Pre-Analysis Plan for "The Impact of Hotspot Policing and Municipal Services on Crime: Experimental Evidence from Bogotá"

Christopher Blattman Donald Green Daniel Ortega Santiago Tobon*

August 30, 2016

Contents

*Blattman (corresponding author): Columbia University, SIPA and Political Science, chrisblattman@columbia.edu; Green: Columbia University, Political Science, dpg2110@columbia.edu; Ortega: Latin America Development Bank - CAF, dortega@caf.com; Tobon: Universidad de los Andes, Economics, s.tobon11@uniandes.edu.co. Research assistance was provided by Juan Carlos Angulo, Daniela Collazos, Eduardo Garcia, Sofia Jaramillo, Richard M. Peck, Patryk Perkowski, and María Aránzazu Rodríguez Uribe.

1 Introduction

1.1 Background

Hotspots In most cities in the world, crime is highly concentrated in a small number of places. For example, out of 136,984 street segments in Bogotá, Colombia, only 2,970 accounted for all homicides between January 2012 and September 2015.[1] Other forms of crime are similarly concentrated. Police usually refer to these areas as crime "hotspots".

Hotspots policing Policing strategies and tactics that focus on such areas are commonly referred to as hotspots policing interventions (Weisburd and Telep, 2016). Hotspot policing has become one of the most common and popular approaches to crime reduction. Hotspots policing interventions cover a range of police responses that share a focus on police and enforcement resources on the locations where crime is highly concentrated (Weisburd and Telep, 2016). The specific strategies can be as simple as drastically increasing the dosage of policing time at crime hotspots (e.g. Sherman and Weisburd (1995) in Minneapolis and our case in Bogotá), or more complex strategies such as the three step program (identifying the problem, developing a tailored response and maintaining crime control gains) in Jersey City (Weisburd and Green, 1995).

In the past two decades, a series of rigorous evaluations have argued that hotspots policing strategies can be effective in reducing crime. Braga et al. (2012) conduct a systematic review of this literature. They identify 25 tests of hotspots policing in 19 experimental or quasi-experimental eligible studies. 20 cases report significant crime control gains. Along with ongoing evaluations in Medellin (Collazos et al., 2016) and Trinidad and Tobago (Sherman et al., 2014), the Bogotá hotspots policing experiment is one of the first interventions of this type in Latin America, the one with the biggest sample size and the first one specifically addressing the issues of network spillovers from the design stage.[2]

"Broken windows" and municipal services Police and scholars have also argued that disorder on the streets, in the form of trash, graffiti, broken windows, and nonfunctional street lights, may promote some forms of criminal behavior – the so-called "broken windows" hypothesis, whose name stems from Wilson and Kelling (1982).

The idea is that disorder contributes to an area becoming a hotspot. The economic theory of crime introduced by Becker (1968) argues that increasing the probability of apprehension and punishment prevents criminals from taking part in illegal activities. The broken windows hypothesis reconciles with this theory by introducing the subjective perception of apprehension and punishment that criminals have. This is, the conditions of an environment may carry signals about social norms and enforcement capacity. If it is the case that the environment presents itself as highly disordered, criminals may subjectively believe that police presence and other enforcement efforts are weak at that location.

Some experimental studies favor the broken windows hypothesis. For instance, Braga et al. (1999) report significant reductions in crime following a combined treatment of intensive arrests, improvements in the physical environment and provision of social services in a randomized controlled trial at Jersey City. Also, Braga and Bond (2008) conducted an experiment in Lowell, Massachusetts aimed at testing the effect of intensive arrests and environmental interventions. These later included surveillance cameras, lighting and clearance of abandoned buildings. The authors report significant reductions in crime stemming from each of these interventions.

Intervention spillovers The success of any of these interventions relies on crime being reduced rather than displaced. Hence measuring spillover effects is essential to any cost-effectiveness analysis.

[1] A street segment is a block of street between two corners.

[2] For instance, previous randomized controlled trials have sample sizes of 110 hotspots (55 treated) in Minneapolis (Sherman and Weisburd, 1995), 56 hotspots (28 treated) in Jersey City (Weisburd and Green, 1995), 24 hotspots (12 treated) in a different intervention in Jersey City (Braga et al., 1999), 207 hotspots (104 treated) in Kansas City (Sherman and Rogan, 1995), 100 hotspots (50 treated) in Oakland (Mazerolle et al., 2000), 24 hotspots (12 treated) in Lowell (Braga and Bond, 2008), 83 hotspots (21 treated with police patrols and 22 with problem oriented policing) in Jacksonville (Taylor et al., 2011), 120 hotpots (60 treated) in Philadelphia (Ratcliffe et al., 2011), 42 hotspots (21 treated) in Sacramento (Telep et al., 2014) and 967 hotspots (384 treated) in Medellín (Collazos et al.).

There are two reasons to focus on spillovers, the first being that spillover from treatment to control hotspots would bias the simplest causal estimate on treated areas. Identification of causal effects when conducting randomized controlled trials rests on the Stable Unit Treatment Value Assumption, or SUTVA (Rubin, 1990). This implies there should be no interference between experimental units, i.e. potential outcomes for a unit should reflect only its treatment status. However, under the rational approach to crime, it would be sensible for criminals to displace whenever their costs rise under a certain context. In particular, if more police presence increases the expected costs of crime for a criminal at his preferred location, it would be rational for him to displace to another less costly place. Moreover, allocation of police efforts at specific places may imply, *ceteris paribus*, that efforts at other places are reduced. Previous experiments generally do not directly address this SUTVA violation. Studies as Weisburd and Green (1995); Braga et al. (1999); Braga and Bond (2008); Ratcliffe et al. (2011); Taylor et al. (2011); Telep et al. (2014) assess spillover effects by comparing crime outcomes at catchment areas surrounding treatment and control units (usually with buffers of one or two blocks, or 500 feet). Some studies such as Braga et al. (1999); Braga and Bond (2008); Taylor et al. (2011); Telep et al. (2014) introduce specific hotspot selection criteria aimed at easing the identification of spillover effects. Specifically, they conduct a cleaning process so that a given hotspot is far enough from any other. However, no such process is able to account for the SUTVA violation. For instance, patrols heading to treated hotspots may treat surrounding street segments, so that comparing neighboring areas of treatment units to neighboring areas of control units does not identify spillover effects.

The second reason to focus on spillovers is to assess whether hotspot policing simply decreases crime to streets outside the experimental sample, in effect creating new hotspots. The existing evidence suggests that immediate spatial spillovers are low, but besides identification problems and statistical power issues with specific studies, other types of spillovers are poorly identified or ignored.

1.2 Aims

The aim of this study is to estimate the direct and indirect (spillover) effects of a large-scale hotspot policing intervention and a large-scale municipal services intervention, and to identify any interaction between the two. Our main focus will be on the effects of the hotspot policing intervention because there may not have been enough time for the municipal services treatment to affect outcomes It is our aim to improve the estimation of direct causal and spillover effects compared to prior studies. In order to do so, it is necessary to model potential outcomes before the intervention begins. For the case of Bogotá, as shown below, we model 16 potential outcomes, according to treatment assignment to both interventions and proximity to treated units. This design allows us to assess whether hotspot policing or municipal services reduce crime in the aggregate or merely displaces crime. Moreover, it allows to identify how large the bias induced by displacement may be in existing impact estimates at hotspots.

1.3 Interventions

The Bogotá police are testing two interventions: hotspots policing and the provision of municipal services, or a municipal services treatment, in a 2×2 factorial design.

The hotspot policing intervention consists of increasing the dosage of police patrolling time from about 55 minutes per day per hotspot street segment (defined in section 2) to 90, divided in six entries of 15 minutes each. This entry time is not arbitrary and is rather based on previous evidence from Koper (1995) and Telep et al. (2014), which find decreasing returns on crime control after 15 minutes of police presence. Moreover, police patrols will be given specific instructions on how to distribute entries during the day. Hotspots located nearby bars and night clubs will have three entries during the day and three during the night. Other hotspots will have five entries during the day and one during the night. For hotspots in the control group, police will not receive any special instructions and will be free to patrol as they see fit. Activities while patrolling are standard, i.e. criminal record checks, door-to-door visits to the community, arrests, drug seizures, etc. Also, when necessary, police patrols will focus on problem-oriented policing strategies with support from different police branches as youth and juvenile, or counter narcotics specialized agents. The intervention began on February 9, 2016 and will continue until October 14. The research team did not see any outcome data before this PAP was registered.

