IMPLEMENTATION AS INTERVENTION: CAN IMPROVING BASIC MANAGEMENT PRACTICES STRENGTHEN POLICING IN CHICAGO?

This draft: September 26, 2019
Latest draft: Available here

Max Kapustin, Terrence D. Neumann, Kimberley Smith, Alexander Heaton and Jens Ludwig

Preliminary draft. Please do not cite or distribute.

The University of Chicago Crime Lab is a privately-funded, independent, non-partisan academic research center founded in 2008 to help cities identify the most effective and humane ways to reduce crime and violence, and the harms associated with the administration of criminal justice. We are grateful to the Chicago Police Department for making available the data upon which much of this analysis is based. We thank AbbVie, the Joyce Foundation, the John D. and Catherine T. Macarthur Foundation, the McCormick Foundation, and the Pritzker Foundation for their support of the University of Chicago Crime Lab and Urban Labs, as well as Susan and Tom Dunn, Ken Griffin, and Ira Handler. We thank Rowan Gledhill, Riddhima Mishra, Michael Ridgway, and Noah Sebek for assistance with the data analysis, and Nikolay Doudchenko, Michael Robbins, and Kaspar Wuthrich for sharing helpful code with us. For useful comments we thank seminar participants at the University of Pennsylvania, NBER 2018 Summer Institute, NYU Wagner, and APPAM. Finally, and most important, we thank Roseanna Ander, Sean Malinowski, Marjolijn Bruggeling-Joyce, Anthony Berglund, Heather Bland, Trayvon Braxton, Amanda Dion, Mariah Farbo, Noe Flores, Jaureese Gaines, Brendan Hall, David Leitson, Kevin Magnan, Emma Marsano, Jacob Miller, Ashley Orosz, Paulina Pogorzelski, Daniel Rosenbaum, Zoe Russek, Thomas Scholten, Lauren Speigel, Diamond Thompson, Michael Thompson, Matthew Triano, and Yida Wang for helpful conversations and comments, and for all of their work implementing the SDSCs themselves. All opinions and any errors are our own and do not necessarily reflect those of our funders or of any government agencies.

ABSTRACT

Numerous studies find that increasing police resources deters crime without increasing arrests and incarceration. But cities struggling to control crime often face significant budget constraints and are unable to substantially increase police resources. Unfortunately, the field's knowledge of how to get more out of existing police spending is limited. In this paper, we study a series of management changes undertaken by the Chicago Police Department (CPD) to improve how it follows the proactive policing "playbook" that many cities around the country have adopted, which involves focusing on the places and people at highest risk of serious crime. These district-level management changes, known as Strategic Decision Support Centers (SDSCs), do not substantially change existing CPD strategies or the resources available to implement them, but instead strengthen the quality of their implementation. We measure the SDSCs' impact using a synthetic controls design, creating for each treated police district a comparison district resembling it. Because treated districts are outliers in their levels of violence, we encounter challenges applying existing methods: some are overly restrictive, while others are flexible but may identify inappropriate comparison units, compromising the reliability of their estimates. We modify existing methods in ways that guard against overfitting, such as by fitting a synthetic control to multiple outcomes simultaneously and altering an estimator to prioritize out-of-sample prediction accuracy. We also propose modifications to existing inference procedures to handle cases where there are few comparison units. Our findings suggest that the SDSC in the 7th police district, historically one of the most violent in Chicago, decreased shootings there by 30 percent and Part I violent felonies by 12 percent, findings that are statistically significant even after accounting for multiple testing and the number of different districts implementing these reforms. These results suggest the promise of interventions focused on management reforms to extract greater performance from police departments and other municipal agencies.

Max Kapustin
University of Chicago Crime Lab
33 North Lasalle, 16th Floor
Chicago, IL 60602
kapustin@uchicago.edu

Terrence D. Neumann
University of Chicago Crime Lab
33 North Lasalle, 16th Floor
Chicago, IL 60602
tdneumann@uchicago.edu

Kimberley Smith
University of Chicago Crime Lab
33 North Lasalle, 16th Floor
Chicago, IL 60602
kimberleys@uchicago.edu

Alexander Heaton
University of Chicago Crime Lab
33 North Lasalle, 16th Floor
Chicago, IL 60602
asheaton@uchicago.edu

Jens Ludwig
Harris School of Public Policy
University of Chicago
1307 East 60th Street
Chicago, IL 60637
and NBER
jludwig@uchicago.edu

I. INTRODUCTION

Spending on policing in the U.S. exceeds $100 billion annually.[1] Research over the last two decades has consistently shown that additional police resources are effective at reducing crime.[2] Accumulating evidence that policing can *prevent* crime without increasing incarceration makes it an especially appealing form of crime control (see, e.g., Durlauf and Nagin 2011).[3] However, large pension debts and other fiscal constraints, in addition to fraught police-community relations, make it difficult for many cities to substantially increase spending on policing.

Yet little is currently known about how to get more "output" from existing spending on policing. Efforts to improve police efficiency tend to focus on "best practice" tactics and strategies that have already been adopted by most big-city police departments in the U.S. (National Academies of Sciences, Engineering, and Medicine 2018). As Bratton and Murad (2018) note, these tactics and strategies, broadly falling under the rubric of "proactive policing" or "precision policing," are to focus police resources on the places, times, and people where, and involving whom, violence is most likely to occur, and to do so in cooperation with the community to strengthen police legitimacy. While some previous studies have examined what happens during police "slowdowns," what remains poorly understood is the degree to which it is possible for a city to intentionally *improve* how policing is carried out, and what the result is for crime.[4]

This paper examines a series of changes to basic management practices meant to bolster proactive policing in the Chicago Police Department (CPD), the second largest in the U.S. Chicago is a particularly important setting for this research: over a one year period from 2015 to 2016, the city's homicide rate per capita increased by fully 60 percent. The combination of the size of this increase and the size of the city of Chicago itself means this local spike single-handedly accounted for 20 percent of the nationwide increase in homicides over that time. Following this surge in violence, and in partnership with Chief Sean Malinowski of the Los Angeles Police Department and our research center (the University of Chicago Crime Lab), CPD implemented a package of management changes that became known as the Strategic Decision Support Centers (SDSCs).

---

[1] https://www.urban.org/policy-centers/cross-center-initiatives/state-and-local-finance-initiative/state-and-local-backgrounders/police-and-corrections-expenditures#Question1Police

[2] See for example Levitt (1997, 2002); DiTella and Schargrodsky (2004); Klick and Tabarrok (2005); Evans and Owens (2007); Draca, Machin, and Witt (2011); Machin and Marie (2011); Owens (2013); Chalfin and McCrary (2018); Mello (2019).

[3] An alternative way to prevent crime is through social policy. But as Levitt (1997) and others has noted, beyond the difficulty of finding promising social programs, there are challenges with implementing them successfully at scale and targeting them toward those population sub-groups that would benefit most from them. Discussion about ways of trying to overcome some of these scale-up challenges are in Al-Ubaydli et al. (2017), Banerjee et al. (2017), Davis et al. (2017), and Muralidharan and Niehaus (2017).

[4] Some studies have shown that police slowdowns—temporary reductions in police activity following public outcry about an incident of perceived police misconduct—may result in increased crime (Shi 2009; Heaton 2010). Evidence from a weeks-long reduction in proactive police activity by the New York Police Department between 2014 and 2015 suggests the opposite (Sullivan and O'Keeffe 2017).

The management changes introduced by the SDSCs are explicitly designed to improve how CPD implements two pillars of the proactive policing strategy: focusing on the places and the people at highest risk of serious crime.[5] Two management changes in particular may help achieve those objectives. First, the SDSCs created the role of a civilian crime analyst tasked with analyzing data and developing insights about local criminal activity to inform resource deployments. Second, the SDSCs established a formal, daily process to incorporate data and analysis into decision-making. Prior to the SDSCs, much of the data collected by the Department went underutilized, and deployments were made in an ad hoc manner relying on information transmitted by word-of-mouth. These two changes are intended to remedy these shortcomings and consistently provide district Commanders with actionable, detailed information when making decisions. Though these changes may be less salient than some of the technology improvements that also accompanied the SDSCs, they represent a significant shift from the status quo at CPD.[6]

Measuring the impact of these management changes is not straightforward. In February 2017, the first SDSCs opened in the two police districts with the highest violent-crime rates—one on the South Side (CPD's 7th district, serving the Englewood and West Englewood neighborhoods) and one on the West Side (CPD's 11th district, serving the Humboldt Park and Garfield Park neighborhoods)—before expanding to four additional districts in March 2017. A randomized controlled trial was not an option given the pressing public safety crisis facing Chicago and the need to prioritize the highest-violence districts for reform. And with only 22 police districts in Chicago, a regression discontinuity design lacks adequate statistical power.

We instead use the synthetic controls method to measure the effects of the SDSCs. Unlike a standard difference-in-differences estimator, a synthetic control uses pre-treatment data to assign individual weights to each control group unit with the goal of approximating the pre-treatment patterns of the treated unit. However, applying the synthetic controls method to our setting is made complicated by the fact that the SDSC districts are significant outliers. The most commonly used synthetic controls estimator, introduced by Abadie, Diamond, and Hainmueller (2010; ADH hereafter), imposes rigid restrictions on the weights that so-called donor units can receive to limit extrapolation beyond the support of the data. A newer estimator introduced by Doudchenko and Imbens (2017; DI hereafter) replaces these rigid constraints with a more flexible, data-driven approach. But by focusing exclusively on matching pre-treatment outcome patterns, this new estimator may increase the likelihood of assigning weight to donor units on the basis of statistical noise rather than a shared relationship with the treated unit. This type of overfitting could undermine the ability of the resulting synthetic control to recover the counterfactual for the treated unit and lead to misestimation of the treatment effect.

We explore the issue of overfitting in synthetic controls and modify existing methods to guard against it. Two modifications concern the role of covariates. Applications of the ADH estimator often rely on the full pre-treatment series of outcome observations to estimate weights;

---

[5] The third pillar, community engagement, is not directly a focus of the SDSC intervention. However, as we describe later, certain district Commanders prioritized this as part of the overall reform introduced by the SDSC.

[6] In addition to management changes, the SDSCs included the adoption of gunshot detection sensors and the expansion of cameras providing high quality, real-time coverage of key locations. Both of these added to an already robust data collection effort at CPD. We discuss these changes in greater detail below.

as Kaul et al. (2018) demonstrate, this renders other covariates used in the estimation irrelevant. The DI estimator leaves out covariates altogether. First, we propose using covariates to limit the pool of donor units to those with characteristics similar to those of the treated unit before assigning any donor weights. Second, we propose a way to fit a synthetic control to multiple outcomes simultaneously, such as a primary outcome and one or more covariates or secondary outcomes. Doing this has two advantages: the likelihood of overfitting is reduced when a synthetic control is fit to multiple outcomes, and it yields a single synthetic control for understanding the role of mediating mechanisms in the observed treatment effect on the primary outcome. The last modification is to the DI estimator itself, aiming to make it more robust to overfitting by prioritizing out-of-sample prediction accuracy. Unlike the original estimator, the version we propose uses a time series cross-validation technique to choose parameters governing weight selection that minimize prediction error in one-step-ahead forecasts.

In addition to modifications to address overfitting, we also introduce a modification to the standard synthetic controls inference method for settings with few comparison areas. The standard approach to inference with synthetic controls is a placebo test: a treatment effect is estimated for each comparison area, the observed effect in the treated unit is compared to the distribution of placebo treatment effects, and the share of placebo effects greater than the observed one is the *p*-value. However, when there are few comparison areas—16 police districts not receiving an SDSC, in our setting—the number of potential *p*-values is small, limiting our ability to understand whether the observed effect is an outlier or not. We address this problem using a resampling method similar in spirit to one used by Robbins, Saunders, and Kilmer (2017): we increase the number of comparison areas by resampling with replacement the more numerous small geographic areas (police beats) within the more limited large geographic areas available (police districts).[7]

We find that the 7th district, one of the first to receive an SDSC, experienced a 30 percent reduction in shooting victims and a 12 percent reduction in Part I violent felony incidents after the SDSC opened. Stated differently, over the 11 post-intervention months of 2017, these results imply that the 7th district had 94 fewer shooting victims as a result of the SDSC. These reductions remain statistically significant even after correcting for multiple comparisons, and remain qualitatively similar when the synthetic control is fit to multiple outcomes simultaneously, which enables us to examine the potential mechanisms of action behind this decrease in serious crime. Accompanying these reductions in the 7th district relative to its synthetic control are increases in gun arrests, warrant arrests, overall arrests, and traffic stops. However, most of these increases are estimated imprecisely or predate the introduction of the SDSC; the closest exception, increased warrant arrests, occurred after the SDSC's introduction and is precisely estimated when considered separately but not after adjusting for multiple comparisons. We find no evidence of increases in officers working in the district, but suggestive evidence of increased positive community interactions recorded by officers.

[7] Robbins, Saunders, and Kilmer (2017) study a neighborhood-level intervention and have data on thousands of sub-neighborhood (e.g., block-level) comparison areas. They use a permutation technique to generate groups of comparison blocks for their placebo test. Because these placebo areas are random assortments of blocks, in contrast to the actual treatment area, they adjust their treatment effect estimates to guard against the resulting bias. In contrast, our permutation technique generates placebo areas that preserve the structure of the real comparison districts by sampling police beats within them with replacement.

In the 11th district, the other to first receive an SDSC, shooting victims declined by 13 percent and Part I violent felonies declined by 10 percent. However, while the reductions in the 11th district are statistically significant when considered separately, they do not remain so after correcting for multiple comparisons. Results from the remaining SDSC districts are more mixed: several show positive treatment effect estimates, implying the intervention increased crime. However, these estimates are either imprecisely estimated or not credible due to the poor fit of the synthetic control, particularly in the last pre-treatment year of 2016. Finally, we detect increases in the numbers of arrests initiated by officers monitoring police video cameras after the introduction of the SDSCs.

Taken together, these results are consistent with the hypothesis that the SDSCs—which represent a very small share of CPD's budget—reduce crime by changing how existing police resources are used, namely through the increased use of technology and data to guide arrests, and potentially also through more community engagement and a greater focusing of resources on the places where crime is most likely to occur. However, the uneven results in other districts suggest that, as a reform intended to improve the implementation of an existing policing strategy, the SDSCs may themselves be subject to imperfect implementation.

