spillovers.

Table C.2.1: Testing for spillovers: F-tests of weighted mean differences between control regions

	p-value from F-test of joint significance			
	Experimental sample (N = 1,919)		Non-experimental sample (N=77,848)	
Outcome	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 1 and the adjoining cells in column 4. This table reports the p-value from the F-test of those four mean differences.

C.3 Program impacts without covariates

Table C.3.1 shows the same results as Table 5, but omits baseline covariates, using only municipal services eligibility, station fixed effects, and an indicator for being in the experimental sample in the right hand side of equation 1. Without baseline covariates, the effect of "both interventions" is greatly diminished, particularly its effect on the number of crimes reported: comparing column 6 in the last row of Table C.3.1 in the main paper to the equivalent cell in Table C.3.1, below.

Tables C.3.2 and C.3.3 are equivalent to Tables 6 and 7, from the main paper, respectively. We multiply each effect by the number of streets in each treatment or spillover status. While Table C.3.1 shows that treatment effect of "both treatments" is sensitive to the inclusion or exclusion of baseline covariates because there are an order of magnitude more spillover than treatment streets, this difference is not reflected in the calculation of aggregate effects. Spillover effects are roughly the same, causing aggregate effects to be similar with and without covariates.

Table C.3.1: Program impacts on security in the experimental sample, accounting for spillovers within 250m, with p-values from randomization inference, with interaction between treatments

Dependent variable	Control mean	Effects without interaction: Any intensive policing	Effects without interaction: Any municipal services	Effects with interaction: Any intensive policing	Effects with interaction: Any municipal services	Effects with interaction: Both interventions	Effects with interaction: Sum of (2), (3), and (4)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Insecurity index, z-score (+ more insecure)	-0.003	-0.077	-0.209	-0.0704	-0.1976	-0.0339	-0.3019
		0.740	0.176	0.7160	0.2010	0.9300	0.3130
Perceived risk index, z-score (+ riskier)	0.049	-0.082	-0.173	-0.0621	-0.1462	-0.0727	-0.2810
		0.540	0.111	0.6310	0.1890	0.6860	0.1300
Crime index, z-score (+ more crime)	-0.054	-0.046	-0.175	-0.0552	-0.1830	0.0163	-0.2219
		0.983	0.371	0.8760	0.3270	0.8470	0.5940
Perceived & actual incidence of crime, z-score (survey)	0.059	-0.006	-0.175	-0.0233	-0.2026	0.0722	-0.1537
		0.747	0.116	0.9850	0.0640	0.3910	0.6540
# crimes reported to police on surveyed segments (admin)	0.743	-0.110	-0.131	-0.1021	-0.1006	-0.0939	-0.2966
		0.739	0.904	0.8520	0.9210	0.6470	0.5740

Table C.3.2: 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			
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	-0.112	*0.337*	756	-84.882	-0.070	*0.718*	756	-52.693
Municipal services	-0.117	*0.326*	201	-23.512	-0.036	*0.870*	201	-7.318
Both					-0.207	***0.316***	75	-15.550
Subtotal				-108.394				-60.011
B. Spillover effect								
Intensive policing	0.015	*0.200*	52095	804.612	0.024	*0.143*	52095	1227.737
Municipal services	0.021	*0.905*	21286	443.620	0.032	*0.290*	21286	689.499
Both					-0.021	*0.277*	15772	-335.874
Subtotal				1248.232				1917.237
Net increase (decrease) in crime				1139.838				1505.801
			95% CI	(-2579 , 2482)			*95% CI*	(-1859 , 2533)
			90% CI	(-1102 2230)			*90% CI*	(-1461 2246)

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.3.3: Aggregate impacts on crimes by type (mean and confidence intervals) without covariates

	without interaction			*with interaction*		
	Effect	95% CI	90% CI	Effect	95% CI	90% CI
	(1)	(2)	(3)	(4)	(5)	(6)
All crime	1139.8	(-1440 ,-2482)	(-1102, 2230)	1505.8	(-1859, 2533)	(-1461, 2246)
Property crime	1152.2	(-827, 2322)	(-521, 2127)	1682.1	(-920, 2664)	(-586, 2383)
Violent crime	-12.3	(-1039, 535)	(-896, 411)	-176.3	(-1312, 312)	(-1183, 177)
Homicides and sexual assaults	-44.9	(-188, 61)	(-165, 41)	-64.1	(-238, 51)	(-210, 24)
Difference between property and violent crime	1164.5			1858.4		
p-value	0.000			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.