The municipal services intervention consists of sending a municipal team to selected hotspots to clean up streets in order to promote more informal social control by residents. The municipal services team is in charge of repairing street lights, tree pruning, cleaning non-artistic graffiti, and collecting garbage. This intervention is conducted in two stages. The first stage is a diagnosis visit in which activities needed at the hotspot are identified. The second stage comes after the diagnosis and consists of the actual intervention which is scheduled and organized by the Mayoral Administration of Bogotá. The Mayoral Administration reports administrative records of the agenda and progress made as a way of measuring compliance. Two different offices within the Mayoral Administration are in charge of conducting this intervention. A first office carries out activities regarding street lights and a second office is in charge of all other activities. Therefore, compliance measures and diagnosis are received separately.The intervention began on April 11, 2016 and will continue until the end of the hotspots policing intervention.

2 Experimental design

2.1 Units of analysis

Bogotá has 19 police stations, one per localidad, and each localidad contains 55 quadrants on average. Quadrants are groups of around 130 street segments and are the basic units for police service planning (but not other municipal services). Each quadrant has a police patrol of two police agents 24 hours a day, divided in three shifts of eight hours each. Quadrants aggregate into CAIs (a local police base), which in turn aggregate into Police Stations.

Our unit of analysis is the street segment.[3] We have geo-coded reported crime data from the National Police starting on January 2012. Specifically, we have each reported crime with its exact location coordinates and type of crime. In order to create a street-segment level crime index, we build a geo-fence of 40 meters around each street segment and assign a crime to a street segment whenever it falls within its geo-fence. If there is a crime within two or more geo-fences we assign the crime to the closest street segment using Euclidean distances.

Note that only major crimes tend to be reported, such as homicides, car thefts, or major property thefts. The majority of assaults, muggings, and petty thefts are not reported. This is because reporting a crime requires the aggrieved party to physically go to one of the 19 police stations and complete a report. It cannot be done by patrolling police officers or at the CAI. The completeness of these major crime data are unknown at present.

Note also that the location of some crimes may be misrecorded. For instance, sometimes the crime is reported as taking place at the police station where it is reported, and homicides may be reported at the place of death (e.g. the hospital) rather than the place where the injury occurred. We plan to hire an RA to correctly geocode this data but this will be contingent on funding and feasibility.

By mapping crimes to their geographic location, we end up with data on reported crime for all 136,984 street segments in the urban area of Bogotá. Table 1 contains street-segment crime data and other descriptive statistics for all street segments in Bogotá from January 2012 to September 2015.[4]

The *aggregate crime index* is a weighted sum of reported homicides, assaults, theft from person, car theft and motorcycle theft. The violent crime index aggregates homicides and assaults, and the property crime index aggregates theft from person, car theft and motorcycle theft. Weights for the indexes are assigned based on the average prison sentence according to Colombian law, which proxy for the social costs of crime[5].

We will evaluate the impact of the intervention by using three different data sources. First, we will use GPS data on police location every 30 seconds to investigate how police reallocate their time spent in each treatment condition.[6] Second, we will analyze crime reports collected by the police. Third, we will conduct

[3]For a discussion on the selection of such units of analysis in assessing crime, see Weisburd et al. (2012).

[4]For 2012 and 2013, about a quarter of reported crimes cannot be geo-coded because of deficiencies in the address data. From 2014 onwards, the crime data come with a geographically coordinate, but in some cases these coordinates do not fall within any 40 meter fence and are therefore "lost."

[5]For the aggregate crime index, weights are: 0.300 for homicides, 0.112 for assaults, 0.116 for theft from person and 0.221 for car and motorcycle theft. The weight for homicides was cut by half in order to avoid every street segment with one homicide in the past four years to become a hotspot. For the violent crime index, weights are: 0.439 for homicides and 0.170 for assaults. For the property crime index weigths are: 0.345 for car theft from person and 0.655 for car and motorcycle theft.

[6]We will collect a similar compliance measure for themunicipal services treatment: the number of times the municipal team visited the treated hotspot.

baseline and follow-up victimization surveys at 1,500 randomly selected hotpots. These surveys include data on perception, victimization, police service and list experiments designed to retrieve answers on commonly underreported crimes as minor assault, muggings, petty theft, extortion and police corruption.

2.2 Site selection

In collaboration with the National Police, we used the aggregate crime index from January 2012 to September 2015 to classify hotspots as street segments in the top two percent of the distribution. These areas accounted for about 33% of the city's aggregate crime index. Next, we verified with each police station to check that these 2,740 street segments were actually crime hotspots. For example some homicides get reported at the hospitals where the person died, not where the assault occurred. Also, some crimes get reported as if they occurred at the police station because that is the location where the individual filled out the actual police report. After verification with the police, our randomization sample was 1,919 hotspots. These hotspots account for 21% of the city's crime index. Figure 1 displays the geographic dispersion of these 1,919 hotspots.

2.3 Randomization

We implemented two different randomization strategies to assign hotspots to either treatment or control for both interventions.

2.3.1 Hotspot policing

The mayor of Bogotá promised to deliver at least 750 treated hotspots, so our goal was to assign between 750 and 770 hotspots to receive at least 90 minutes of policing per day, with the remaining hotspots assigned to a control group for which the police station would receive no special instructions but would be free to patrol them as they saw fit. Restrictions on the operational capacity of the police implied that any given quadrant could not have more than two treated hotspots. We began by randomly assigning each quadrant with at least one hotspot to either treatment or control with a treatment probability of 0.60, blocking by police station. For quadrants assigned to treatment, we then assigned hotspots to treatment or control using the following rule:

- Quadrants with one or two hotspots: assign both to treatment.
- Quadrants with more than two hotspots: randomly assign two to treatment and the rest to control.[7]

A randomization was deemed successful only if the number of hotspot segments assigned to treatment was between 750 and 770.[8] Our randomization procedure assigned 756 hotspots to treatment and 1,163 to control.[9] Treated hotspots account for 24% of the aggregate crime index, while untreated hotspots account for 31% of the aggregate crime index. Section 6.7 includes power calculations for this allocation. The statistical approach described below takes the complexities of the randomization procedure into account when calculating point estimates and p-values.

2.3.2 Municipal services

Determining eligibility for the municipal services treatment required a segment-level measure of disorder. For this reason, we sent out enumerators to all 1,919 hotspots to take five photographs and rate hotspots on presence of graffiti, municipal services, garbage, boarded-up buildings, and run-down buildings. Enumerators were able to reach 1,534/1,919 hotspots.[10]

[7] This restriction was introduced because of operational constraints from the National Police: no quadrant should have more than two treated hotspots.

[8] Because of our blocking strategy, the number of treated hotspots falls below 750 in some randomizations. The mayor of Bogotá promised to deliver 750 treated hotspots, so we restricted our randomizations to only those where at least 750 hotspots were assigned to treatment, but no more than 770 (the maximum the police could handle).

[9] After randomization, it was discovered that one quadrant assigned to treatment was actually not a part of our experimental sample because it was exclusively policed by a special narcotics unit. We keep the hotspot segments associated with this quadrant (2 treatment, 1 control) in our sample and treat them as non-compliant.