The remainder of the paper is organized as follows. Section II provides basic background on crime and the criminal justice in Chicago, focusing on very recent trends that serve as a backdrop for the management changes we examine. Section III reviews the available research basis for the current conventional wisdom regarding policing best practices. Section IV describes how CPD was implementing these practices prior to adoption of the SDSCs and how their introduction changed those practices. Section V discusses the intuition behind the synthetic controls method, pitfalls in applying it, and how we navigate those pitfalls in our setting. Section VI discusses our estimates of the SDSCs' impact on crime, including what we can say about specific management changes and policing practices that may be responsible for any observed declines. Section VII concludes.

II. BACKGROUND ON CRIME AND CRIMINAL JUSTICE IN CHICAGO

Chicago is a city of 2.7 million people, divided roughly evenly between non-Hispanic whites (32.7 percent of total residents), non-Hispanic blacks (30.1 percent) and Hispanics (29.0 percent).[8] Like other American cities, Chicago remains very racially segregated, as shown in Figure 1: the African-American population is disproportionately concentrated on the city's South and West Sides, the Hispanic population in a handful of neighborhoods southwest and northwest of the downtown central business district, and the white population in neighborhoods on the north side and in the surrounding suburbs. Segregation by income is closely related to segregation by race, as shown in Figure 2.

Like other American cities, Chicago also struggles with a gun violence problem that is concentrated in the city's most disadvantaged neighborhoods (Figure 3). For most of the last decade, Chicago has experienced more murders than any other city in the country. However, this is partly a function of Chicago's size; among cities with a population in excess of 250,000,

[8] https://factfinder.census.gov/faces/nav/jsf/pages/community_facts.xhtml

Chicago's homicide rate per capita has ranked anywhere between the 9th and 20th highest since 2010, and has consistently been below cities like Baltimore, Detroit, New Orleans, Newark, and St. Louis. Among large American cities, most of the variation across in homicide rates is accounted for by gun homicides, while the rates for non-gun homicide are much more similar (Figure 4).

Chicago has long had the reputation of being a large city with a high rate of serious violence, but the long-term picture is more nuanced and revealing. Figure 5 shows the homicide rates per 100,000 residents in Chicago and two of its peer cities, New York City and Los Angeles, over the last 130 years. A striking feature of this graph is how generally similar the homicide rates for all three cities have been for most of this period; as recently as the early 1990s, they were virtually indistinguishable. But equally striking are two notable periods of divergence, during which Chicago experienced much higher homicide rates than its peers: the 1920s during the Prohibition era, and the last 25 years.

The gulf between Chicago and its peer cities reached its widest point in 2016, when Chicago experienced 765 homicides, an increase of almost 60 percent from 2015. This spike in homicides in Chicago single-handedly accounted for 20 percent of the nationwide increase in homicides in 2016. Annual increases of this enormous magnitude are exceedingly rare for a city of Chicago's size. This unprecedented surge in gun violence was the impetus for the SDSC intervention.

Why gun violence increased so substantially from 2015 to 2016 remains unclear; most existing explanations are, in our view, either wrong or incomplete.[9] One common hypothesis points to social conditions like poverty that are believed to be key "root causes" of crime and violence.[10] Such root causes are surely important for explaining variation across areas in crime rates at a point in time, but they are less helpful in understanding what happened in 2016 given how *sudden* the increase was in Chicago's gun violence. Even with the widely-reported budget problems facing the city of Chicago, Cook County, and the state of Illinois, key root causes generally change slowly over time and, as best we can tell in available data, they did not change rapidly at the end of 2015.[11] Moreover, the 2016 crime increase was mostly confined to gun crimes: shootings rose by 43 percent and gun robberies rose by 26 percent, while non-gun violent crimes only rose by 10 percent and property crime incidents by 6 percent. We would expect important changes in major social conditions to have a more broad-based impact across a range of crime types.

[9] Much of the data and arguments we present here about the change in crime in Chicago from 2015 to 2016 are taken from Kapustin et al. (2017).

[10] See, for example, NPR's interview with a number of leading advocates and researchers in Chicago: https://www.npr.org/2016/09/03/492549546/examining-the-reasons-for-chicagos-violence

[11] One variant of these "root cause" arguments is that government funding for violence mediation and street outreach programming was cut in March 2015. Though gun violence increased following those cuts, those increases are of a magnitude typically seen in Chicago as the weather warms. Adjusting for seasonality, the significant increases in Chicago's gun violence began at the start of 2016.

Other explanations focus on changes in the policing environment in Chicago. For example, Cassell and Fowles (2018) point to a decline in street stops following the consent decree entered into by the American Civil Liberties Union (ACLU) and CPD as the cause of the gun violence spike. Because they focus on changes that occurred proximate to the gun violence increase and that affected one of the key "inputs" to crime control—policing—this explanation is more plausible than those focused on slowly-evolving root causes. However, because there were so many changes occurring within such a short period of time in late 2015 that affected CPD, there is no way to determine how much any single change contributed to the subsequent violence increase.[12]

Since the decline in Chicago's street stops happened citywide, it is hard to isolate its causal effect on crime. But evidence from other cities suggests the relationship between street stops and violent crime may be weaker than commonly hypothesized. For example, the rate of street stops in New York City declined by over 95 percent from its peak in 2011 in response to legal challenges, with no attendant increase in the city's homicide rate (Figure 6). Conversely, Milwaukee is another Midwestern city with a homicide trend that closely follows that of Chicago, including a 60 percent homicide rate increase that pre-dated Chicago's by one year (Figure 7, Panel A). Yet Milwaukee experienced a much more modest decline in its rate of street stops than did Chicago (Figure 7, Panel B).

III. PROACTIVE POLICING: HIGH-RISK PLACES AND PEOPLE

For much of policing's history in the U.S., the focus was on "reactive policing": responding quickly to the scene and successfully investigating the case following a call for service. But starting in the 1960s, departments across the country began shifting their focus to "proactive policing," which a recent National Academy of Sciences committee on the topic defines as "all policing strategies that have as one of their goals the prevention or reduction of crime and disorder and that are not reactive in terms of focusing primarily on uncovering ongoing crime or on investigating or responding to crimes once they have occurred" (National Academies of Sciences, Engineering, and Medicine 2018, p. 1). The standard playbook for modern-day proactive policing is to concentrate police resources on high-risk places and people, in addition to improving relations with the community and emphasizing police problem-solving (National Academies of Sciences, Engineering, and Medicine 2018, p. 2).

Like most other police departments in the U.S., CPD has for years adopted a proactive policing strategy. However, as we discuss later, the Department's implementation of that strategy left room for improvement in how it carried out two of its core elements: concentrating police resources on high-risk places and people. The SDSCs are designed to provide those improvements.

[12] On November 24, 2015, the dash cam video of the shooting of Laquan McDonald by CPD officer Jason Van Dyke was released; on December 1, 2015, CPD Superintendent Garry McCarthy was fired and Mayor Rahm Emanuel created the Chicago Police Accountability Task Force; on December 7, 2015, the U.S. Department of Justice Civil Rights Division opened a civil pattern or practice investigation into CPD; on December 9, 2015, Mayor Rahm Emanuel delivered a speech criticizing decades of police corruption; on January 1, 2016, the consent decree between the ACLU and CPD went into effect, requiring an independent evaluation of CPD practices and procedures, and introducing a new Investigatory Stop Report (ISR) that officers were required to complete after every street stop; also on January 1, 2016, Illinois SB1304 went into effect, also requiring police to record information about each investigative street stop and to issue a receipt after each such stop.

In what follows, we discuss the basis for concentrating police resources on high-risk places and people, drawing on evidence from the larger research literature.

High-risk places

Crime, particularly violent crime, is very geographically concentrated. As a result, it is often suggested that crime prevention resources can be used most effectively by deploying them to a handful of "high crime" areas. However, it is important to note that, while crime may be concentrated *ex post*, what is relevant for crime prevention is the degree to which the risk of crime is concentrated *ex ante*. For example, suppose that residential burglaries within a city occur at just 2 percent of addresses. If the annual burglary rate in the city is 2 per 100 households, then this concentration of burglaries is not surprising; in fact, burglaries could have occurred, at most, at just 2 percent of addresses within a year. In contrast to this *ex post* perspective, which notes the concentration of crimes after they occurred, the relevant perspective for crime prevention is an *ex ante* one: predict the likelihood that a given address or block will experience a burglary, and prioritize resources to those at highest risk.

One of the most common place-based strategies used by police departments is to concentrate extra police resources on places with elevated risk of violence, sometimes called "hot spot" policing. Departments vary with respect to whether they do this targeting using an explicit statistical prediction model of where crime will occur (e.g., using commercial software like PredPol or HunchLab), or an implicit prediction model of where crime will occur (e.g., using the locations of similar crimes in the recent past). Several randomized controlled trials (RCTs) within criminology suggest that increased police presence leads to fewer crimes in the locations that are targeted (see, e.g., Braga, Papachristos, and Hureau 2012 for a review).

What has been more difficult to determine is the degree to which hot spot policing actually prevents crime overall, rather than just displacing it to areas with a relatively smaller police presence. Underlying the displacement issue is in part the question of whether the features of a given place can make it unusually criminogenic. While some places surely have such features—such as a street corner used to sell drugs on Chicago's west side that is adjacent to the access road running alongside I-290, the interstate known as the "heroin highway"—we have limited conceptual guidance, and no reliable empirical evidence, about which factors make some places criminogenic.[13]

The hot spot policing RCTs that have been carried out typically do not find evidence of displacement. But they also typically have lower statistical power to detect displacement than to

[13] A literature within criminology often labeled under the umbrella of "opportunity theory" suggests the importance of some places in bringing together motivated offenders with suitable targets in situations where there is limited guardianship (see, e.g., National Academies of Sciences, Engineering, and Medicine 2018, p. 46).

detect impacts in the hot spots themselves.[14] One of the best recent studies from a methodological perspective is by Blattman et al. (2019), who find evidence of displacement for property crimes, though their experiment is carried out in Bogotá, Colombia, so the applicability of this finding to hot spot policing in the U.S. is unclear.

A related place-based strategy uses technology, such as closed-circuit television (CCTV) cameras, to help decide where to concentrate police resources in something closer to real-time, as opposed to using statistical models that rely on data recorded with some lag. Cameras can also have other effects on crime as well, such as through potential deterrence and assistance with investigations. The available research suggests the effect of CCTV cameras is a "modest but significant desirable effect on crime" (Welsh and Farrington 2008, p. 2), with impacts particularly concentrated on motor vehicle thefts, though efficacy may also depend on the other crime-control strategies that are in place. As we discuss below, the use of CCTV cameras is increasingly widespread, including in Chicago.

High-risk people

Criminologists have known since the birth cohort studies of Marvin Wolfgang in Philadelphia that the *ex post* (realized) risk of violence involvement is very concentrated within the population (Wolfgang et al. 1972). He found, for example, that about half of all offenses (arrests) were accounted for by 18 percent of the birth cohort. Similar results have subsequently been replicated in a number of other settings. However, as with the burglary example earlier, this is an *ex post* calculation of concentration; it is distinct from, and arguably less useful for crime prevention than, being able to determine which specific youth are at elevated *ex ante* risk for engaging in future violence.

An example of how police departments operationalize this insight is the practice of stepped up enforcement and surveillance of individuals on probation or parole. Because these individuals are often considered to be at higher risk of committing future crimes, and the terms of their release require their compliance with certain conditions, they may receive greater attention from the police. For example, if suspected gang members are on probation and must observe a curfew or refrain from having drugs and guns at home, the police may use this information to monitor them more closely, with the goal of deterring them from engaging in crime.

A different approach to identifying high-risk people is to develop an explicit statistical prediction model of who will engage in crime, the person-based analog to commercially available place-based prediction tools. The Crime and Victimization Risk Model (CVRM), formerly known as the Strategic Subjects List (SSL), is an example of such a model that was developed by Professor Miles Wernick of the Illinois Institute of Technology for CPD. The CVRM predicts an individual's risk of arrest or victimization for a future shooting or a homicide, relying on predictors such as a

[14] We thank Justin McCrary for this observation. As a way to see the issue, imagine we have one street block that is identified as a hot spot surrounded by ten adjacent street blocks. If our prior is that crime will be displaced uniformly into the ten adjacent blocks when extra police resources are deployed to the hot spot, then we would need to be able to detect an increase in crime rates in these adjacent areas that is one-tenth as large as the reduction in crime in the hot spot. While the sample size for detecting that displacement effect is ten times as large as for detecting the impact in the hot spot itself, the statistical power for that displacement impact increases only in the square root of the number of displacement areas.

person's age at the time of their most recent arrest, incidents in which they were the victim of a shooting, and incidents in which they were arrested for illegal possession of a gun.[15]

The CVRM, and previously the SSL, is used by CPD to help identify people for its "custom notification" program and "call-in" sessions. During a custom notification, an officer and a community group representative refer a person to social services and warn them about the dangers of continuing to engage in risky behavior.[16] Similarly, call-in sessions involve delivering a message of focused deterrence—heightened penalties for continuing to engage in violence—to a group of individuals, alongside offers of social services.[17] A quasi-experimental study of an earlier version of CPD's call-ins, implemented as part of the local U.S. Attorney's Project Safe Neighborhoods initiative, suggests they may reduce crime in Chicago (Papachristos, Fagan, and Meares 2007). However, it is unclear whether the research design is capable of isolating the effects of the intervention itself due to the difficulty of finding a comparable comparison group, a point we return to below. A different quasi-experimental evaluation of CPD's use of an earlier version of the SSL finds that being identified by the model is not associated with a change in the likelihood of being a homicide or shooting victim, though subjects are more likely to be arrested for a shooting (Saunders, Hunt, and Hollywood 2016). Here, too, it is difficult to know whether the design is able to credibly isolate the effect of being identified as high-risk by the SSL (e.g., being above some threshold). In assessing the effectiveness of any person-based approach to crime prevention, the results of an evaluation may say as much about the intervention being delivered as they do about the method of identifying the people to receive it.

More work is necessary to understand the potential and limitations of using statistical models that predict individuals' risk to target crime-prevention efforts. For example, one study currently underway in Chicago uses, in part, a statistical model developed with data from CPD to identify men at the highest risk for gun violence involvement and refer them to an intensive social service intervention designed to reduce this risk (Bertrand et al., forthcoming). Among the issues that researchers and practitioners must contend with are the limitations of the data used to predict risk. For example, in Chicago, the arrest clearance rate for a serious offense like homicide is below the national average, and the arrest clearance rate for non-fatal shootings is in the single digits. This makes it challenging to measure, and therefore predict, alleged perpetration of serious violence. Finally, and relatedly, the risk of offending is likely to be more concentrated than the risk of victimization due to the small number of people willing to become high-volume shooters, based on anecdotal reports from officers at CPD. In contrast, a shooter may be willing to target anyone in an opposing gang and could also end up injuring innocent bystanders in the process, resulting in a diffusion of victimization risk relative to offending risk.