C.4 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.4.1 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 within 250m of a treated street.

Results from both tables are similar, hence we focus on the no-spillovers case (Table C.4.1). 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 interaction).

Table C.4.1: Instrumental variables results (full sample)

	All experimental streets			Excluding control hotspots within 250m of treated hotspots		
		Instrumental variables			Instrumental variables	
Dependent variable	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.

C.5 Back-of-the-envelope calculation of the effects of scaling the dual treatment and no treatment

In Table C.5.1 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 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.

Table C.5.1: 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.

C.6 Aggregate effects for crime subgroups

Table 7 in the main paper shows the aggregate effects of property crime, violent crime, and homicides and compares their effects. In Tables C.6.1, C.6.2, and C.6.3, we display the aggregate effects on crime subgroups: property crime, violent crimes, and homicides and rapes only, like Table 6 shows in the main paper for the total number of reported crimes. We show the aggregate for each treatment and spillover condition separately, and how they add up to the aggregate effect.

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			
Impacts of treatment	Coeff. (1)	RI p-value (2)	# segments (3)	Estimated total impact = $(1) \times (3)$ (4)	Coeff. (5)	RI p-value (6)	# segments (7)	Estimated total impact = $(5) \times (7)$ (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	(-316, 1941)			*95% CI*	(-394, 2413)
			90% CI	(-167, 1769)			*90% CI*	(-136, 2193)

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 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	(-894, 347)			*95% CI*	(-1126, 234)
			90% CI	(-786, 250)			*90% CI*	(-1012, 115)

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.6.3: 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			
	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	-0.016	*0.058*	756.000	-11.868	-0.010	*0.299*	756.000	-7.616
Municipal services	-0.002	*0.920*	201.000	-0.329	0.007	*0.531*	201.000	1.353
Both					-0.022	*0.215*	75.000	-1.644
Subtotal				-12.197				-6.262
B. Spillover, experimental sample								
Intensive policing	0.00	*0.49*	52095.00	-48.16	0.00	*0.26*	52095.00	-84.98
Municipal services	0.00	*0.69*	21286.00	0.72	0.00	*0.46*	21286.00	-16.61
Both					0.00	*0.48*	15772.00	23.52
Subtotal				-47.45				-101.60
Net increase (decrease) in crime				-59.64				-85.98
			95% CI	(-176, 56)			*95% CI*	(-229, 43)
			90% CI	(-159, 43)			*90% CI*	(-200, 24)

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

Table C.7.1 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 estimated separately. Conclusions do not change materially.

Table C.7.1: 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.

C.8 Program effects without re-weighting and randomization inference

Tables and C.8.1 and C.8.2 reproduce Table 5 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.8.1: Naïve treatment effects without interaction

Dependent variable	Control mean	ITT of assignment to: Any intensive policing	ITT of assignment to: Any municipal services	Impact of spillovers <250: Any intensive policing	Impact of spillovers <250: Any municipal services
	(1)	(2)	(3)	(6)	(7)
Insecurity index, z-score (+ more insecure)	0.066	0.007	-0.099	0.117	0.207
		[.056]	[.070]	[.059]**	[.053]***
Perceived risk index, z-score (+ riskier)	-0.018	-0.022	-0.113	0.053	0.129
		[.056]	[.068]*	[.059]	[.049]***
Crime index, z-score (+ more crime)	0.128	0.035	-0.052	0.142	0.216
		[.054]	[.072]	[.061]**	[.059]***
Perceived & actual incidence of crime, z-score	0.002	0.019	-0.078	0.085	0.147
		[.060]	[.073]	[.063]	[.053]***
# crimes reported to police on street segment	1.334	0.055	0.010	0.214	0.296
		[.089]	[.119]	[.110]*	[.118]**

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). Standard errors are clustered using the following rules: For all treated segments except with cluster size 2, each segment is a cluster. For all other untreated segments, each segment gets its own cluster ID. For entirely untreated quadrants, they form a cluster. For quadrants with exactly 2 units assigned to treatment, those units form a cluster.