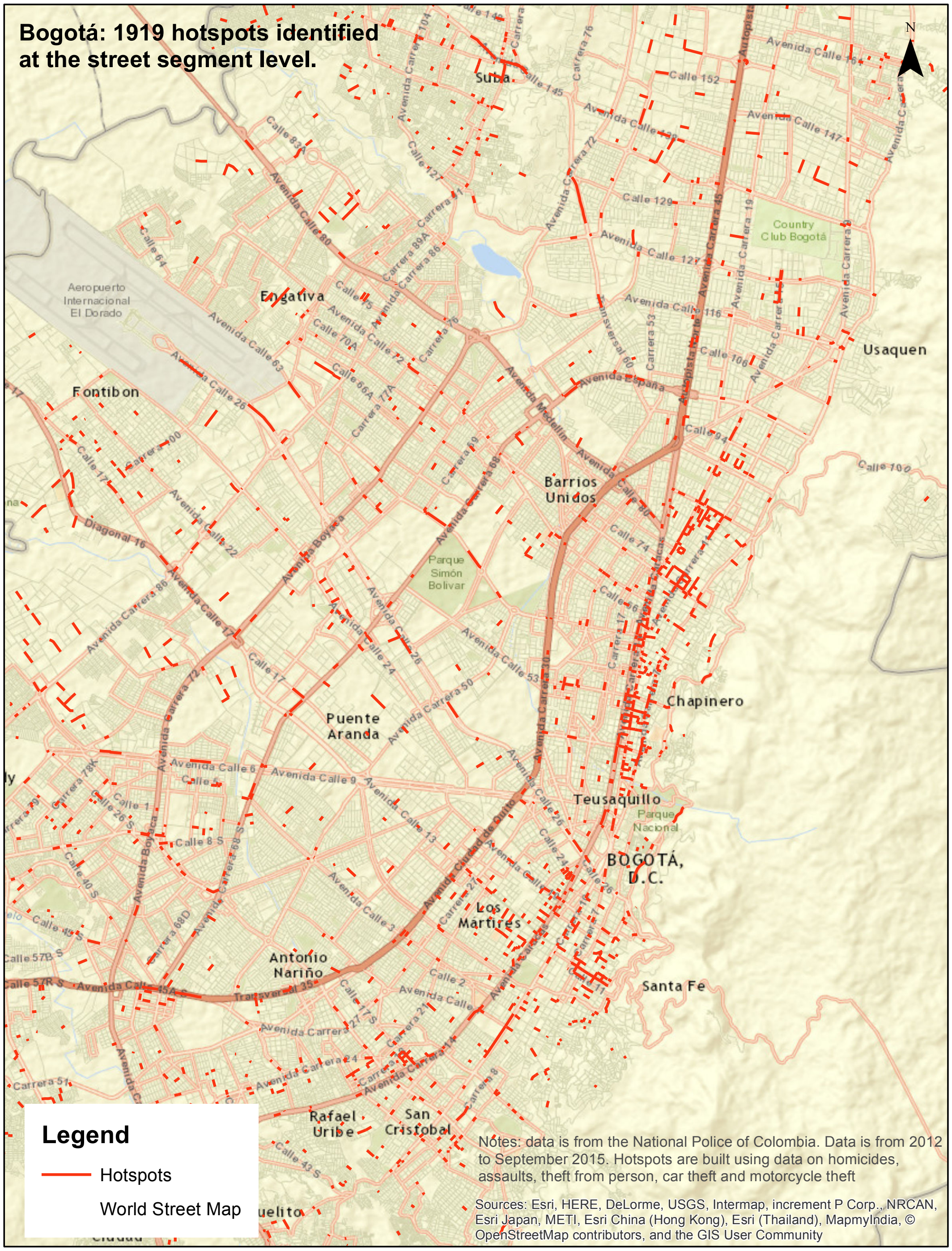
[10] Inaccessible hotspots had less overall crime and property crime, but more violent crime. They were also farther away from social environments like churches and shopping centers (Summary statistics not shown).

Table 1: Aggregate crime levels and other characteristics at the street segment level. Bogotá, Colombia.

Variable	Summary statistics				
	Street segments	Mean	S.D.	Min	Max
Crime levels at the street segment level (Jan. 2012 - Sept. 2015)					
Homicides (reported cases)	136,984	0.031	0.323	0	25
Assaults (reported cases)	136,984	0.082	26.938	0	9,409
Violent crime index	136,984	0.153	4.589	0	1,600
Theft from person (reported cases)	136,984	1.718	46.705	0	10,404
Car theft (reported cases)	136,984	0.083	1.613	0	169
Motorcycle theft (reported cases)	136,984	0.063	0.605	0	81
Property crime index	136,984	0.688	16.173	0	3,589
Aggregate crime index	136,984	0.332	6.567	0	1,229
Other characteristics					
Length (m.)	136,984	58.896	50.004	2	1,480
Distance to closest police facility (m.)	136,984	621.160	621.160	0.213	4,781.751
Distance to closest commercial facility (m.)	136,984	835.389	777.438	0.326	5,114.206
Distance to closest educational facility (m.)	136,984	326.723	266.421	0.612	3,833.580
Distance to closest park or recreational facility (m.)	136,984	704.265	426.896	0.536	3,811.705
Distance to closest religious or cultural facility (m.)	136,984	524.071	524.071	0.368	5,670.104
Distance to closest health facility (m.)	136,984	961.426	961.426	0.248	9,180.748
Distance to closest transportation facility (m.)	136,984	104.304	95.000	0.002	2,757.408
Distance to other services facility (m.)	136,984	769.454	769.454	0.481	7,360.323
Urban density (sqm built within 100m over length)	136,984	23,886.240	29,032.290	0	956,128
Main usage					
Housing	136,984	0.722	0.448	0	1
Industry or commerce	136,984	0.203	0.403	0	1
Services	136,984	0.074	0.448	0	1
Income level					
Low	136,984	0.495	0.500	0	1
Medium	136,984	0.441	0.497	0	1
High	136,984	0.064	0.245	0	1

Note: Data is by street segment from January 2012 to September 2015. The aggregate crime index is a weighted sum of homicides, assaults, theft from person, car theft and motorcycle theft. The violent crime index aggregates homicides and assaults, and the property crime index aggregates theft from person, car theft and motorcycle theft. Weights are assigned based on the average prison sentence according to Colombian law. For the aggregate crime index, weights are: 0.300 for homicides, 0.112 for assaults, 0.116 for theft from person and 0.221 for car and motorcycle theft. The weight for homicides was cut by half in order to avoid every street segment with one homicide in the past four years becoming a hotspot. For the violent crime index, weights are: 0.439 for homicides and 0.170 for assaults. For the property crime index weights are: 0.345 for car theft from person and 0.655 for car and motorcycle theft. Other services facility refers to services as justice, employment or registry offices. Urban density is measured by adding up all square meters built 100 meters around the street segment, normalized by the segment length. Income levels are based on Estratos, an instrument used in Colombia to target public policies. Estratos are updated by law once every 10 years at most. The last update conducted in Bogotá was in 2013.

Figure 1: Map of hotspots



Notes: This map shows most of Bogotá, although far off zones both north and south are not included. These areas also have hotspots.

Table 2: Breakdown of municipal services need

Index of need (0-5)	#
0	460
1	655
2	348
3	57
4	13
5	1
Total rated	1,534

Notes: The index of need is defined as the sum of graffiti presence, broken lights, garbage on the street, boarded-up buildings and run-down buildings. Photographers could only reach 1,534/1,919 hotspots due to safety concerns.

Table 3: Distribution of hotspots assigned to broken windows

	Assigned to treatment			Assigned to treatment	
Name	#	% of all hotspots	Name	#	% of all hotspots
	(1)	(2)		(3)	(4)
Antonio Narino	15	0.29	Puente Aranda	13	0.29
Barrios Unidos	11	0.17	Rafael Uribe	26	0.26
Bosa	15	0.26	San Cristobal	15	0.32
Candelaria	16	0.27	Santa Fe	23	0.27
Chapinero	35	0.13	Suba	46	0.19
Ciudad Bolivar	33	0.22	Teusaquillo	16	0.25
Engativa	21	0.21	Tunjuelito	15	0.31
Fontibon	21	0.20	Usaquen	21	0.19
Kennedy	43	0.22	Usme	8	0.26
Los Martires	22	0.25			

Notes: The table displays the distribution of hotspots assigned to receive municipal services by localidad. The first and third columns represent the total number of hotspots receiving the treatment in each localidad, while the second and fourth columns display the percentage of all hotspots that receive the treatment.