IV. STRENGTHENING PROACTIVE POLICING IN CHICAGO THROUGH BETTER MANAGEMENT

---

[15] https://home.chicagopolice.org/wp-content/uploads/2019/01/FACT-SHEET-Crime-and-Victimization-Risk-Model-1.pdf

[16] http://directives.chicagopolice.org/directives/data/a7a57bf0-1456faf9-bfa14-570a-a2deebf33c56ae59.html

[17] http://directives.chicagopolice.org/directives/data/a7a57bf0-136d1d31-16513-6d1d-382b311ddf65fd3a.pdf

Implementing a proactive policing strategy that concentrates resources on places and people at high *ex ante* risk requires three components: data, analytical capacity, and a process for incorporating analysis into decision-making. In this section, we describe how the SDSCs improved CPD's implementation of a proactive policing strategy by making improvements in these three areas. Because the Department had invested significantly in its data infrastructure for many years, the contributions made by the SDSCs in this area were comparatively more minor, despite their most outwardly visible aspect being a dedicated room within a police station outfitted with computers and large monitors. Instead, where we believe the SDSCs made the largest improvements is to CPD's *management process*: how data are analyzed, by whom, and for what purpose. Though less salient than the technology investments that accompanied the SDSCs, these improvements may represent a more substantial change to the status quo at CPD.

In this section, we first provide necessary background on how the patrol function at CPD is structured, since it is patrol that is ultimately tasked with executing any proactive policing strategy. Then, we discuss each area the SDSCs seek to improve, including their status quo prior to the implementation of the SDSCs and how the SDSCs changed CPD's practices. Finally, we describe how the SDSCs were rolled out across the Department's highest-violence police districts.

Structure of patrol at the Chicago Police Department

Before describing how the SDSCs changed practices at CPD, it is helpful to first understand how the Department's patrol function is structured. CPD has approximately 13,000 officers, over half of whom are within the Bureau of Patrol. To manage the task of patrolling Chicago's roughly 230 square miles, the city is divided into 22 police districts (analogous to precincts in other cities) that each fall within one of three areas: North, Central, and South (Figure 8).[18]

Each police district is overseen by a Commander who decides how to task the patrol resources at his or her disposal. These resources include beat officers, discretionary officers, tactical officers, and intelligence officers. Beat officers patrol the smallest unit of geography—a beat, usually less than 1/10th the area of the district—and respond to calls for service. Discretionary officers are similar to beat officers, except they can be deployed anywhere in the district; Commanders usually have between 3 and 9 discretionary officers to work with per shift. Tactical officers typically do not respond to calls for service except those that are highest priority (e.g., shots fired), wear plainclothes, and focus primarily on gangs, executing search warrants, and making gun and other high-profile arrests. Finally, intelligence officers gather information from local sources to provide the Commander a more complete picture of crime in the district; they wear plainclothes and typically do not make arrests.

In addition to units operating within a district that are overseen by the Commander, each Area is assigned its own units that can be deployed across the districts in that Area. Overseen by the Area's Deputy Chief, these include saturation teams, gun teams, and gang enforcement teams. Saturation teams are deployed as-needed to parts of the Area experiencing heightened activity,

[18] In addition to encompassing the city's north side, Area North includes downtown Chicago and much of what is considered the city's West Side. Area Central includes part of the near south side, and Area South covers the remainder of the city's South Side.

acting as a rapid-response force. Gun teams are similar to districts' tactical units and focus on gun recoveries and arrests. Finally, gang teams are deployed to areas of gang activity.

Due to their varying size and the nature of crime that occurs there, the 22 districts function like 22 different small towns and cities, with Commanders as their police chiefs. A smaller district like the 15th, covering the West Side neighborhood of Austin, is under four square miles and contains 60,000 residents. A larger district like the 8th, covering several South Side neighborhoods including Chicago Lawn, is roughly 24 square miles and contains almost 250,000 residents (about the population of Buffalo, New York). Some districts experience little violent crime, while others have homicide rates rivaling some of the most dangerous cities on earth. Even among the districts experiencing serious violent crime, its nature can vary significantly: gun violence in West Side districts is reputed to be tied to the narcotics trade operating there, while in South Side districts it is thought to be driven more by interpersonal disputes. As a result, each Commander must develop his or her own strategy to address the concerns in their own district, using the information and patrol resources available to them.

Available data

Concentrating resources on places and people at high *ex ante* risk requires having access to historical (or, if available, real-time) data on the factors that predict where crime will occur and whom it will involve. Prior to the SDSCs, CPD's infrastructure for recording administrative data was already very robust. The Department's data systems track detailed information about reported crimes, victims, and calls for service, in addition to measures of officer activity ranging from arrests, investigatory street stops (Terry stops), traffic stops, and information on officers' locations from GPS units in their vehicles. These data are relatively clean and well-structured, and in most cases go back several years.

The SDSCs expanded CPD's data collection in two ways. First, the Department introduced ShotSpotter's acoustic gunshot detection sensors to each SDSC district. Without ShotSpotter, officers typically rely on calls for service to alert them to shootings. According to officers, it is not uncommon for there to be a 5 to 10-minute delay between when a shooting takes place and when someone calls 911 to report shots fired, often with an incorrect address or location for where they think it occurred. Other times, no call is received at all. After ShotSpotter detects a shooting and a human validates the audio recording, an alert is sent to the SDSC room and to the mobile phones of officers containing the incident's precise location as well as the associated audio. In addition to allowing officers to respond more quickly to the scene and gather evidence, ShotSpotter also provides the Department with an additional and more complete source of data about patterns of gun violence in the city.

Second, SDSC districts expanded their use of CCTV cameras, called Police Observation Devices (PODs). Prior to the SDSCs, monitoring of PODs was very limited due to the cumbersome software available for doing so at the time. The SDSCs introduced additional PODs, which now number in excess of 30,000; upgraded many cameras from standard definition to high definition; and, perhaps most important, integrated all of the Department's PODs into an easy-to-use platform called Genetec. Officers in the SDSC room use Genetec to quickly toggle between cameras, rotate their field of vision, and review stored video footage. When officers in the field respond to a situation, or when a ShotSpotter alert goes off, officers in the SDSC room use the nearest PODs to monitor it and provide information to field units.

In contrast to administrative data, information collected from human sources has historically been more difficult to record and make available. District intelligence officers (DIOs), the primary source of human intelligence within a district, often record information on paper that is stored in binders, making it difficult to access. Beat and tactical officers have no system for logging information gleaned while on patrol or from interactions with residents. Detectives, who also gather information pertinent to understanding crime in the district while investigating cases, are not assigned to districts and do not routinely share their insights with DIOs or patrol officers.

The SDSCs appear to have helped facilitate some additional information collection and sharing, though have probably not changed matters dramatically. The SDSC room itself can serve as a focal point in a district. For example, DIOs or patrol officers sometimes drop by and share pertinent information they have obtained, which is then incorporated into analyses and disseminated within the district. The process of developing a daily briefing, described below, is another mechanism to elicit information about crime from officers. Finally, though not explicitly tied to the SDSCs, the Department has prioritized the collection of data on the frequency of officers' interactions with the community. These data are captured through a call category called a positive community interaction (PCI), which an officer calls in after engaging in such an interaction. The PCI call category has existed since 2015, however it was prioritized significantly, and tracked as part of the Department's CompStat process, beginning in 2017.

Analytical capacity

Despite having a well-developed data infrastructure, CPD's ability to translate data into analyses useful to Commanders for directing resources was very limited prior to the SDSCs' adoption. Crime analysis was carried out by a handful of officers at headquarters who were responsible for collecting intelligence and identifying crime patterns and trends across all 22 districts. As a result, analysis products tailored to each district's unique needs were not available to Commanders. Furthermore, although CPD made available several home-grown software tools for accessing and summarizing the vast quantity of data it collects, including a powerful mapping tool, these were seldom used by officers in police districts. This underutilization of the available data—either due to the cumbersome user interface of the available software, the lack of specific training on crime analysis, the absence of a role dedicated to performing this function, or some combination of all three—created a type of "last mile" problem that made local crime analysis the exception rather than the rule at CPD.

The introduction of the SDSCs significantly increased the analytical capacity available to district Commanders. For the first time in the Department's history, a civilian role specifically devoted to crime analysis was created. The crime analyst works alongside officers and the Commander to develop analytical products that describe recent patterns of criminal activity in the district. The analyst is capable of using all of CPD's existing software tools—as well as several new ones created specifically to automate common tasks—to make detailed maps of reported crimes or calls for service, or profiles detailing where individuals with open warrants had recently been stopped. Along with the officers staffing the SDSC room, the analyst helps draft a daily briefing for the Commander, described in further detail below.

An example of the type of work performed by crime analysts is presented in Figure 9. In 2017, then-Commander of the 7th district, Kenneth Johnson, noting that stolen vehicles were often used as a platform for armed robberies and shootings, asked the district's analyst to examine whether there was an underlying pattern to these thefts. The analyst gathered data on the locations from which cars were stolen and the locations at which they were recovered. This identified a cluster of 18 cars that were all recovered near the same intersection in the 7th district and that had been stolen over a six-month period from commercial and residential areas in the adjacent district. The Commander ordered increased patrols in the area where the cars were recovered, and patrol officers were provided information about the pattern. Shortly thereafter, an individual was arrested who had both an extensive history of motor vehicle theft and a close connection to the victims of a quadruple homicide in a neighboring district. The motor vehicle theft pattern subsided after the arrest.

In addition to crime analysts, the SDSCs also introduced HunchLab, a place-based predictive policing software. While the analytical products developed by the analysts make implicit predictions about where crime is likely to occur based on recent data and intelligence gathered from officers, HunchLab applies a statistical model to data on prior crimes and calls for service, as well as geographic and other features, in order to make explicit predictions about where crime is likely to occur. HunchLab divides each district into approximately 500-foot by 500-foot cells to which it assigns predicted probabilities that certain crimes will occur, separately for each of the day's three shifts (watches). Beat officers can access a map on their Department-issued smartphones on which several "mission boxes" are placed on cells with high predicted probabilities; in the highest-violence districts, these mission boxes indicate a high risk of shootings or robberies. Officers are asked to patrol each HunchLab mission box in their beat in three 15-minute intervals, totaling 45 minutes during their shift. In addition, officers in the SDSC room monitor HunchLab mission boxes using POD cameras accessed through Genetec.

Decision-making processes

Prior to the SDSCs, each Commander had a different process for incorporating information into their decisions about which resources to deploy and how. However, because they lacked access to high-quality analysis of local crime patterns, the resulting deployment decisions were often haphazard. Or, as Commander Kenneth Johnson put it, officers were "just patrolling randomly" and "riding around rubber-necking on the street waiting for something to happen."[19] Commanders relied on informal systems to consult with their officers, tactical teams, and other specialized units about deployment plans for that day or that week; these conversations often happened in passing, and rarely offered the Commander a systematic view into the district's recent criminal activity. To the extent that there was a process in place for shaping patrol's response to recent events, it most commonly took the form of "post-shooting missions": increased patrols within a very small perimeter around the location of a recent shooting that lasted between 3-5 days. This type of mission appeared to generate little buy-in from officers, who estimated the likelihood

[19] https://chicago.suntimes.com/2017/3/29/18358635/violent-crime-falls-in-2-districts-run-by-the-johnson-brothers

of retaliation occurring in such a small area to be low. As a result, and aided by a lack of follow-up and accountability, mission fulfillment may have been lacking.

The SDSCs introduced a structure designed to give Commanders comprehensive, high-quality information in a consistent format each day for use in making patrol decisions. This structure primarily takes the form of the Commander's daily briefing, which usually occurs at 1:00PM. Prepared by the SDSC officers, and featuring the work of the crime analyst, the presentation covers:

- recent crime trends and high-profile arrests
- high-priority open warrants
- deeper analyses into areas of interest, including those raised at previous briefings
- a comprehensive overview of available discretionary resources, including Area units, as well as their current assigned deployment locations
- the locations of HunchLab mission boxes.

In addition to the Commander, SDSC officers, and the crime analyst, others who attend the briefing include district intelligence and tactical officers, as well as non-district personnel such as Area units and detectives, whose input is valued by the Commander and who can relay information back to their respective teams.

The output of the daily briefing is a set of missions ordered by the Commander and information for dissemination to field units. Missions can vary in their complexity. For example, in response to an active conflict between two gang factions, the Commander may order traffic missions on a busy thoroughfare connecting the rival areas in order to interdict firearms that could be used for retaliatory shootings. An upcoming anniversary of a slain gang member's death might result in heightened patrol activity in areas associated with the gang member's rivals. Recent shootings in the vicinity of a gas station or other private business whose owner filed a criminal trespass affidavit with the Department could result in a mission to enforce the affidavit by entering the premises and dispersing trespassers and others engaged in criminal activity.[20] The information produced by the SDSC room is shared with patrol and tactical units during roll calls at the start of each watch. For example, in one district, SDSC officers prepare a binder with the high-priority open warrants discussed in the briefing for tactical officers to review.

In addition to daily briefings, some districts participate in weekly briefings focused specifically on sharing information about recent shootings. In preparation for these weekly briefings, SDSC staff review, in detail, every shooting and homicide that occurred in the district during the past week. Attendees at these briefings include key district personnel, Area units (gang, detectives), and representatives from law enforcement agencies at the County-level (State's Attorney's Office, Probation), State-level (Illinois State Police), and federal-level (U.S. Attorney's Office, FBI, ATF). The goal of these briefings is to assess the likelihood of retaliation following a given shooting, and to focus prevention efforts on those with the highest likelihood of retaliation.

Rollout of the SDSCs

[20] http://directives.chicagopolice.org/directives/data/a7a57b38-14a39f6a-cc814-a39f-8ee270ffb8fdfe8c.html?ownapi=1

The impetus for developing and introducing the SDSC model was the unprecedented increase in gun violence that Chicago experienced in 2016. In September 2016, at the request of the CPD Superintendent Eddie Johnson, the Bureau of Justice Assistance (BJA) at the U.S. Department of Justice engaged Chief Sean Malinowski, then the Chief of Staff to Los Angeles Police Chief Charlie Beck, to lead a team of law enforcement experts to assess CPD's crime fighting strategy. After several months of site visits and interviews, the BJA team concluded its work and produced a set of recommendations for how the Department could improve its crime fighting strategy, particularly at the district level. These recommendations called for a combination of improvements to physical infrastructure, targeted investments in technology, the systematic use of data to inform deployment decisions, and a streamlined intelligence gathering process. Collectively, these reforms were adopted by the Department and came to be known as the SDSCs.