Table C.8.2: Naïve treatment effects with interaction

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.

C.9 Continuous spillover results with interaction term

Table C.9.1 reports the same set of specifications as Table 8 in the main paper, but with the interaction between policing and municipal services included for all regressions. As mentioned when discussing Table 8 , the results of this robustness test are broadly similar across specifications: without the interaction term, direct treatment effects have negative and non-significant effects crime. The interacted specification, on the other hand, shows these effects to be concentrated in streets that received both policing and municipal services.

Table C.9.1: Estimated direct and spillover effects using alternative covariate adjustments and methods of spillover estimation with an interaction term, with RI p-values

		ITT of assignment to:				Impact of spillovers <250m:			
Dependent variable	Control mean	Any intensive policing	Any municipal services	Both interventions	Sum of (2), (3), and (4)	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	0.743	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	0.250	-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	0.494	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 indicator, without covariates									
# of total crimes	0.743	-0.070	-0.036	-0.207	-0.313	0.024	0.032	-0.021	0.035
		0.718	*0.870*	*0.316*	***0.082***	*0.143*	*0.290*	*0.277*	***0.089***
# of violent crimes	0.250	-0.047	-0.030	-0.072	-0.150	0.000	-0.012	0.008	-0.004
		0.268	*0.695*	*0.284*	***0.006***	*0.535*	*0.265*	*0.716*	*0.960*
# of property crimes	0.494	-0.022	-0.007	-0.135	-0.164	0.023	0.045	-0.029	0.039
		0.962	*0.982*	*0.430*	*0.252*	*0.090*	***0.073***	*0.112*	***0.036***
C. Spillover intensity									
# of total crimes	0.743	0.034	0.101	-0.413	-0.278	-0.008	0.004	-0.008	-0.013
		0.851	0.525	***0.079***	***0.043***	0.951	0.947	*0.978*	*0.991*
# of violent crimes	0.250	-0.011	0.009	-0.108	-0.110	-0.008	-0.004	0.007	-0.005
		0.600	0.815	*0.112*	***0.018***	0.910	0.930	*0.813*	*0.977*
# of property crimes	0.494	0.045	0.092	-0.305	-0.168	0.000	0.007	-0.015	-0.008
		0.999	0.507	*0.155*	*0.146*	0.940	0.982	*0.973*	*0.970*
D. Exponential decay									
# of total crimes	0.743	0.025	0.065	-0.408	-0.318	0.086	-0.014	-0.041	0.031
		0.807	0.762	***0.090***	***0.018***	**0.067**	0.643	*0.143*	*0.595*
# of violent crimes	0.250	-0.016	0.002	-0.102	-0.116	0.010	0.002	0.004	0.016
		0.484	0.966	*0.135*	*0.013*	0.550	0.889	*0.700*	*0.425*
# of property crimes	0.494	0.041	0.062	-0.306	-0.202	0.076	-0.016	-0.045	0.015
		0.984	0.758	*0.161*	***0.098***	**0.051**	0.545	***0.081***	*0.784*
E. Linear decay									
# of total crimes	0.743	0.027	0.075	-0.407	-0.305	0.026	-0.017	-0.001	0.007
		0.824	0.708	*0.092*	***0.027***	0.287	0.221	*0.920*	*0.802*
# of violent crimes	0.250	-0.016	0.002	-0.101	-0.115	0.005	0.002	-0.001	0.006
		0.502	0.988	*0.144*	***0.016***	0.530	0.761	*0.751*	*0.516*
# of property crimes	0.494	0.043	0.073	-0.306	-0.190	0.021	-0.019	0.000	0.001
		0.996	0.659	*0.164*	*0.119*	0.332	0.110	*0.979*	*0.964*

Notes: Randomization inference p-values are in italics. This table estimates the coefficients on spillovers, $\check{\lambda}$, using equation 1 for panels A and B, and equation 2 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.

References

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

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.

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*.

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.

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

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.

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.

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.

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. 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.