We created a 0–5 index of need for the broken windows treatment using the data mentioned above. Table 2 displays the breakdown of this index for the 1,534 segments our enumerators were able to rate. 30% of these hotspots show no need for the municipal services treatment.

We restricted eligibility for the municipal services treatment to hotspots with an index score of 1 or greater, and all hotspots our enumerators could not access. To randomize segments, we assigned eligible hotspots into treatment with a probability of $p = 0.25$, blocking by station and hotspot policing assignment (treated, <250m spillover, >250m and <500m, and >500m from a treated unit). We then randomized these selected hotspots into different batches to roll-out the intervention over time. Table 3 displays the distribution of hotspots assigned to broken windows by station. The average localidad has 22 hotspots assigned to receive the treatment.

We batched the units receiving municipal services into two groups. The first group started receiving treatment on April 11. We sent photographers to analyze compliance with the intervention beginning July 1. After analyzing the data, we decided to not move onto the second batch but increase the intensity for the first batch.

2.4 Spillover units

In order to measure spatial spillovers and retrieve the direct causal effect of hotspots policing/broken windows on crime, we differentiate between different control units depending on their distance to treated hotspots. Table 4 breaks down how hotspots are distributed in 16 potential outcomes.

Out of the sample of 1,919 hotspot segments, 756 are assigned to hotspot treatment, 705 are spillover

Table 4: Distribution of treatment and spillover

Distribution of treatment assignments						
		Broken windows assignment				
		Treated	<250m	250m-500m	>500m	All
	Treated	158	263	180	155	756
Hotspot policing assignment	<250m	149	374	125	57	705
	250m-500m	67	77	108	42	294
	>500m	41	28	24	71	164
	All	415	742	437	325	1919

Notes: The table breaks down our sample of 1,919 hotspots into 16 groups based on treatment assignment and distance to other treated units.

segments within 250m of a treated hotspot, 294 are spillover segments between 250m and 500m of a treated hotspot, and 164 are controls greater than 500m from any treated hotspot. Similarly, 415 are treated by the broken windows treatment, 742 are within 250m of a treated hotspot, 437 are between 250m and 500m of a treated hotspot, and 325 are controls greater than 500m away from any hotspot receiving the broken windows treatment. 71 units are considered "pure control" in that they are greater than 500m away from any hotspot receiving either treatment.

3 Baseline data

3.1 Administrative data

We currently have administrative data on crimes, police patrolling time, socio-economic characteristics of all land plots in Bogotá, geo-coded urban infrastructure and location of public surveillance cameras. Specifically, we have:

- Geo-coded data on reported crime from January 2012 to September 2015. Each crime event has information on the location and the type of crime, i.e. if it was a homicide, an assault, theft from person, car theft or motorcycle theft.
- Geo-coded data on previous patrolling activity, specifically for one week on November 2015.
- Geo-coded data on the location of all police facilities (police stations and CAIs, a police station is composed by an aggregation of CAIs).
- Geo-coded data on all land plots in Bogotá, with details on the economic destination of each land plot, a proxy for income at each plot (in Colombia it is called *Estrato* and goes from 1 to 6, 1 being low-low income and 6 high income, it is a tool used for targeted public policies), and number of square meters constructed at each land plot.
- Geo-coded data on the location of all mayor commercial spots in Bogotá.
- Geo-coded data on the location of all educational facilities in Bogotá.
- Geo-coded data on the location of all religious and cultural facilities in Bogotá.
- Geo-coded data on the location of all health facilities in Bogotá.
- Geo-coded data on the location of all transport infrastructure in Bogotá (as bus and BRT stations).
- Geo-coded data on the location of public surveillance cameras monitored by the Police.

With this information, we construct the following administrative baseline data at the hotspot level:

- By using a geo-fence of 40m around each hotspot, for every hotspot we have: number of homicides, assaults, theft from person, car theft, motorcycle theft, aggregate crime index, violent crime index and property crime index, for all crimes committed between January 2012 and September 2015.
- By using a geo-fence of 40m around each hotspot, we have patrolling time (in minutes) for one week of November 2015 for every hotspot.
- Distance to the closest police facility.
- A density measure (number of square meters constructed 100 meters around the hotspot, normalized by the length of the hotspot.
- A predominant income level measure, based on the *Estrato* level of closest land plot.
- A predominant economic destination measure, based on the economic destination of the closest land plots (economic destinations are: housing, industry or commerce, and services).
- Distance to the closest commercial spot.
- Distance to the closest educational facility.
- Distance to the closest religious or cultural facility.
- Distance to the closest health facility.
- Distance to the closest transport infrastructure.
- A dummy variable indicating whether or not the hotspot is within the range of 40m. around a public surveillance camera.

3.2 Survey sampling and subjects

The baseline survey was conducted at 1,500 randomly selected hotspots from January 30 to February 14 2016. We collected data in just three quarters of streets in large part due to short-term budget and time constraints. For the same reasons, the survey extended five days into the hotspots policing intervention. However, since all questions were retrospective we believe contamination stemming from increased police presence may be minimal. We used a complete random assignment procedure: with the experimental sample of 1,919 hotspots we generated uniformly distributed random variables on the interval [0,1) for each hotspot, we sorted the hotspots and selected the top 1,500 for surveys. Two surveys were conducted at each hotspot. People was intercepted in the street. Minors and those who did not live or work in the area were filtered out of the survey. Key variables of interest in the survey were:

- Victimization and security perception:
  - How safe do you feel in this street segment?
  - Have the security conditions improved in this street segment during the past year?
  - Have the security conditions improved in this street segment during the past six months?
  - Have you been a victim of any crime at this street segment during the past six months?
  - Do you have any knowledge of any person being a victim of a crime at this street segment during the past six months?
  - If you were victim of a crime, did you report the crime?
- Perception about policing activities:
  - Do you trust in police surveillance?
  - Do you consider police activity to be good to society?
  - Do you consider the presence of police patrols to be active in this street segment?

  - Are you satisfied with the police response at this street segment?

- List experiments on commonly under-reported crimes:

  - Presence of gangs at the hotspot.
  - Gangs charging for private security at the hotspot.
  - Police corruption at the hotspot.
  - Illegal drug sales at the hotspot.
  - Extortion at the hotspot.

Our endline survey will be conducted at 2,399 segments (1,919 hotspot segments and 480 non-experimental segments) using a survey based largely on the baseline survey.

3.3 Covariate selection

To measure the prognostic ability of our baseline data, we regressed the 2015 aggregate crime index on an indicator for treatment and each of the 2^{17} combinations of covariates for 100 different treatment assignments, weighting by the inverse of the probability of being in its observed experimental condition and restricting the sample to only hotspots with a non-zero probability of being assigned to both treatment and control.

Our analysis revealed that including all covariates worked about as well as including just the optimal covariates so we will include all 17, which are:

- Index of aggregate crime from January 2012 to September 2014;
- Index of violent crime from January 2012 to September 2014;
- Index of property crime from January 2012 to September 2014;
- Total patrolling time from November 19, 2015 to November 29, 2015;[11]
- Urban density;
- Predominant income measure (high, medium, low);
- Predominant economic destination measure (housing, transport, other);
- Distance from police station;
- Distance from commercial area;
- Distance from school;
- Distance from religious center;
- Distance from health center;
- Distance from transportation;
- Distance from other services like justice.
- Dummy variable indicating whether the hotspot is within a 40m. range of any public surveillance camera.

These 17 variables predict around 55% of the variability in the aggregate 2015 crime index.

We also plan to include the following police-level covariates in all of our specifications:

- Police grade of Police Station and CAI commanders (which proxies for police capacity).

[11] In theory it would be better to use the mean or median value for the months in 2015 during which the experiment will take place in 2016. However, this was the only pre-treatment patrolling time data we have access to.

- Time in office (which proxies for knowledge of the location).