The decision was made to first pilot the SDSCs in the 7th and 11th districts, located on the South and West sides, respectively. The 7th and 11th districts have long experienced some of the highest levels of violence in Chicago. In 2016, each district saw its homicide rate roughly double, and collectively these two districts accounted for almost a quarter of the city's homicides that year, making them the Department's and the City's top priority. Building out the physical SDSC rooms began near the start of 2017 and was complete by February. Within weeks of becoming operational, the Department and the City decided to expand SDSCs to the four remaining so-called Tier 1 districts: the 6th, 9th, 10th, and 15th districts, in addition to the 7th and 11th. By mid-March, the SDSC rooms in all Tier 1 districts were operational.

Despite their rapid build-out of the SDSCs, the Department struggled to quickly implement one component of the model: crime analysts. No such role previously existed at CPD, and there was no precedent for how to create and fill it. Further, the City estimated that the earliest it could hire analysts for the Tier 1 districts was June 2017. Because of the crucial role they play in the SDSC model, and because, as late as December 2016, the surge in Chicago's violence showed no signs of abating, CPD asked the University of Chicago Crime Lab to provide civilian analyst support while the City's hiring process ran its course. The Crime Lab agreed to help and identified two of its analysts to work with CPD in the 7th and 11th districts. At the invitation of Chief Malinowski, these two analysts visited LAPD to observe their implementation of a similar data-driven policing model.

The Crime Lab analysts spent the month of January 2017 embedded in the 7th and 11th districts, speaking with officers, participating in ride-alongs, attending roll calls, meeting with Commanders, and familiarizing themselves with the Department's data infrastructure. As the first civilian analysts to work alongside officers at CPD, the Crime Lab analysts created prototypes of analysis products and daily briefings for Commanders, and developed new tools to automate common tasks and make possible others that were too tedious to do previously. By February 2017, when the SDSCs in the 7th and 11th districts were fully operational, the Crime Lab analysts were working with district leadership to use the new technology and existing data sources to monitor changes in criminal activity and gang conflicts in close to real-time. When the SDSCs in the remaining Tier 1 districts opened in mid-March 2017, the Crime Lab deployed more of its existing analysts and hired additional ones to help staff these new SDSCs while the City continued to ramp up its own hiring. Similarly, when the Tier 2 districts' SDSCs began coming online in early 2018, in the 2nd, 3rd, 4th, 5th, 8th, 12th, and 25th districts, the Crime Lab deployed its analysts to staff these as well until the City's hiring process was able to catch up.

Soon after the SDSC initiative got underway, the Department and the Crime Lab recognized a need for analysis support at the Area level. Approximately 60 percent of CPD's officers work under the direction of district Commanders; the majority of remaining officers are assigned to specialized units, like gang and narcotics teams, managed by Area Deputy Chiefs. In September 2017, the Crime Lab embedded one analyst in each of the three Areas to work closely with those Deputy Chiefs, providing analytical capacity similar to what was being provided to district Commanders. This was designed to reinforce the district-level SDSCs, with Crime Lab Area analysts looking across district boundaries to spot emerging crime patterns and advising Deputy Chiefs' potential resource deployments.

In addition to providing stop-gap crime analysis capacity to CPD, the Crime Lab analysts provided extensive training to SDSC staff. Initially, this training focused on familiarizing SDSC officers with Microsoft Office software (Word, Excel, PowerPoint) to help them with the preparation of the daily briefing. Over time, however, it came to include training in the Department's existing, but underutilized, data infrastructure and the software designed to access it. When CPD's own crime analysts were hired in June 2017, Crime Lab analysts worked closely with them to share best practices learned during the first few months of the SDSCs' existence. Since then, the Crime Lab has organized several half-day training sessions for CPD's crime analysts, covering such topics as place-based analysis techniques and visualizing geospatial data, and convenes CPD's analysts twice a month to discuss new methods of data analysis, updates to tools and dashboards, and to generate ideas for future trainings.

The Crime Lab's role in the expansion of the SDSCs is unique: it provided the organization (and this research team) unprecedented insight into the functioning of the Department, both prior to and after the intervention's launch, while also shaping the intervention itself. The outsize role played by Crime Lab analysts in training their CPD counterparts, the lasting remnants of which include numerous training materials and input incorporated into CPD's standard operating procedures, will continue to have an impact on CPD's SDSCs even after their support ends.

V. SYNTHETIC CONTROLS AND THE SDSCs

A randomized controlled trial to measure the impact of the SDSCs was not an option due to the public safety crisis facing Chicago in early 2017, which resulted in the City prioritizing for reform the police districts experiencing the most gun violence. Nor is it likely that SDSCs' impact could be measured using a regression discontinuity design because Chicago is divided into only 22 police districts.[21] Instead, we turn to a non-experimental panel data evaluation method to estimate the SDSCs' effect on gun violence: synthetic controls.

---

[21] As noted by Schochet (2009), Wing and Cook (2013), and others, regression discontinuity designs have much lower statistical power to detect treatment effects than randomized controlled trials.

Unlike a standard difference-in-differences estimator, which implicitly assigns equal weights to all control group units,[22] the synthetic controls method uses pre-treatment data to assign individual weights to each control group unit, such that their weighted sum—the synthetic control—resembles the treated unit in its pre-treatment characteristics.[23] Due to the variability in how districts implemented the SDSC intervention, we apply this method separately to each of the six Tier 1 police districts that received an SDSC in 2017. We return to this implementation variability in our discussion of the results, as it may provide a source of leverage to better understand what specific police activities are most useful in reducing crime.

Synthetic controls: intuition and pitfalls

Before describing how we apply the synthetic controls method to measure the SDSCs' impact on gun violence, we first outline the intuition behind the method. We pay particular attention to the issue of overfitting, which we describe below, and which we suspect will become more common as this method is applied to increasingly large datasets with many observed and untreated units.

The synthetic controls method, introduced by Abadie and Gardeazabal (2003) and ADH (2010), relies on pre-period observations of the outcome of interest for the treated unit and a set of untreated donor units. Based on these data, the method chooses a weight for each donor unit such that the weighted sum of donor units' observations in each pre-period is close to that of the treated unit. For one treated unit ($j = 0$) and J donor units ($j = 1, \dots, J$), using the observed outcome of interest (Y_{jt}) in the pre-periods ($t \leq T_0$), a synthetic controls estimator chooses a vector of weights $\boldsymbol{w} = (w_1, \dots, w_J)$ to solve an objective function like the following one:

$$\boldsymbol{w}^* = \arg\min_{\boldsymbol{w}} \sum_{t=1}^{T_0} \left(Y_{0t} - \sum_{j=1}^{J} w_j Y_{jt} \right)^2 \tag{1}$$

The synthetic control tries to mimic, as closely as possible, the pre-period observations of the treated unit. The implicit assumption is that, by ensuring a good "pre-period fit," the synthetic control's post-period observations ($t > T_0$) provide a good counterfactual for the missing potential outcome observations of the treated unit in the absence of treatment (Rubin 1974; Imbens and Rubin 2015). This assumption is crucial for the validity of the synthetic controls method, but it is impossible to test directly. Instead, when it is applied in practice, the validity of a synthetic control is typically judged by how closely it mimics the pre-period observations of the treated unit, rather than how reliably it is able to recover the missing potential outcome observations of the treated unit (ADH 2015).[24]

[22] In settings where treatment occurs at different times in different units, difference-in-differences may not assign equal weight to all control group units (see, e.g., Goodman-Bacon 2019).

[23] In most applications of the synthetic controls method, the pre-treatment characteristics of interest are the observed values of the primary outcome. We return to this point later in the section.

[24] "The credibility of a synthetic control depends upon how well it tracks the treated unit's characteristics and outcomes over an extended period of time prior to the treatment" (ADH 2015, p. 500).

Good pre-period fit is an important but potentially misleading indicator of a synthetic control's ability to reliably recover the missing potential outcome observations of the treated unit. A good fit may arise when a synthetic control places weight on donor units that share no underlying relationship with the treated unit but have similar pre-period observations due to idiosyncratic errors. Such donors are chosen on the basis of noise and may not provide any useful signal about the future potential outcome observations of the treated unit, threatening the assumption underlying the validity of the synthetic control. This risk of overfitting is likely to grow with the size of the candidate donor pool, which increases the odds that one or more donors will appear to be a good match for the treated unit due to noise and be assigned weight as a result.

As a stylized example, consider a treated unit with an outcome of interest determined by a vector of unit-specific characteristics, $\boldsymbol{X}_0$, in each period; a treatment effect, α, in post-periods; and an idiosyncratic error:

$$Y_{0t} = f(\boldsymbol{X}_0, t) + \mathbf{1}[t > T_0]\alpha + \varepsilon_{0t}$$

The objective of the synthetic control is to recover the treated unit's missing potential outcome observations, $f(\boldsymbol{X}_0, t)$ for $t > T_0$, using a weighted sum of donors. Suppose there is one signal donor unit available ($j = 1$) with shared unit-specific characteristics ($\boldsymbol{X}_1 = \boldsymbol{X}_0$), and $J - 1$ noise donor units available with different unit-specific characteristics ($\boldsymbol{X}_j \neq \boldsymbol{X}_0, j > 1$). Let the outcome of interest for all donor units be determined by:

$$Y_{jt} = f\left(\boldsymbol{X}_j, t\right) + \varepsilon_{jt}$$

The pre-period observations of the noise donors will likely differ from those of the treated unit due to their different unit-specific characteristics, but some may be similar due to their realizations of the stochastic error term.[25] In contrast, the pre-period observations of the signal donor are likely to be similar to those of the treated unit, but for different realizations of its stochastic error term. Further, the signal donor's post-period observations are, in expectation, equivalent to the missing potential outcome observations of the treated unit. Therefore, the optimal vector of donor weights, $\boldsymbol{w} = (w_1, w_2, \dots, w_J)$, is $\boldsymbol{w} = (1,0, \dots ,0)$. However, there is no guarantee that a synthetic controls estimator will recover this weight vector, particularly if the number of noise donors is large, the length of the pre-period is short, or the variance of the stochastic error distribution is high.[26] In

[25] A donor may also have outcome observations similar to those of the treated unit for reasons other than realizations of the stochastic error term. For example, consider two neighborhoods experiencing low rates of crime: one is socioeconomically advantaged and has low police presence, the other is disadvantaged but has high police presence that deters criminal activity. Both may appear to be plausible candidate donor units for a treated unit also experiencing low rates of crime. However, depending on the socioeconomic conditions and police presence in the treated unit, one donor may be better able to recover the counterfactual than the other, underscoring the potential importance of considering not only pre-period outcome observations but also covariates in assigning donor weights.

[26] For example, holding other factors constant, as J grows very large, so too does the probability of a noise donor with pre-period outcomes that are, by chance, closer to those of the treated unit than the signal donor: $\sum_{t=1}^{T_0}\left(Y_{0t} - Y_{jt}\right)^2 < \sum_{t=1}^{T_0}(Y_{0t} - Y_{1t})^2$ for $j > 1$.

trying to match the pre-period observations of the treated unit using those of the available donors, the synthetic control may overfit by assigning weight to noise donors, undermining its ability to predict the missing potential outcome observations of the treated unit in the post-period.

Although we are not the first to point out this issue, guidance for applied researchers on how to guard against it is lacking. For example, ADH (2010, 2015) warn about the potential for "interpolation" bias if the candidate donor pool contains units with characteristics different from those of the treated unit. However, they offer no recommendation on how to choose candidate donors for the pool. Instead, the estimator that ADH propose seems to incorporate this insight: part of how it chooses weights is to minimize the distance between a vector of characteristics for the treated unit and a vector of characteristics for the synthetic control.[27] But this vector of characteristics can (and in practice often does) include the full set of pre-period observations of the outcome of interest, in addition to other covariates. As demonstrated by Kaul et al. (2018), when the full set of pre-period outcome observations is included, the addition of other covariates proves to be irrelevant; the ADH estimator, by ultimately seeking to minimize the mean squared error between the outcome observations of the treated unit and the synthetic control during the pre-period, ignores the covariates.[28] In other words, in many applications of synthetic controls, the donor pool may contain units that differ substantially from the treated unit, yet the estimator discards covariate information when selecting weights.

Synthetic controls: choices that may affect overfitting

In selecting a synthetic control that can reliably recover the counterfactual, a researcher faces choices in several areas about how best to incorporate the available data, both pre-period outcome observations and covariates. We focus on three areas where the decisions made can affect overfitting: the choice of the candidate donor pool, the choice of the estimator, and the choice of how many outcome variables to use when fitting a synthetic control. Each of these is described in greater detail below.[29]

[27] The ADH estimator relies on a nested optimization for choosing weights. The inner optimization chooses a donor weight vector, W, to minimize the distance between the treated unit and the synthetic control in X, a covariate vector containing candidate predictors, and potentially pre-period observations of Y, the outcome of interest. The outer optimization chooses a matrix, V, that assigns predictor weights to each element of X when choosing W. For a given value of V, $W^*(V)$ is the optimal donor weight vector chosen by the inner optimization. The outer optimization chooses a V^* that minimizes the distance between the treated unit and the synthetic control in Y, using the donor weight vector $W^*(V^*)$.

[28] As Kaul et al. (2018) note, there are alternatives to including the full set of pre-period outcome observations in the vector of characteristics, such as including their average or the last pre-period observation. These alternatives increase the likelihood that the synthetic control assigns weight to donor units with similar covariates to the treated unit, in addition to those with similar pre-period outcome observations. However, it is unclear how to choose from among these ad hoc approaches, each of which discard potentially useful pre-period outcome data.

[29] A fourth factor that affects the likelihood of overfitting is the length of the pre-period. All things being equal, a longer pre-period reduces the likelihood of noise donors being assigned weight by a given estimator. However, in most settings, researchers already use all the pre-periods in the data they have available. It may be possible in some contexts to obtain additional outcome observations measured at shorter time intervals, but doing so can result in a noisier outcome measure, potentially offsetting any benefit from this approach.

Although mentioned as a possible strategy by ADH, limiting the candidate donor pool to units with characteristics similar to the treated unit is a technique that appears to seldom be used in practice.[30] Instead, researchers commonly include all available donor units in the pool, perhaps in part because synthetic controls is often applied in settings with few available donor units to begin with. Rather than limiting the donor pool, researchers typically rely on the inclusion of predictor covariates when applying the ADH estimator to guide its choice of donor unit weights—which, as noted above, Kaul et al. (2018) find to be irrelevant when paired with all available pre-period observations of the outcome of interest, as is often done.

An alternative approach is to limit the candidate donor pool prior to choosing the synthetic control weights. For example, techniques from the matching literature, like propensity score estimation, exist to address a similar problem: finding control observations comparable to treatment observations when there are multiple dimensions on which their similarity can be assessed. Using available baseline covariate predictors, one could estimate a propensity score for each donor unit and limit the candidate donor pool to those with scores within some bandwidth around the treated unit. If sufficient pre-period data are available, they can be used to inform the bandwidth choice. In addition to guarding against overfitting, limiting the size of the candidate donor pool, particularly when it is large relative to the number of available pre-period outcome observations, increases the likelihood of there being a unique solution to the synthetic control estimation.

Another choice that can affect the degree of overfitting is that of the estimator used to generate weights. Recent developments in panel data causal inference methods provide researchers with alternatives to the canonical ADH synthetic controls estimator.[31] We consider one such alternative: the estimator proposed by DI (2017). The choice of estimator bears on the issue of overfitting in several ways, but we focus here on how they differ in the restrictions they place on the choice of weights. Restrictions impose a 'cost' to assigning non-zero weight to a donor. In the stylized example of signal and noise donors, when assigning weight is costless and the goal is to produce a good pre-period fit, the estimator will be unconstrained; it will assign weight to noise donors if they can improve pre-period fit, even marginally.[32] When assigning weight is costly, however, the estimator will first assign weight to donors that are similar to the treated unit across the full series of pre-period outcome observations, which are more likely to be signal donors.

The ADH estimator imposes weight restrictions that make it unsuitable when the treated unit is an outlier relative to the candidate donor pool, as is the case in our setting. The two

[30] One example is Billmeier and Nannicini (2013), who estimate the impact of economic liberalization on a country's real GDP per capita. Because their study includes cases of economic liberalization from countries all over the world, they consider two approaches to choosing the candidate donor pool for each treated unit: limiting donors to other countries in the same region as the treated country (type A), or not limiting donors to only countries in the same region (type B). While the authors acknowledge the trade-offs between these two approaches, they nevertheless rely on the type B donor pool to generate their treatment effect estimates in cases where the pre-period fit using the type A donor pool is poor.

[31] See, e.g., Athey et al. (2018); Ben-Michael, Feller, and Rothstein (2018); Powell (2018); Arkhangelsky et al. (2019).

[32] If the number of donor units exceeds the number of pre-periods, the resulting fit from an unconstrained estimator will be perfect.

restrictions imposed by ADH are that weights must be non-negative ($w_j \geq 0$) and sum to one ($\sum w_j = 1$). Together, these conditions restrict the synthetic control to fall within the convex hull of donor units and limit extrapolation outside the support of the data. As a result, if the treated unit is an outlier and falls outside the convex hull of the donor units, the ADH synthetic control will exhibit poor pre-period fit.

In contrast, the DI estimator produces a synthetic control with better pre-period fit when the treated unit is an outlier by relaxing the ADH constraints. However, it may be more susceptible to overfitting because it does not consider covariates. Instead of requiring that weights be non-negative and sum to one, the DI estimator uses a flexible, data-driven approach—elastic net regularization—to assign smaller weights to each donor unit and to assign non-zero weight to fewer donor units. Specifically, the DI estimator chooses both a vector of weights and an intercept to solve:

$$(\boldsymbol{w}^*, \mu^*) = \underset{\boldsymbol{w},\mu}{\arg\min} \sum_{t=1}^{T_0} \left(Y_{0t} - \mu - \sum_{j=1}^{J} w_j Y_{jt} \right)^2 + \lambda \left(\frac{1-\alpha}{2} \|\boldsymbol{w}\|_2^2 + \alpha \|\boldsymbol{w}\|_1 \right) \quad (2)$$

Unlike the ADH estimator, the DI estimator imposes no direct constraint on the weights, which can be negative and sum to any value; instead, their value is indirectly constrained by an elastic net penalty term that prioritizes weight vectors with fewer non-zero ($\|\boldsymbol{w}\|_1 = \sum |\omega_j|$) and smaller ($\|\boldsymbol{w}\|_2^2 = \sum \omega_j^2$) entries.[33] Furthermore, the DI estimator does not incorporate covariates in any way, which allows for the choice of an intercept term but may make it particularly susceptible to overfitting.[34] We return to this point below when we discuss how we apply the DI estimator, including a modification we make intended to guard against overfitting.

Finally, the choice of how many outcome variables to use when fitting a synthetic control can also affect overfitting. Until now, we have assumed a single outcome of interest, as is the case in most applications of synthetic controls. However, fitting a synthetic control to multiple outcomes simultaneously reduces the likelihood of assigning weight to noise donors, if the idiosyncratic errors across outcomes are uncorrelated or weakly correlated. Intuitively, if a donor

---

[33] The magnitude of the elastic net penalty, and the relative weight placed on the Lasso and ridge regression terms, are determined by a pair of hyperparameters (α^*, λ^*) chosen through a cross-validation procedure. In this procedure, for a given pair of (α, λ), first estimate a weight vector and intercept for each donor unit j: ($\widehat{\boldsymbol{w}}(j; \alpha, \lambda)$, $\hat{\mu}(j; \alpha, \lambda)$). Then, using the estimated weight vector and intercept, calculate the predicted post-period ($t > T_0$) outcome series for each donor unit: $\hat{Y}_{jt}(\widehat{\boldsymbol{w}}, \mu) = \hat{\mu}(j; \alpha, \lambda) + \sum_{i \neq j}^{J} \widehat{w}_i(\alpha, \lambda) Y_{it}$. Finally, calculate the cross-validation error across all donor units as the average difference between the observed and predicted post-period outcome series: $CV(\alpha, \lambda) = \frac{1}{J} \sum_{j=1}^{J} \left(Y_{jt} - \hat{Y}_{jt} \right)^2$. The values of ($\alpha^*$, λ^*) minimize this cross-validation error.

[34] As noted by DI (2017), the ADH estimator does not permit the choice of an intercept because the inner optimization it uses to choose a donor weight vector, W, seeks to minimize the distance between the treated unit and the synthetic control in X, a covariate vector containing candidate predictors that may be qualitatively different from each other. An intercept would create a fixed difference between the treated unit and the synthetic control for all covariates, violating scale invariance. In contrast, the DI estimator only seeks to minimize the distance between the treated unit and the synthetic control in pre-period observations of a single variable, Y, the outcome of interest.

unit's realizations of one outcome are similar to those of the treated unit due to idiosyncratic errors, it is less likely that realizations of a different outcome are also similar to those of the treated unit.

Fitting a synthetic control to multiple outcomes simultaneously has a second benefit as well. When the researcher can observe both a primary outcome of interest and secondary outcomes representing potential mediating mechanisms through which the treatment may affect the primary outcome, obtaining a single vector of synthetic control weights facilitates interpretation of those mediating mechanisms. For example, consider an intervention intended to reduce mortality through the provision of health insurance. The synthetic control used to estimate the treatment effect on mortality (primary outcome) will likely differ from the synthetic control used to estimate the treatment effect on health insurance receipt (secondary outcome). This makes it difficult to know whether a reduction in mortality is due to greater health insurance receipt, since the two effects are derived from comparisons with two different synthetic controls.

Fitting a synthetic control to multiple outcomes simultaneously turns out to be relatively straightforward. Using the DI estimator as an example, for outcomes $n = 1, \ldots, N$, the estimator chooses a $1 \times J$ vector of weights and a $1 \times N$ vector of intercepts to solve:

$$(\boldsymbol{w}^*, \boldsymbol{\mu}^*) = \underset{\boldsymbol{w},\boldsymbol{\mu}}{\arg\min} \sum_{n=1}^{N} \frac{1}{T_{0n}} \sum_{t=1}^{T_{0n}} \left(\frac{Y_{0nt} - \mu_n - \sum_{j=1}^{J} w_j Y_{jnt}}{\frac{1}{J} \sum_{j=0}^{J} Y_{jnt}} \right)^2 + \lambda \left(\frac{1-\alpha}{2} \|\boldsymbol{w}\|_2^2 + \alpha \|\boldsymbol{w}\|_1 \right) \qquad (3)$$

Relative to the original DI estimator in equation (2), this version replaces the mean squared error (MSE) term with a mean squared percentage error, to reflect the fact that the different outcomes Y_{jnt} may be scaled differently. Failing to account for these scale differences will implicitly give greater priority to weight vectors that minimize the errors of outcomes with larger absolute deviations.[35] In practice, one can operationalize this approach by providing the DI estimator with $N * T_{0n}$ pre-period observations, with each set of T_{0n} observations corresponding to a different outcome, a technique similar to one used by Robbins, Saunders, and Kilmer (2017).[36]

Synthetic controls: application to the SDSCs

We start our evaluation by applying the ADH estimator to each SDSC district. Because our focus is their impact on gun violence, we use a count of shooting victims as our primary outcome, scaling by a district's population to account for differences in size.[37] Reliable data on shooting

[35] An adjustment is also made to account for the possibility that each outcome Y_{jnt} may be observed for a different number of pre-periods, T_{0n}.

[36] Though motivated here, in part, by the desire for a single synthetic control to facilitate comparisons of treatment effects across primary and secondary/mediating outcomes, this approach can be used to include any kind of covariate into the DI estimator, helping to address the weakness described earlier.

[37] We have access to data on each shooting incident in the city, which provides us flexibility in how we aggregate them. However, we have found that the synthetic controls method performs poorly when time periods are shorter and, correspondingly, many of them contain no shooting incidents for the treated district.

victims are available going back to 2010, providing us with 14 biannual pre-period observations through the beginning of 2017 when the first SDSCs were implemented in the 7th and 11th districts.[38] We use all of the available pre-period outcome observations when applying the ADH estimator and, consequently, do not incorporate any covariates; we revisit the role of covariates in our application below.

The ADH synthetic control is unable to produce a good pre-period fit (Figure 10, using the 7th district as an example). This is not surprising: the SDSC districts are outliers compared to the 16 donor districts, the ADH estimator constrains the synthetic control to be within the convex hull of the donor districts, and it does not permit an intercept. Improving this situation will require either turning to a more flexible estimator or finding a candidate donor pool with donor units experiencing levels of gun violence more comparable to those of the SDSC districts.

Before turning to estimators that may be better equipped to handle outlier treated units, we first consider whether a more diverse candidate donor pool is available. Relying on the fact that Chicago's 22 police districts are subdivided into approximately a dozen beats each (Figure 11), and that some of the 16 donor districts contain beats with high rates of gun violence, we redefine our pool as the 196 donor beats within the 16 donor districts. Comparing the shootings per capita of SDSC districts to those of donor districts and donor beats, respectively, we see that the SDSC districts are not outliers to the same degree relative to the donor beat distribution (Figure 12). Applying the ADH estimator with the pool of donor beats yields synthetic controls with much better pre-period fit, though still with significant deviations (Figure 13). Though a more diverse candidate donor pool improves the performance of the ADH estimator, it remains true that the SDSC districts are outliers relative to the donor beats in their rates of gun violence.

We next turn to the DI estimator and apply it to each SDSC district, using the pool of candidate donor beats.[39] By relaxing the constraints imposed by the ADH estimator, the DI synthetic control achieves a much better pre-period fit (Figure 14). However, we might be worried that, with weaker constraints and so many donor beats from which to choose, the DI estimator achieves this impressive pre-period fit by overfitting, which would compromise the reliability of the synthetic control in the post-period.

This concern is exacerbated in our settings, with multiple treated units that are outliers and may be differentially susceptible to overfitting, because of how the DI estimator chooses two key parameters that govern how weights are determined. As proposed by DI, the two hyperparameters, (α, λ), are chosen using a cross-validation procedure to minimize error in the *post-period* among

[38] The biannual periods are from March through August and September through February. In all calculations and figures, we report monthly averages within each period. Because the first SDSCs opened in February 2017, and expanded to the control districts beginning in January 2018, the last pre-period is from September 2016 through January 2017 (5 months), the first post-period is from February 2017 through August 2017 (7 months), and the last post-period is from September 2017 through December 2017 (4 months).

[39] The cross-validation procedure outlined by DI (2017) requires estimating a synthetic control for each donor unit to choose the hyperparameters (α^*, λ^*). Due to difficulty estimating a synthetic control for each donor beat, we instead estimate one for each of the 16 donor districts as part of the cross-validation procedure.

the donor units.[40] This procedure yields a single pair of hyperparameters to be used in determining weights for all the treated units, relying only on data from the donor units. This may be particularly problematic here, where our six treated units are outlier police districts on the South and West Sides of Chicago. Furthermore, due to differences in the availability of local non-treated beats in the donor pool, we suspect that some of our treated districts may be more susceptible to overfitting than others.[41]

We propose a modification to the DI estimator that aims to address these concerns. This modification departs from the original DI estimator by choosing hyperparameters separately for each treated unit using a time-series cross-validation technique.[42] Unlike the original DI estimator, which chooses hyperparameters that minimize error in the *post-period* among the donor units, the modified version chooses hyperparameters that minimize error in the *pre-period* for just the treated unit in question. The implicit assumption in the modified approach (that treatment has not yet occurred in the pre-period) is considerably weaker than the one in the original DI approach (that the donor units are unaffected by treatment in the post-period). Furthermore, by choosing hyperparameters that minimize prediction error within the treated unit across a series of one-step-ahead forecasts, the estimator builds in some additional protection against overfitting.

We apply the modified DI estimator along with two of the additional safeguards against overfitting described earlier: limiting the candidate donor pool and fitting a synthetic control to multiple outcomes simultaneously. To limit the candidate donor pool, we estimate a propensity score for each candidate donor beat using a vector of characteristics drawn from Census data.[43] In the modified DI estimator cross-validation step, in addition to searching over possible values of the hyperparameters (α, λ), we also search over potential bandwidths that determine how many of the donor beats that are "closest" in characteristics-space to include in the estimation. After

[40] This approach relies on the assumption that, absent any treatment effect, synthetic controls chosen for the donor units should closely track their observed outcomes in the post-period. This is a strong assumption, particularly in settings where it is plausible that treatment may result in spillover effects on donor units, as is the case here.