We do not plan to use the community survey questions as baseline control variables in all of our specifications. We only surveyed 1,500 hotspot segments at random with two surveys at each hotspot, so we are missing data for 419 hotspots. We only plan to use these as baseline controls when testing the correlation between treatment status and likeliness for an individual to report a crime (see section **??** for more details).

4 Outcomes

4.1 Outcome variables

4.1.1 Primary outcomes

We have two principal outcomes of interest. The first is an index of perceived street risk, which will consist of a 0-4 index of average risk. Respondents will be asked to rate perceived risk on the segment on a scale from 1 to 4 where 1 is "very unsafe" and 4 is "very safe" in the following situations:

- during the day
- at dusk
- for someone to talk on their smartphone on this street
- for a young woman to walk alone after dark on this street
- for a young man to walk alone after dark on this street

The index will be the average across all 5 categories. In order to map survey data from individuals into outcome data for segments, we will take the average score across all surveyed individuals in the segment.

The second index will be an index of crime that includes three equally-weighted components, like in Kling et al. (2007). The first component will be perceived incidence of crime. We will ask how often a list of criminal activities has taken place on this block. For each activity, we will create a 0-7 index (everyday = 7, never = 0). We will then aggregate the average of activities into 3 categories: property crime, violent crime, and victimless crime. The perceived risk component will be the average of these 3 components.

The second component of the crime index will be personal victimization. We ask how often individuals themselves or someone in their family have been victims of a variety of crimes. Our index will either be the percentage of respondent-crime pairs that occurred on the segment or the percentage of respondents that were victimized on the segment (depending on the frequency of the data). We will limit our results to crimes that occurred on the segment for identification purposes. Like before, we will take the average score across all surveyed individuals in the segment. The personal victimization component will be the average of the violent crime subcomponent and the property crime subcomponent.

The third component of the crime index will come from administrative crime data. We will calculate either the number of total incidents or a binary indicator for having any crime during the treatment period (depending on the frequency of data). Unlike the baseline data, we will not weight the crimes by the average prison sentence according to Colombian law. We will divide the data into property crime and violent crime and look within the index to see if treatment was more likely to affect a specific type of crime.[12][13]

[12] One potential issue to consider is whether crime reporting in the administrative data is correlated with treatment (and in what direction). It's important to note that, for the major crimes that are currently part of the aggregate index, some are likely to be reported in most cases–homicides but also vehicle theft, for insurance purposes. For less serious crimes, it's unclear whether we should expect treatment to increase reporting. On the one hand, individuals cannot report crimes directly to the police patrols or municipal teams, so crime reporting seems unlikely to be correlated with treatment mechanically because police are physically present more. This, however, could be the case if treatment led to an increase in the public's trust in the police, knowledge about how to report crimes, or encouragement from a police officer to report crimes, causing individuals to report crimes that would previously go unreported. On the other hand, treatment may cause cops to see more serious crimes being committed, which would lead to increased crime if these crimes were previously going unreported. All of these reporting issues would bias our results to the null, however. Endline surveying will help us assess the completeness of reporting and it's correlation with treatment.

[13] We are considering breaking up crimes into those that are likely to be deterred by policing versus those that are not (i.e. exclude involuntary manslaughter from homicide data, but keep in crimes of passion). We are not sure what data we will have

We are mostly interested in the effect of hotspot policing because there may not have been enough time for the municipal services intervention to affect these outcomes. We will also look at differences in the effects by crime type (violent versus property).

4.1.2 Secondary outcomes

To better understand the effect we observe on our main outcome, we will consider three secondary outcomes.

The first secondary outcome will be an index of police trust and satisfaction created by taking the average of the following questions:

- On a scale from 1 to 4, where 4 is "a lot of trust" and 1 is "no trust", how much trust do you have in the Policía Metropolitana de Bogotá?
- On a scale from 1 to 4, where 4 is "very good" and 1 is "very bad", how would you rate the work the Policía Metropolitana de Bogotá does?
- On a scale from 1 to 4, where 4 is "very likely" and 1 is "very unlikely", how likely would you be to provide information to the Metropolitan Police to help them improve the security of your neighborhood?
- On a scale from 1 to 4, where 4 is "very satisfied" and 1 is "very unsatisfied", how satisfied are you with the Metropolitan Police of Bogotá?

The second secondary outcome will be an index of Mayor's office trust and satisfaction, based on similar questions to the index of police trust:

- On a scale from 1 to 4, where 4 is "a lot of trust" and 1 is "no trust", how much trust do you have in the Mayor's Office?
- On a scale from 1 to 4, where 4 is "very good" and 1 is "very bad", how would you rate the work the Mayor's Office does?
- On a scale from 1 to 4, where 4 is "very likely" and 1 is "very unlikely", how likely would you be to provide information to the Mayor's Office to help them improve street conditions?
- On a scale from 1 to 4, where 4 is "very satisfied" and 1 is "very unsatisfied", how satisfied are you with the Mayor's Office?

The third secondary outcome will be an index of reporting from the following question:

- On a scale from 1 to 4, where 4 is "very likely" and 1 is "very unlikely", how likely would you be toreport a crime to the police or other authorities?

If we get the data and there is enough variability in it, we will also include two other components: the percentage of crimes that are reported (extensive margin, from the survey) while the second will be the average time between crime occurring and crime reporting (intensive margin, from administrative data).

4.1.3 First stage results

We plan to create a first-stage index for policing and one for municipal services. The index for policing will be the average of three normalized components. The first will be policing time, which will be an index of 3 components: patrolling time per day, recorded entrances per day, and total days with no recorded patrolling time. The second will be an index of policing activities, which will be an index of 2 components: police charges and police operative cases such as drug seizures or car recoveries. The third will be a 0-2 index from the following question:

- From Christmas until now, do you think that police presence has increased, stayed the same or decreased in this block?

available, so we are not sure how this analysis will look at the time of the pre-analysis plan publication.

The first-stage index for the municipal services will consist of two components. The first will be the amount of municipal service visits (either the number or a 0-1 indicator, depending on the final data we have available), while the second component will be a 0-2 index from the following question:

- From Christmas until now, do you think that city cleaning services has increased, stayed the same or decreased in this block?

These are not primary or secondary outcomes so we will not adjust p-values for these results. Instead, we will use them to rescale our ITT estimates by the number of additional minutes of policing/municipal service visits induced by treatment. We will use instrumental variables regression to rescale our ITT estimate by these outcomes.

4.2 Multiple comparisons

In order to deal with multiple comparisons, we do the following. We first create family indices of similar variables using equally weighted averages of z-scores, like in Kling et al. (2007). Next we group our outcomes into three types: main outcomes (perceptions of street risk and crime), secondary outcomes (police trust, government trust, and crime reporting), and first stage outcomes (policing first stage and municipal services first stage).

Although p-value adjustment has become a popular approach in dealing with multiple comparisons, we do not believe this is appropriate for our study for the following reasons. Unlike studies that collect many outcomes, we are primarily interested in just two outcomes. As is common in medical trials, we do not adjust for secondary outcomes as these are much more exploatory and we do not have as strong priors about them. We also do not adjust the p-vals for the first stage because these are just manipulation tests and we will use it to rescale any effects we get on our two main outcomes. Finally, we do not plan to adjust p-values for our 2 main outcomes – since we are only concerned with two outcomes, p-hacking is not likely to be an issue.[14][15]

We should note that it is also possible to adjust across potential outcomes (treated, inner spillover, outter spillover) or units (experimental vs non-experimental). We refrain from doing so because these are distinct samples with their own randomization procedures.

5 Hypotheses

5.1 Intention to Treat (ITT) Effects

Our analysis will focus on Intention-to-Treat (ITT) estimates to identify the impact of our intervention regardless of whether the police and/or municipal team complied with the treatment status of each street segment. This analysis addresses the policy question of whether a hotspots policing strategy/broken windows strategy has effects on crime months after the intervention. We hypothesize that not only will treatment affect segments assigned to 90 additional minutes of policing, but there will also be spillover effects onto nearby untreated units.

5.1.1 Hotspot policing

We hypothesize that treatment will increase policing time in treated hotspots and decrease policing time in non-treated spillover segments. We also hypothesize that treatment will decrease crime in segments assigned to hotspot policing. We will test all of these hypotheses using a one-tailed test. For spillover control units, the direction of the crime effect is uncertain: spillover segments might see increases in crime due to displacement or decreases in crime due to deterrence and a diffusion of benefits. For this reason we will use a two-tailed test.

[14] It is possible to make these into a family index of one outcome. This is what other studies like Casey, Glennerster, Miguel's Sierra Leone CDR experiment, or the Oregon health experiment, or Kling and Katz's MTO experiment have done, but these studies had many primary outcomes. Since we only have two, we think this is unneccessary.

[15] To add more evidence that we are not p-hacking we will compare the distribution of p-values in the pre-analysis plan versus those displayed in the paper, as done in Berge et al. 2015

Furthermore, the effects on spillover segments may vary across time. We hypothesize that if the effect of police presence on crime is mostly through deterrence, the reallocation of police presence should only have a transitory effect on crime unless this reallocation leads to a permanent shift in offenders' perceived probability of apprehension. Even if there are positive spillovers in the short run, there may still be some crime displacement over a longer stretch of time, as offenders seek out new low cost locations.

5.1.2 Municipal services/broken windows

We hypothesize that treatment will increase the number of times a segment is visited by the municipal team, and may also reduce crime. We will test both of these hypotheses using a one-tailed test. We anticipate decreases in crime for spillover units and will also test this using a one-sided test.

However, we hypothesize that the spillover effect will be less pronounced than that in the hotspot policing case. If we cannot reject the null that there is no effect for spillover hotspots between 250m-500m away from a hotspot treated with broken windows, we will transfer these units into the pure control group and only consider one spillover: <250m (described more in section 6.2).

5.1.3 Interaction

We hypothesize that segments receiving both the hotspots policing and broken windows treatments will see larger decreases in crime than segments receiving just one. We explain our empirical strategy to estimate this interaction in section 6.3. We will test this hypothesis using a one-tailed test.

5.1.4 Bias

While our randomization procedure is correct, there are certain features in our design, like the geographic clustering of crime in downtown, that may introduce bias in our experiment. This bias isn't an issue when you're randomly assigning individuals to treatment versus control for a cash transfer program, for example. In our experiment, however, a segment's treatment status also depends on the treatment assignment of its neighbors, so the clustering of high crime units makes causal inference a little more difficult.

Luckily, there is a pretty straightforward way of dealing with this. In order to remove the bias in our estimates, we repeat our randomization procedure 1,000 times to get 1,000 new treatment assignments. We use randomization inference and IPW pairwise regressions to regress crime indicators on treatment, baseline controls, and block fixed effects for each set of treatment assignments. We then take the average of all 1,000 point estimates on treatment. This effect should be 0 under the sharp null of no effect for any units, so the average of all these point estimates gives us the bias associated with our design.

We plan to remove this bias for each of the treatment effects we display in the final paper.

5.2 Heterogeneity

5.2.1 Hotspot policing

We plan to examine heterogeneity in the hotspot policing treatment by one main characteristic: baseline level of crime .

The direction of the impact baseline crime level on treatment is unclear. On the one hand, areas with higher crime will experience larger declines because our crime index is bounded by zero, so crime reductions in relatively lower-crime areas will be more difficult. On the other hand, additional policing by two officers may not be enough to break up certain cartels, meaning that areas with lower crime are easier to treat. Therefore we will test this hypothesis with a two-sided test. The heterogeneity analysis are not as important as the key parameters described below because (i) there is a great deal of noise in baseline crime, both from the survey and administrative data, and (ii) since we are intervening on high crime areas only, there is less variation in the "signal", and so the noise to signal ratio is potentially quite great.

5.2.2 Municipal services/broken windows

We plan to examine heterogeneity by baseline level of crime and initial need for broken windows treatment (score on the 0–5 index). We expect that segments with higher crime and larger need will experience larger reductions in crime. We will test both of these hypotheses with one-sided tests. Like the above heterogeneity analysis, the results are not as important as the key parameters.

6 Analysis

6.1 Key parameters

While there are various parameters we could estimate, the following are key parameters in order of importance:

1. Direct treatment effects of the policing intervention;
2. Spillovers from the policing intervention, especially the inner ring, for both the experimental and non-experimental sample
3. Direct treatment effects of broken windows intervention
4. Interaction between direct treatment effects of policing and broken windows
5. Spillovers from the broken windows intervention, for both the experimental and non-experimental sample

Parameters one, two, three and five will be estimated using pairwise regressions described in section 6.2, while parameter four will be estimated using the regression described in section 6.3.

6.2 Pairwise regressions

Consider a hotspot street segment s in quadrant q in police station p. We will calculate Intent-to-Treat (ITT) estimate via the pairwise weighted least squares (WLS) regression:

$$Y_{sqp} = \beta_0 + \theta_{EC} * EC_s + \beta * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{1}$$

where EC_s is the experimental condition of segment s, X is a vector of segment, quadrant, or police station controls, and γ_p is a vector of block fixed effects. We will weight each observation by the inverse of the probability of being observed in its assigned experimental condition, restricting our attention to segments whose probabilities fall between 0 and 1 (See section 6.5 for more details).

Suppose we were to test the direct treatment effect of the policing intervention (parameter one above). Then we would run the regression

$$Y_{sqp} = \beta_0 + \theta_H * H_s + \beta * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{2}$$

where H is an indicator for assignment to hotspot policing. The coefficient of interest is θ_H, which represents the ITT estimate of receiving the hotspot policing treatment on outcome Y relative to segments greater than 500m away from any treated hotspot. Because we use inverse probability weights, we restrict our sample to hotspots which have a non-zero probability of both being treated by the policing intervention and being greater than 500m away from any hotspot receiving the policing intervention.[16]

We can run a similar pairwise regression to test parameter 4: whether the broken windows treatment affected outcomes in spillover units within 250m of treated hotspots:

$$Y_{sqp} = \beta_0 + \theta_{S_{B,250}} * S_{B,250} + \beta * X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{3}$$

In the above, $S_{B,250}$ is an indicator for not receiving the broken windows treatment but being within 250m of a unit assigned to the treatment. Therefore, $\theta_{S_{B,250}}$ represents how being within 250m of a unit receiving broken windows affects outcome Y relative to segments greater than 500m away from any hotspot receiving the broken windows treatment. We will restrict equation 3 to units with a non-zero probability of being assigned to be a spillover unit within 250m of a treated hotspot and a control unit greater than 500m

[16] See section 6.5 for the distribution of probabilities by experimental condition.

away from any hotspot receiving the broken windows treatment. We will use such pairwise regressions to test parameters one, two, and four.

Note that in the previous examples, we considered units more than 500m away from a treated hotspot to be our comparison group. Our preanalysis plan specifies two rings of spillovers (<250m, and 250m-500m) for treatment, but treatment may not produce spillover effects in such ranges. For example, the correct spillover radius for the broken windows treatment may just be the ring of radius 250m. To check whether our spillover ranges are accurate, we first plan to run equation 3 for each level of spillovers. We will then conduct a two-tailed test to determine whether there are differences in outcomes between hotspots in this range versus hotspots greater than 500m away from treated hotspots. If the two-tailed t-test returns $p \geq 0.10$, we will consider the spillover region to be equivalent to the segment >500m away, and collapse the spillover units into pure control units. For example, if we run equation 3 and find no outcome differences between segments between 250m and 500m away from treated hotspots and segments >500m from treated hotspots, we would only have three experimental conditions for the broken windows treatment: treated, <250m, and >250m. Then if we were to estimate the effect of broken windows by running equation 2, θ_H would represent the ITT effect of receiving treatment relative to hotspots >250m (not >500m) away from treated units.[17]

This technique help us in two ways. First it boosts our statistical power, especially when we estimate parameter three. Second, it helps ensure there are meaningful differences in our specified spillover regions. While parameter one is of most importance, we will run parameters two and four first to test our spillover regions and then we will estimate parameters one and three.