[41] Gun violence on the West Side is thought to be driven by the narcotics trade that occurs along the I-290 corridor, while gun violence on the South Side is considered more idiosyncratic and less purposeful in nature, often revolving around personal disputes. This characterization is a gross oversimplification, but it highlights the fact that West Side donor beats may be poor candidates for a South Side treated unit, and South Side donor beats may be poor candidates for a West Side treated unit. Furthermore, three of the SDSC districts—the 15th, 11th, and 10th—encompass most of the gun violence on the West Side, leaving relatively few non-treated West Side districts from which donor beats can be drawn compared to the South Side, where the three SDSC districts—the 6th, 7th, and 9th—have a comparatively larger pool of non-treated South Side donors in the candidate pool.

[42] For each treated unit, use a subset of sequential pre-periods as a training set, with the one pre-period following as a test set. For a given value of (α, λ), estimate a weight vector and intercept using the training set, and use these to predict the outcome in the single-period test set, recording the MSE. Then, lengthen the training set by one period, using the next available pre-period as the new test set. Repeat this process until the last pre-period is the test set. Calculate the cross-validation error as the average MSE across all test sets. The values of (α^*, λ^*) minimize this cross-validation error.

[43] Propensity scores are estimated using a logistic regression on the sample of candidate donor beats $(D = 0)$ and beats in the treated district $(D = 1)$. Predictors include total population, fraction of households with annual income under $40,000, fraction African American, and fraction Hispanic.

choosing hyperparameters and a donor pool to minimize out-of-sample prediction error, we fit a synthetic control to six outcomes simultaneously: two primary outcomes of interest (shooting victims per capita, Part I violent felony incidents per capita) and four secondary outcomes that measure potential mediating mechanisms (warrant arrests per capita, gun arrests per capita, overall arrests per capita, and traffic stops per capita). For a given outcome n of a treated unit, Y_{0nt}, and the associated synthetic control, $\hat{Y}_{0nt} = \sum_{j=1}^{J^*} w_j^* Y_{jnt}$, we estimate the treatment effect in period t as:

$$\hat{\tau}_{nt} = Y_{0nt} - \hat{Y}_{0nt} \tag{4}$$

Having outlined how we choose our synthetic control and generate treatment effect estimates, we now describe how we assess the significance of those estimates. We start with the placebo test suggested by ADH (2010, 2015), in which a synthetic control is calculated for each donor unit, along with a placebo treatment effect. Under the null hypothesis that the donor units are unexposed to any treatment, the estimated treatment effect for the treated unit is compared to this distribution of placebo treatment effects. The fraction of placebo treatment effects greater than or equal to the estimated treatment effect is the *p*-value.

Due to difficulty estimating a synthetic control for each donor beat, we instead estimate one for each of the 16 donor districts. But this yields a distribution of placebo treatment effects with only 16 values, limiting our ability to assess precisely how extreme the estimated treatment effect is. For example, if the estimated treatment effect in district A is just slightly larger than the largest placebo treatment effect, and the estimated treatment effect in district B is much larger, both will receive *p*-values of zero. Under different realizations of the placebo treatment effects distribution, this *p*-value for district B may hold, but the one for district A may not.

We resolve the inference challenge posed by having too few donor districts through the use of a resampling method. Using the fact that each police district is comprised of individual beats, we create resampled donor police districts by sampling beats with replacement from within each actual donor district. These resampled police districts are slightly perturbed versions of the original donor districts. As a result, in the example above, the estimated treatment effect in district A may no longer be larger than the largest placebo treatment effect, increasing its *p*-value. Using resampled police districts dramatically increases the size the donor district pool for the inference procedure, while preserving the geographic clustering of beats by resampling within districts rather than across them, which may be important in environments in which statistical noise has a strong patterning by place and time.[44]

[44] This is related to but slightly different from the approach employed by Robbins, Saunders, and Kilmer (2017), who generate placebo areas using a permutation technique that groups together many comparison areas that are smaller (block-level) than their treatment area (neighborhood-level). As a result, their placebo areas are random assortments of small comparison areas that lack the structure of the treatment area, requiring them to standardize their effect estimates to guard against the resulting bias. In contrast, we avoid this issue by creating placebo districts using a resampling procedure, wherein we resample beats from existing comparison districts with replacement, preserving the structure of those comparison districts in the process.

Spillovers

One concern worth noting is the potential for bias due to treatment spillover from the introduction of the SDSCs, particularly given the close geographic proximity of some candidate donor units. This potential spillover could operate through several channels. For example, if treatment induces a reallocation of resources across districts, then the outcomes of donor districts may be affected by treatment even if they did not receive an SDSC.[45] Even if resources across districts remain fixed, individuals deterred from engaging in crime in SDSC districts may do so elsewhere.

Though important to take seriously, we think the magnitude of any spillover effect is likely to be small for the outcomes of interest, for two reasons. First, the SDSC intervention is designed to change within-district resource allocations—which district units are deployed where and when—rather than across-district allocations. As discussed in Section IV, the SDSC provides a district Commander with information, analytical capacity, and a management process with which to implement a proactive policing strategy. The district Commander does not have command over the Area units operating in his or her district, nor do data from the GPS units within each vehicle in CPD's fleet and attendance records suggest there was an influx of resources into districts, either from the Area or from elsewhere in the city, following the SDSCs' introduction (see, e.g., 7th district, Figure 15).

Second, the primary outcome of interest—violent crime, and particularly gun violence—is among the crimes least likely to be susceptible to displacement by enhanced policing. Gun violence is often tied to a feature of a specific location that is immobile in the short run, such as a lucrative drug-selling corner or the block of a rival street organization member, or it occurs following an altercation. In neither case is such gun violence likely to move several blocks away if deterred from occurring in its original location. This is consistent with the best available evidence on crime displacement, which detects adverse spillovers for property crime but not violent crime (Blattman et al. 2019).

VI. THE IMPACT OF AN SDSC ON SERIOUS CRIME

We turn to estimating the impact of the SDSCs on serious crime in each of the six Tier 1 districts using the modified DI estimator described in the previous section. Our focus is on two primary measures of serious crime: shooting victims and Part I violent felonies per capita.[46] We first estimate synthetic controls separately for each of these outcomes in each of the SDSC districts. Then, we estimate a single synthetic control for each SDSC district that fits to these primary outcomes along with four secondary outcomes: gun arrests, warrant arrests, overall arrests, and

[45] The sign of any bias resulting from such a reallocation is ambiguous. If the SDSC reduces violence in a treated district, then resources may be shifted from there to districts without an SDSC. On the other hand, the Department may have an incentive to bolster the SDSC effort by allocating additional officers to those districts.

[46] Part I violent felonies include homicide, criminal sexual assault, robbery, aggravated assault, and aggravated battery.

traffic stops per capita.[47] The first set of estimates serve as our main results for the impact of the SDSCs on serious crime. The second set of estimates serve as both a robustness check and as a way to understand more about the potential mechanisms driving the primary results.

Because we conduct multiple hypothesis tests throughout this analysis, we must correct for multiple comparisons. We do this in two ways. First, we control for the familywise error rate (FWER), or the probability of falsely rejecting one or more true null hypotheses (type I error). Second, we control for the false discovery rate (FDR), or the proportion of true null hypotheses that are falsely rejected among the total number of rejected hypotheses ("discoveries"). Controlling for the FWER is generally the more conservative multiple comparison procedure, as it limits the probability of making any type I error. In contrast, controlling for the FDR is generally the less conservative multiple comparison procedure, as it limits the fraction of false rejections among the set of discoveries. We correct for multiple comparisons across districts and outcomes separately for the main and secondary sets of results, reporting multiple comparisons-corrected q-values alongside uncorrected p-values.[48]

Table 1 reports estimates of the SDSCs' impacts across several outcomes in the 7th district. The 7th district experienced a 30 percent decline in shooting victims and a 12 percent decline in Part I violent felony incidents following the introduction of the SDSC. This implies that there were 94 fewer shooting victims in the 7th district during the 11 months of 2017 after the SDSC was introduced than there would have been otherwise. For both outcomes, the synthetic controls are able to closely match the observed pre-treatment patterns in the 7th district, and the placebo tests make clear that these reductions are outliers relative to their respective placebo effect distributions (Figures 16 and 17). Both estimates are highly statistically significant even after adjusting for multiple comparisons. When the synthetic control is estimated jointly across both the primary and secondary outcomes, the treatment effects for the primary outcomes remain qualitatively similar and the fit of the synthetic control remains good (Figure 18), but the precision of the estimates erodes after adjusting for multiple comparisons. The treatment effect estimates for the secondary outcomes suggest substantial increases: 27 percent (gun arrests), 36 percent (warrant arrests), 20 percent (overall arrests), and 78 percent (traffic stops). However, the estimates for gun arrests and overall arrests are imprecisely estimated, and the increase in traffic stops predates the introduction of the SDSC (Figure 18). In contrast, the estimated increase in warrant arrests occurs after the SDSC's introduction and is precise when considered separately (column 2), but we cannot be confident that it differs from zero after adjusting for multiple comparisons (columns 3-4).

---

[47] A warrant arrest is an arrest made on the basis of an outstanding warrant being located in the Law Enforcement Agencies Data System (LEADS) database. A gun arrest is an arrest of a person found to illegally possess a firearm or after the use of a firearm in a shooting that did not result in injury to a person.

[48] To control for the FWER, we use the Holm-Bonferroni correction (Holm 1979), which is uniformly more powerful than the classic Bonferroni correction and does not rely on assumptions about the independence of the tests being performed, as does the more powerful procedure suggested by Hochberg (1988). To control for the FDR, we use the Benjamini-Yekutieli procedure, which similarly does not rely on assumptions about the independence of the tests being performed (Benjamini and Yekutieli 2001). Both corrections are implemented using the Stata `qqvalue` package. For the main set of results, these corrections account for 12 comparisons (6 districts, 2 primary outcomes per district). For the secondary set of results, these corrections account for 36 comparisons (6 districts, 2 primary and 4 secondary outcomes per district).

Table 2 reports analogous estimates for the 11th district. Relative to the 7th district, the 11th district experienced a more modest 13 percent reduction in shooting victims and a comparable 10 percent reduction in Part I violent felony incidents, although both estimates are not statistically distinguishable from zero after adjusting for multiple comparisons. Neither are these estimates robust to estimation of the synthetic control jointly across both primary and secondary outcomes, with the estimated treatment effects changing sign. Similarly, the quality of the synthetic control fit, while high when estimated separately for the primary outcomes (Figures 19 and 20), degrades when estimated jointly across primary and secondary outcomes (Figure 21).

Tables 3-6 report estimates for the remaining SDSC districts: the 6th, 9th, 10th, and 15th. Most of the treatment effect estimates for these districts, across both primary and secondary outcomes, are very imprecisely estimated. The notable exception is the 15th district, where the treatment effect estimate suggests that shooting victims *increased* following the introduction of the SDSC. To assess the credibility of this result, we must examine the quality of the pre-period fit to see whether the synthetic control is able to approximately match the outcome series for shooting victims, particularly immediately prior to the treatment. When the synthetic control is estimated separately for the shooting victim outcome, it appears to underestimate the observed pattern in the 15th district in 2016, just prior to the SDSC's introduction (Figure 22). When the synthetic control is estimated jointly for the primary and secondary outcomes, this prediction error increases substantially (Figure 23). These deviations cause the synthetic control to significantly underestimate the true series, partially explaining why we estimate an adverse treatment effect in the post-period.

Perhaps the clearest change in police behavior induced by the SDSCs is in arrests initiated by officers monitoring the Department's POD cameras. As described in Section IV, POD cameras predated the SDSCs, though their number and quality—from standard definition to high definition—both expanded as part of the intervention. Furthermore, the cumbersome software previously available for monitoring the PODs was replaced by an easy-to-use platform that made them more accessible to officers. As a result, with each wave of new SDSCs—in Tier 1 districts in early 2017, and in Tier 2 districts in early 2018—the number of POD-initiated arrests increased substantially (Figure 24).

VII. CONCLUSION

We study the impact of SDSCs: a package of reforms to basic management practices within several high-violence police districts in Chicago. These reforms are intended to help CPD execute a proactive policing strategy that calls for focusing resources on the places and people at highest risk for serious crime. Though outwardly an intervention focused on data and technology enhancements, at their core the SDSCs mostly layer management improvements—tasking one person to oversee data analysis, setting up a process for analysis to inform decision-making—atop a data infrastructure that predates them.

Our results suggest that any reductions in serious crime stemming from the SDSCs were not universally shared across the districts that received them. The 7th district experienced substantial and precisely estimated decreases in shooting victims and Part I violent felonies, while estimates of the reductions in the 11th district were smaller and less precise. Results in the other five SDSC districts were either imprecisely estimated or inconclusive. This underscores an important point about the SDSC intervention: as a reform meant to improve implementation of an existing policing strategy, it, too, may be subject to imperfect implementation.