6.3 Interaction between treatments

In order to test parameter three, or the interaction between treatment effects, we will run the following regression:

$$Y_{sqp} = \theta_H H_s + \theta_B B_s + \theta_{HB} H \times B_s + \beta X_{sqp} + \gamma_p + \varepsilon_{sqp} \tag{4}$$

where H is an indicator for assignment to hotspot policing and B is an indicator for assignment to broken windows. X is a vector of segment, quadrant, or police station controls, and is a vector of block (police station) fixed effects. Like in our pairwise case, we weight by the inverse of the probability of being in the observed treatment condition.

θ_H would therefore represent the effect of receiving just hotspot policing and being greater than 500m from any unit receiving broken windows, relative to units more than 500 meters away from any unit receiving either treatment. This estimate should be similar to that from equation 2 above, but there are key differences:

- In the pairwise regression, we condition on segments having a non-zero probability of receiving the hotspots policing treatment and being >500m. In this example, we condition on not only hotspots policing assignment but *also broken windows assignment.* So while a segment receiving hotspots policing but being within 250m of a segment receiving the would help estimate the effect of receiving hotspot policing in the pairwise case, it would not be considered in the effect of hotspot policing in equation 3. Only segments receiving policing but being greater than 500m away from any segment receiving broken windows would determine the effect of hotspot policing in the interaction regression.
- In the pairwise regression, we restrict our sample to observations that have a non-zero probability of receiving the policing treatment and being greater than 500m from a unit receiving that treatment. Now we now restrict to units with a non-zero probability of:
  - Receiving the hotspots policing treatment and being >500m away from any units receiving the broken windows treatment
  - Receiving the broken windows treatment and being >500m away from any units receiving the hotspots policing treatment
  - Receiving both the hotspots policing and broken windows treatments
  - Being more than 500m away from any unit receiving either the hotspots policing intervention or the broken windows intervention.

[17] In this case we would recalculate our probabilities of treatment.

These differences will lead to power concerns as our sample size will decrease substantially when we estimate the interaction effect. Despite these power concerns, the parameter of interest is θ_{HB}, which represents the additional change in Y for units assigned to both treatments concurrently.

Another way we will test for an interaction effect is by using the timing of the intervention. The broken windows intervention began two months after the hotspot policing intervention, which was around the halfway mark of the policing intervention. This time lapse will allow us to compare crime reductions from hotspot policing before broken windows began to after using the pairwise regressions described in section 6.2. In addition, we phased-in the broken windows intervention in batches, allowing us to further use this time variation to evaluate the interactive effect. One benefit of this method is that it allows us to estimate the interaction effect without the power concerns discussed above. However, we still plan to estimate the interactive effect both ways, as our power calculations suggest we can detect an effect of about 0.12 SD on the interaction term in equation 3.[18]

6.4 Spillover on nonexperimental sample

While only two percent of street segments in Bogotá are eligible to be assigned to treatment, the others may experience spillover effects based off their proximity to treated hotspots. In order to estimate the spillover effects on non-experimental segments (segments under the 98th percentile of the aggregate crime index distribution), we plan to run pairwise regressions similar to those described in section 6.2.

6.5 Weights

As discussed above, we will weight each observation by the inverse of the probability of being observed in its assigned experimental condition. In equation 2, for example, we weight each hotspot assigned to pure control as the inverse of the probability of that hotspot being assigned to pure control. Units with weights of zero for certain conditions (e.g., a unit that can never be assigned to treatment) are excluded from estimation of the effect of that condition.[19] Therefore the more experimental conditions measured in a regression, the smaller the sample size that those estimates are based off.

To calculate these probabilities, we ran 10,000 simulations of the randomization procedure described above. Figure 2 displays the distribution of probabilities of hotspot assignment to each of the four conditions for hotspot segments while Figure 3 displays the distribution of probabilities for being treated in the broken windows treatment. Almost half of the hotspots have zero probability of being a pure control because they are less than 500m away from other hotspots.

6.6 Compliance

A central component of the hotspot policing intervention is the reallocation of time police spend in each of the four treatment arms. Although the treatment is meant to provide each hotspot with 90 minutes of policing per day, there is no guarantee that police will comply with the treatment status assigned to each segment.

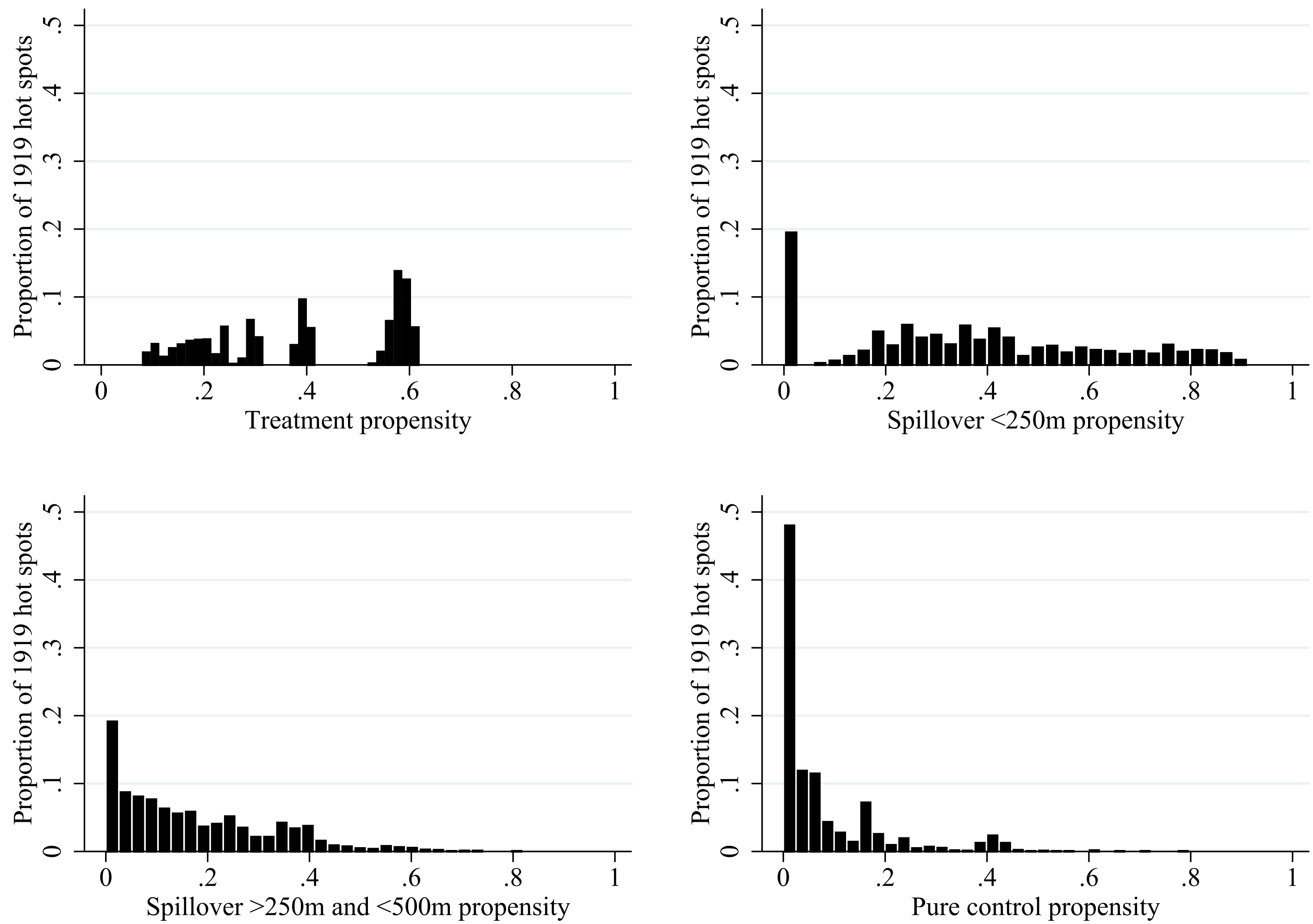
In order to measure compliance, we will use a PDC device that will send a signal with the exact location coordinates of the officers every 30 seconds. We construct a geo-fence of 40 meters around each hotspot,[20] and assign as patrolling time every time the signal falls into this geo-fence. Whenever the signal is within the geo-fences of two hotspots, the decision rule is to assign it to both geo-fences. By using this information, we will generate two reports every week that summarize the levels of compliance per shift, per hotspot, per quadrant, per police station, and overall for the police. While our preferred treatment effect is an ITT estimate, we will use this information as "first stage" information that will aid in the interpretation of our treatment effect on overall crime.

[18] Power calculations are discussed more in section 6.7.

[19] In practice we will exclude weights such that $p_{assignment} < \frac{1}{simulations}$, where simulations is the amount of randomization simulations run to estimate treatment probabilities. We will also top-code our weights at 20 so that observations are not given undue weight. We plan to conduct a sensitivity analysis with different values to ensure that this topcoding is not driving our results.

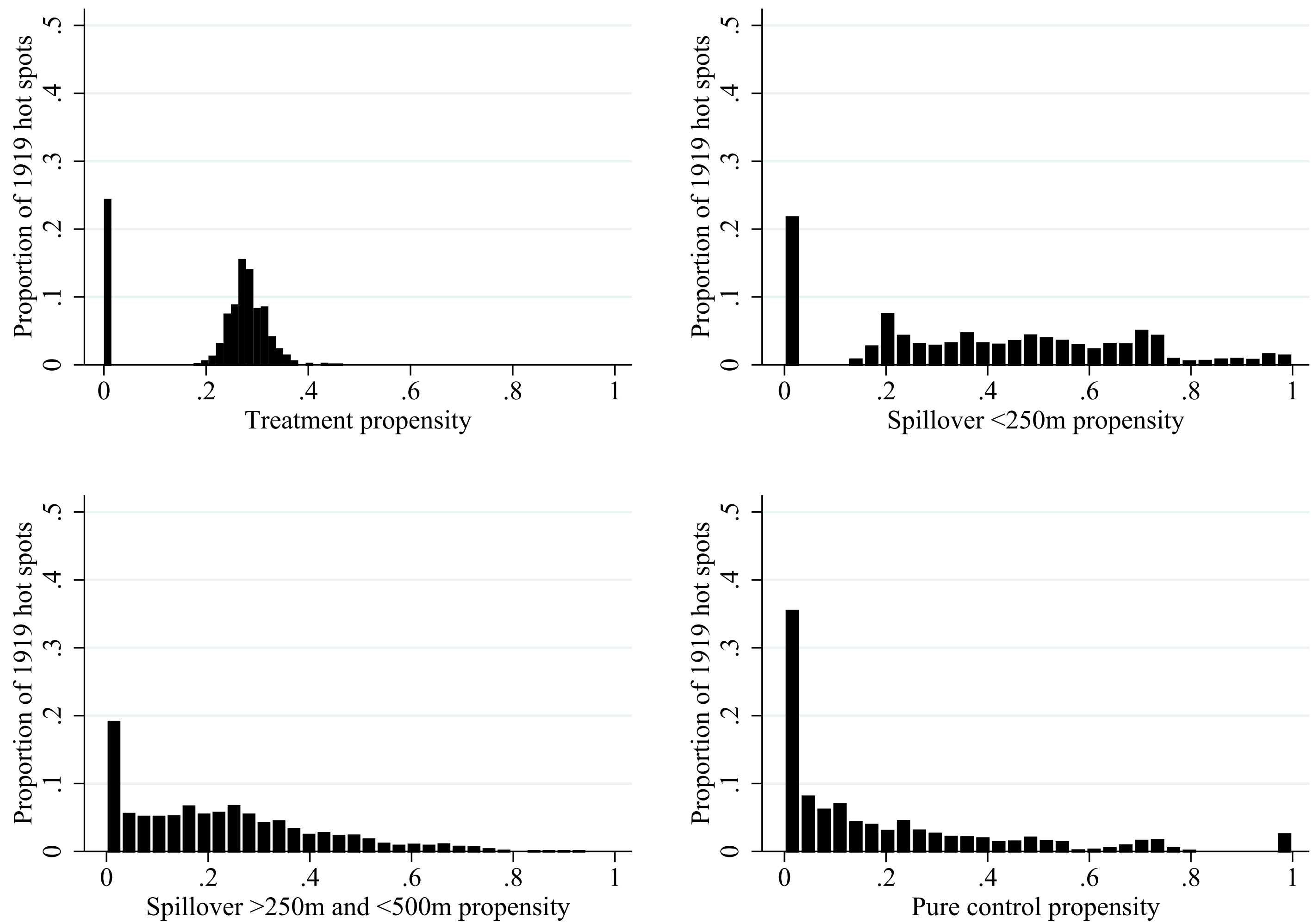
[20] Street segments are lines in geographic information systems. This geo-fence covers the resulting area of drawing circles with a radius of 40 meters at each point in the line.

Figure 2: Probability of treatment, hotspot policing



Notes: Treatment probabilities based off 10,000 simulations of the randomization procedure described in section 2.3.

Figure 3: Probability of treatment, broken windows



Notes: Treatment probabilities based off 10,000 simulations of the randomization procedure described in section 2.3.

Table 5: Minimum detectable effects

	Minimum detectable effect (SD)		
	Hotspot policing	Broken windows	Interaction
	(1)	(2)	(3)
Treatment	0.097	0.083	0.124
<250m spillover	0.131	0.112	
250m-500m spillover	0.122	0.098	

Notes: This table displays the minimum detectable effects for our design based of 1,000 simulations. We estimate each pairwise regression (or interaction regression for column (3)), take the empirical standard deviation of all 1,000 estimated β's and transform them into an MDE in standard deviations.

Compliance will be easier to measure in the broken windows intervention: we will measure the total number of days the municipal team visited the segment over the course of the study.

Because the amount of time that police spent at the assigned hotspot is encouraged rather than determined, we plan to estimate the average effect of actual surveillance time using instrumental variables regression. The outcome is regressed on surveillance crime at each hotspot, which is in turn instrumented using the random assignment to treatment.

Similarly, visits by the municipal cleanup team are encouraged rather than determined. Our instrumental variables regression estimates the average effect of each visit by the cleanup team by regressing outcomes on visits, which is in turn instrumented using the random assignment to cleanup treatment.

In order to estimate the average effect of the interaction between actual police surveillance time and the actual number of cleanups, we using IV regression to regress outcomes on actual surveillance time, actual number of cleanups, and the product of the two, using assigned surveillance, assigned cleanup, and the product of the two assignments as instrumental variables.

6.7 Power calculations

We ran 1,000 simulations to try to estimate the minimum detectable effects (MDEs) associated with our design. For each T = {treatment, short-range spillover, long-range spillover}, we ran the following pairwise regression:

$$crime_{15} = \beta_0 + \beta_1 * T + C_1 * S + C_2 * X \tag{5}$$

where S is block (police station) fixed effects and X includes all the covariates described in section 3.3. We weighted each observation by the inverse of its probability of being assigned to their respective treatment assignment. To make sure we were comparing like units we restricted our sample to observations where the probability of being assigned to pure control and the probability of being assigned to treatment T were both greater than 0, and to observations where control = 1 or $T = 1$. We then took the standard deviation of all 1000 of our $\beta_1's$ and transformed them into an MDE in standard deviations.

Table 5 displays the MDE's associated with our design. For hotspot segments with two spillovers and all covariates, we are powered to detect effects of about 0.10 SD on treatment and 0.13 SD on both spillovers. For segments receiving the broken windows treatment, these numbers are 0.08 SD and 0.10 SD, respectively. We have the power to detect an effect of about 0.12 SD on the interaction between treatments.

6.8 Omissions

Omissions will be adjudicated using the default Standard Operating Procedures found here: `https://github.com/acoppock/Green-Lab-SOP`.

References

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

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.

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

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

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.

Rubin, D. B. (1990). Formal Modes of Statistical Inference for Causal Effects. *Journal of Statistical Planning and Inference 25*(3), 279–292.

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.

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

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

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

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