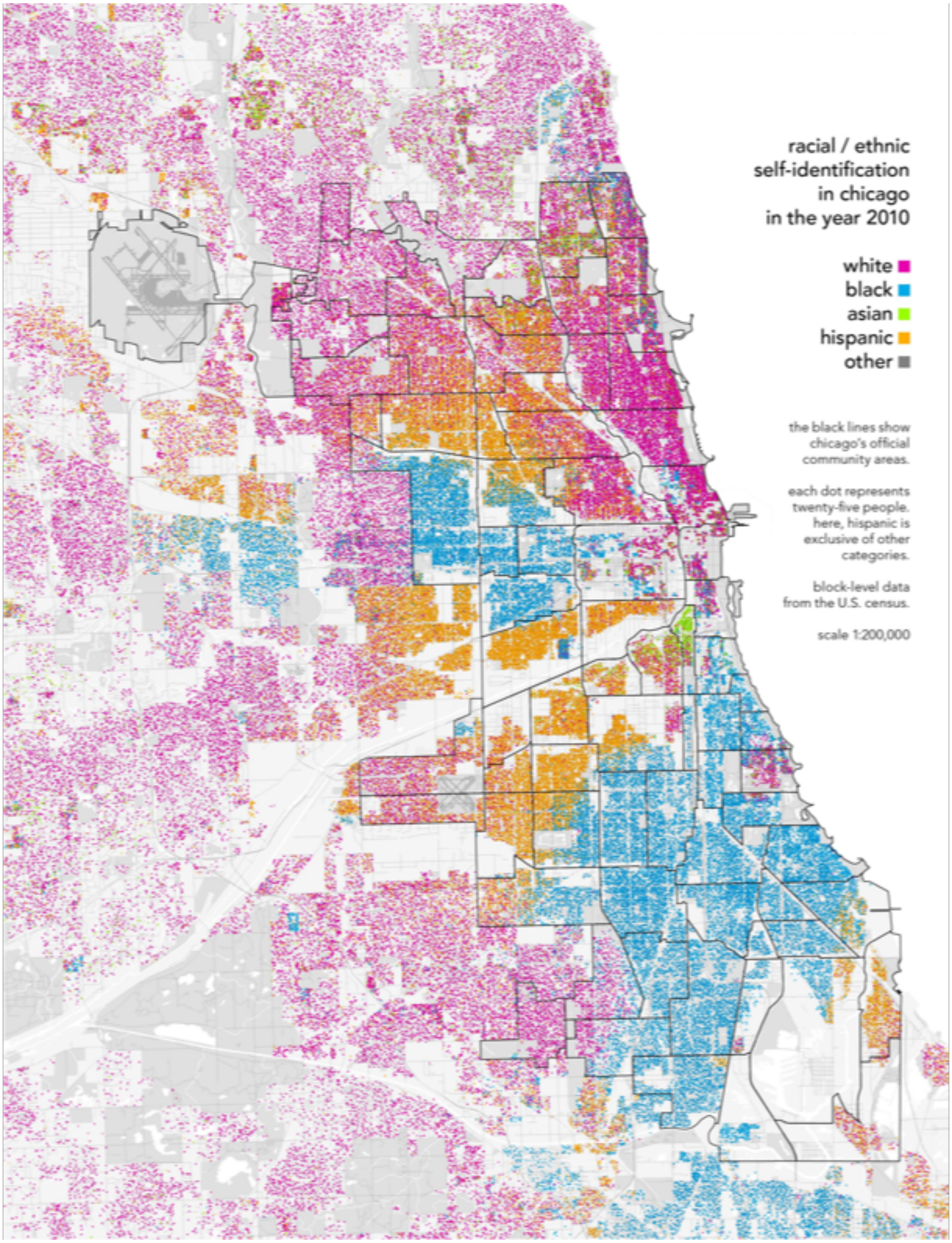
We will continue to study which aspects of the SDSC appear to drive the crime improvements experienced in the 7th district. One potentially important factor is community engagement, the third pillar of the proactive policing strategy described earlier. Although not directly a focus of the SDSC intervention, some district Commanders, particularly then-Commander Kenneth Johnson of the 7th district, made a point of emphasizing the importance of officers engaging in positive interactions with members of the community. Commander Johnson's emphasis that his officers do this may be reflected, in part, in the staggering increase in the volume of community interaction calls, or PCIs, they made following the SDSC's implementation (Figure 25). Of course, we do not mean to suggest that this increase is entirely or even mostly reflective of greater engagement with community members by officers. However, it is noteworthy that this increase was by far the largest in the 7th district, suggesting either a real change in officer behavior, a greater willingness to follow through on the Commander's orders, or some combination of the two that merits further study to understand what role, if any, they played in the serious crime reduction there.

These results also highlight an important dimension of potential policing reform that seldom receives attention in the academic literature. Many studies on policing and crime control focus on estimating how increasing the *size* of the former may affect the latter. A separate literature is concerned with the efficacy of the specific tactics and strategies employed by police (Weisburd and Eck 2004). Both of these literatures offer practical guidance to policymakers, but they are of limited use when the level of policing resources in the short run is fixed, and when existing strategy and tactics already reflect best practices. In this scenario, which Chicago and other cities find themselves in, the available margin for improvement is to strengthen the implementation of the existing strategy, such as through reforms to management practices that can improve productivity.

Perhaps the literature most relevant for understanding which interventions might improve policing and reduce crime on the implementation margin is not the literature about policing and crime at all, but rather the literature about the relationship between management practices and productivity (Bloom and Van Reenen 2007; Bloom et al. 2013). This literature, broadly speaking, suggests that adopting superior management practices can have a sizeable effect on firms' productivity. It also suggests that these practices may not be adopted on their own due to information barriers, a phenomenon that is common not only in policing but other municipal agencies as well.

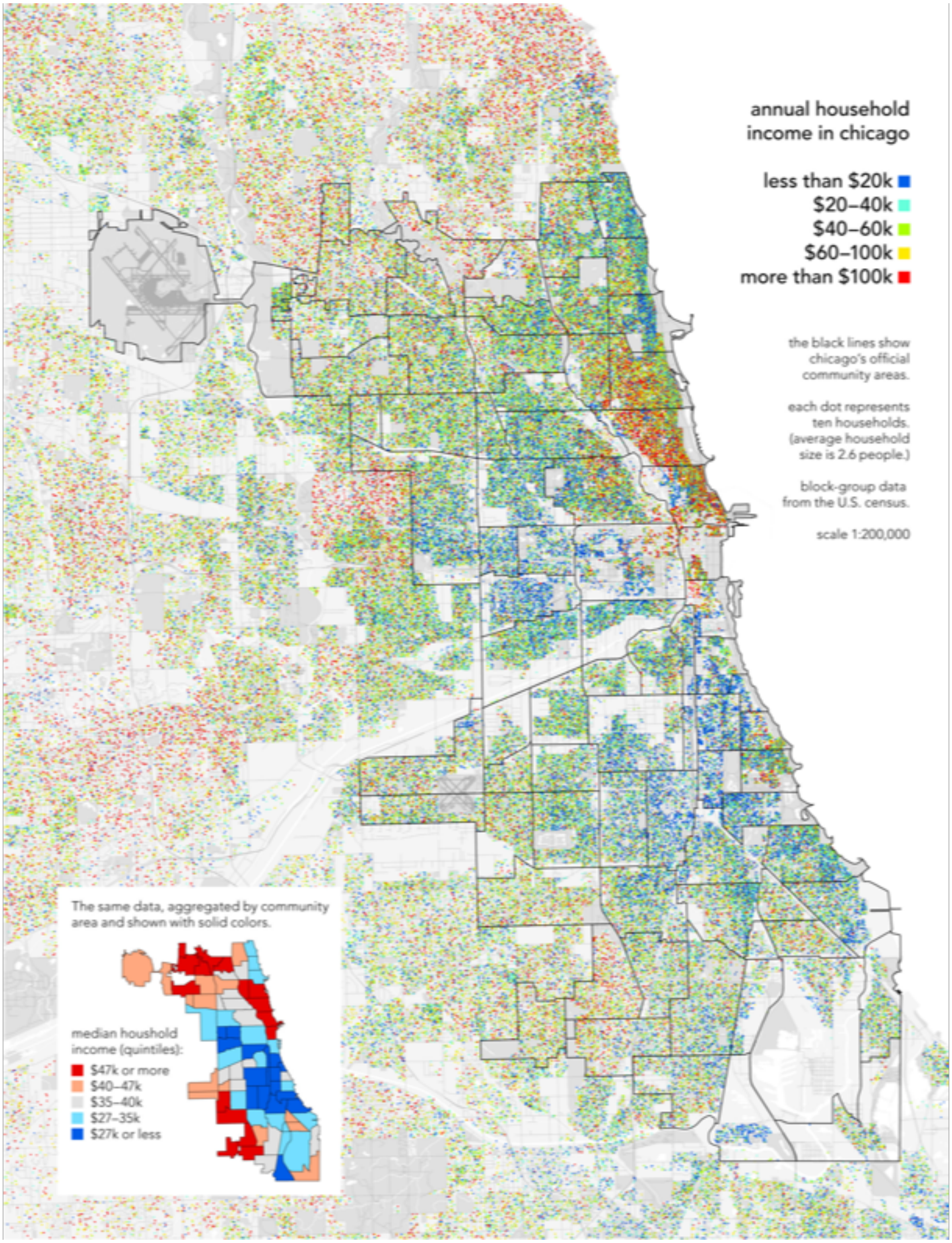
Finally, it is worth noting that, even if the SDSCs' implementation was uneven, they may still represent a worthwhile investment. The cost to build an SDSC is approximately $2 million, with ongoing operating costs likely to be substantially less than this figure. If the SDSC in the 7th district was able to prevent 94 shooting victimizations from occurring, then using an estimated cost per gunshot injury derived from contingent valuation methods of $1.5 million in 2019 dollars (Cook and Ludwig 2000), that one SDSC generated savings of $141 million in a single year. Compared to CPD's annual budget of $1.6 billion in 2018, such investments in better management are modest and seem likely to pay for themselves.

Figure 1. Racial segregation across Chicago’s community areas



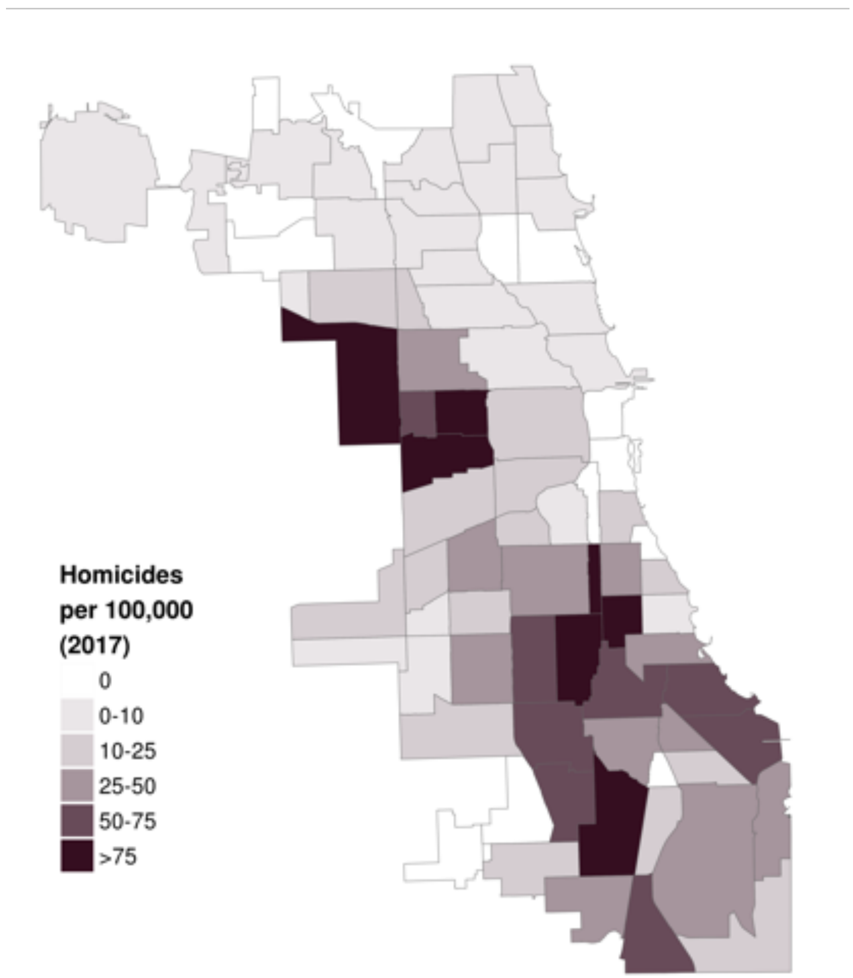
Notes: Bill Rankin (2009). http://www.radicalcartography.net/index.html?chicagodots

Figure 2. Income segregation across Chicago's community areas



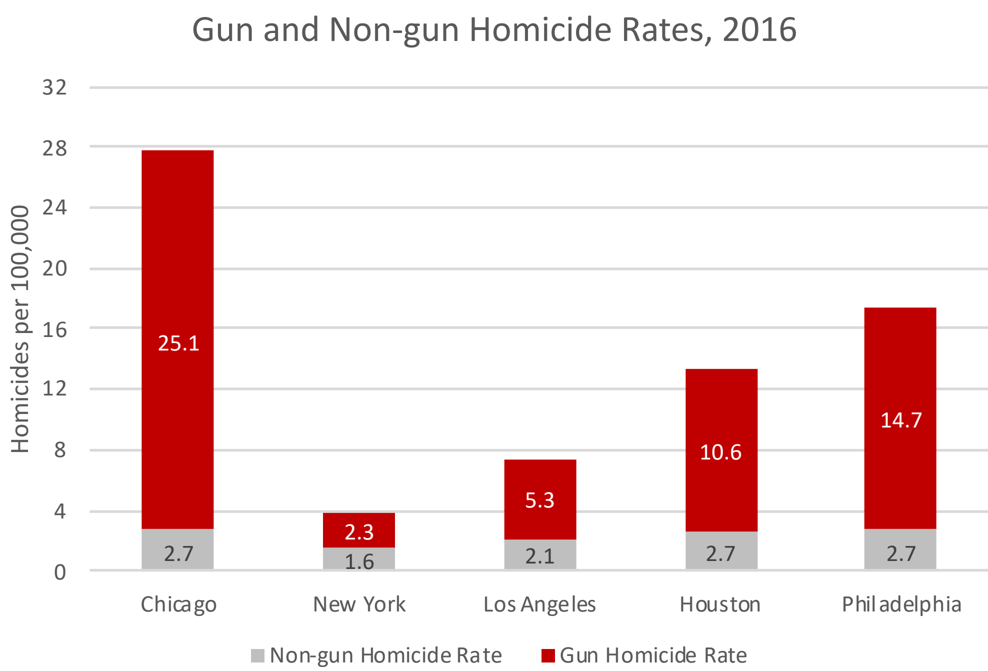
Notes: Bill Rankin (2009). http://www.radicalcartography.net/index.html?chicagodots

Figure 3. Homicide victims per 100,000 across Chicago's community areas, 2017



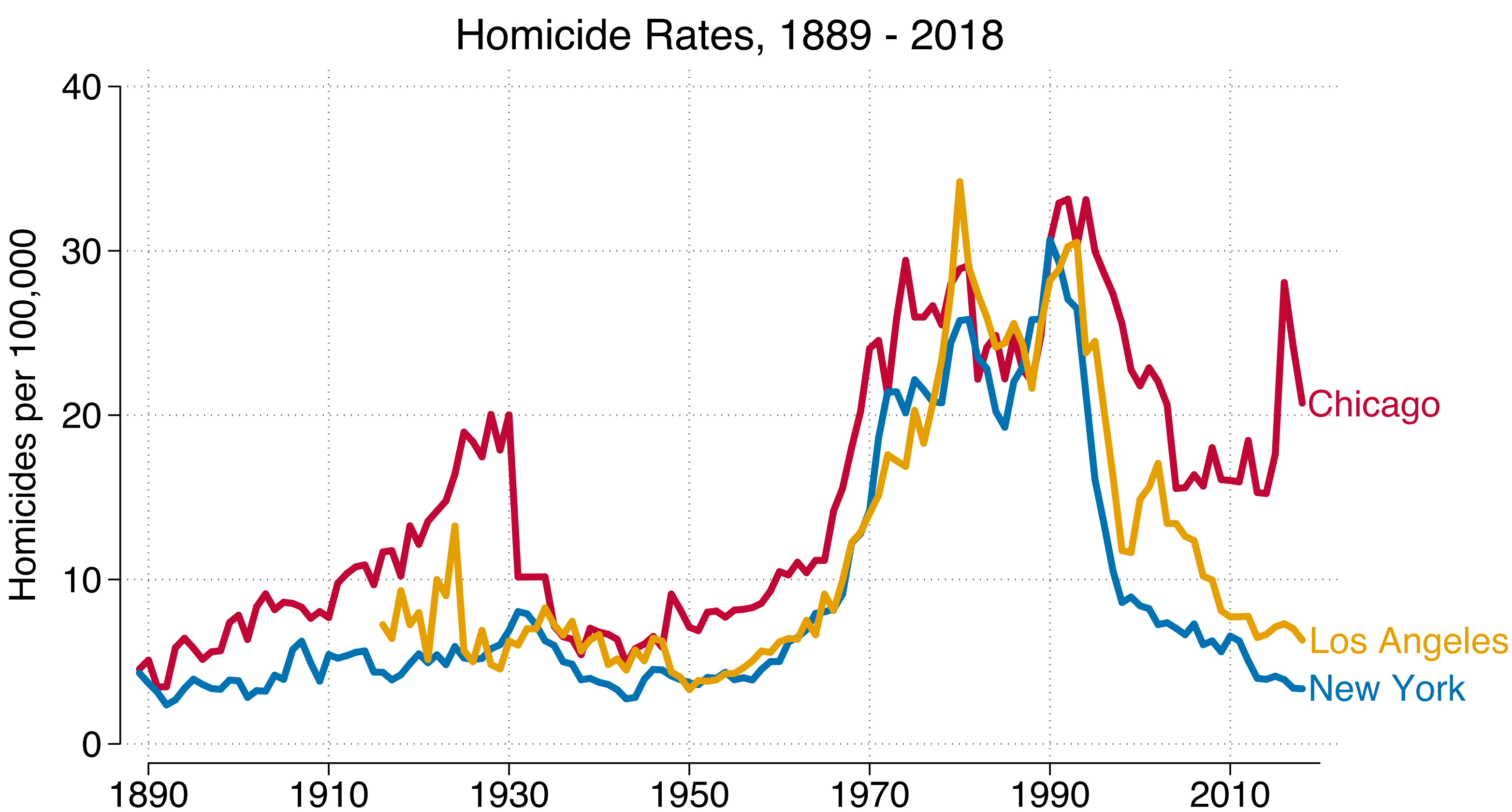
Notes: CPD data.

Figure 4. Gun and non-gun homicide rates among large American cities, 2016



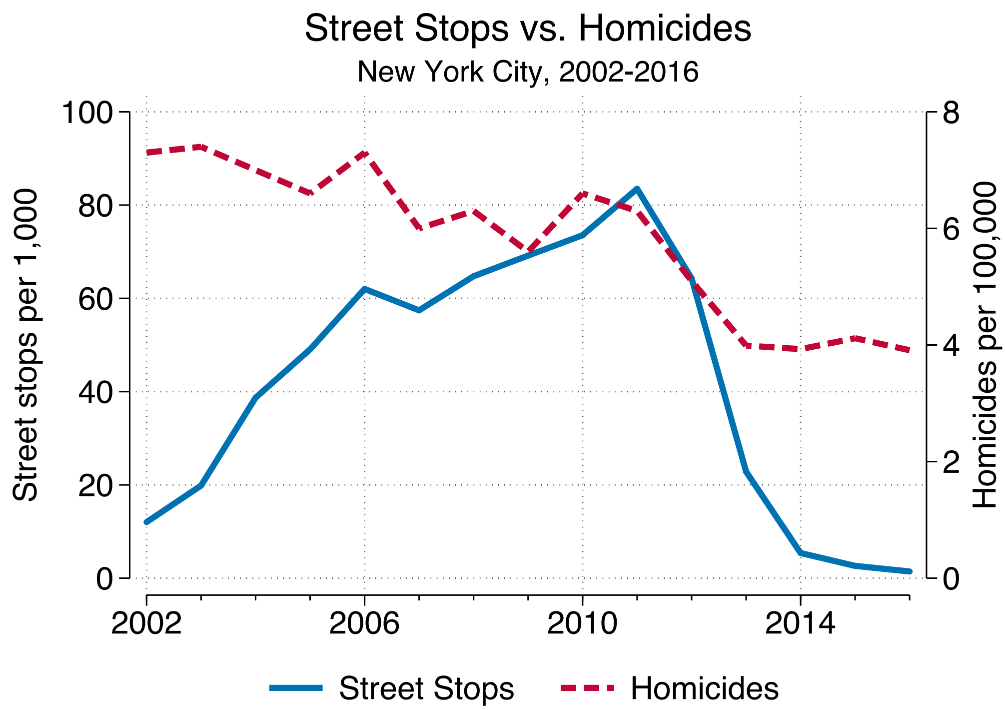
Notes: Records provided by the police departments of Chicago, New York City, Los Angeles, Houston, and Philadelphia.

Figure 5. Long-run homicide rates in New York City, Los Angeles, and Chicago, 1889 – 2018



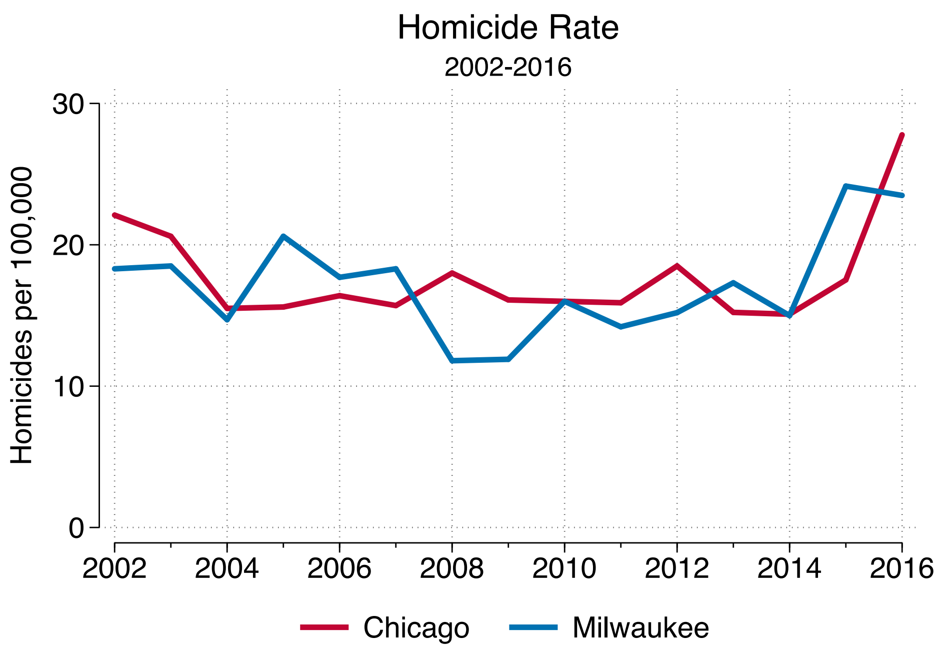
Notes: Chicago homicide data for 1889 through 1930 from the Chicago Historical Homicide Project at Northwestern University (http://homicide.northwestern.edu/). Chicago homicide data for 1930 through 1959 from the FBI's Uniform Crime Reports (ICPSR 3666). Los Angeles homicide data for 1916 through 1959 from the Historical Violence Database at the Criminal Justice Research Center, the Ohio State University (https://cjrc.osu.edu/research/interdisciplinary/hvd). New York City homicide data for 1890 through 1959 from the National Institute of Justice (ICPSR 3226). Homicide data for 1960 through 2017 from the FBI's Uniform Crime Reports (Open ICPSR). Homicide data for 2018 from the police departments of Chicago, New York City, and Los Angeles.

Figure 6. Street stops and homicides in New York City, 2002 – 2016

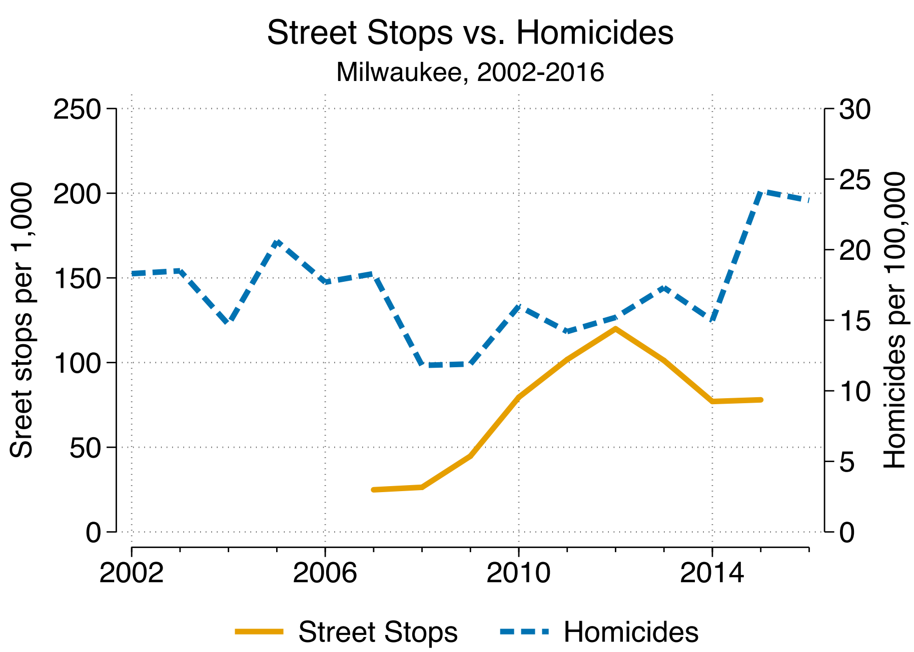


Notes: Homicide data from the FBI's Uniform Crime Reports. Stop, question, and frisk data from the New York Police Department.

Figure 7. Comparison of homicide rates in Chicago and Milwaukee, and street stop rates in Milwaukee, 2002 – 2016



Panel A



Panel B

Notes: Homicide data from the FBI's Uniform Crime Reports. Stop data from Milwaukee PD.

Figure 8. Chicago Police Department patrol areas and districts

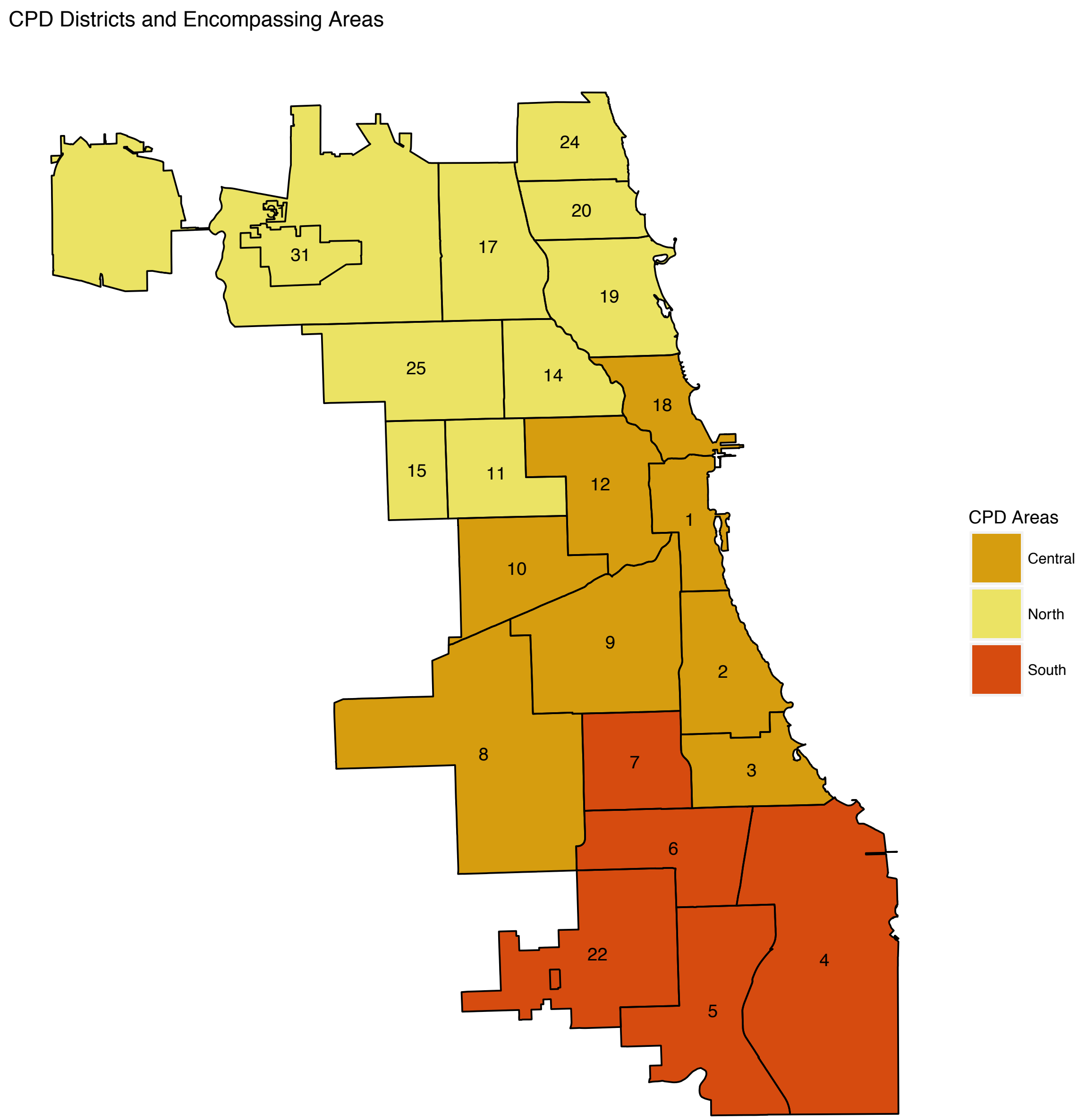


Figure 9. Sample analysis involving motor vehicle thefts from the 7th district

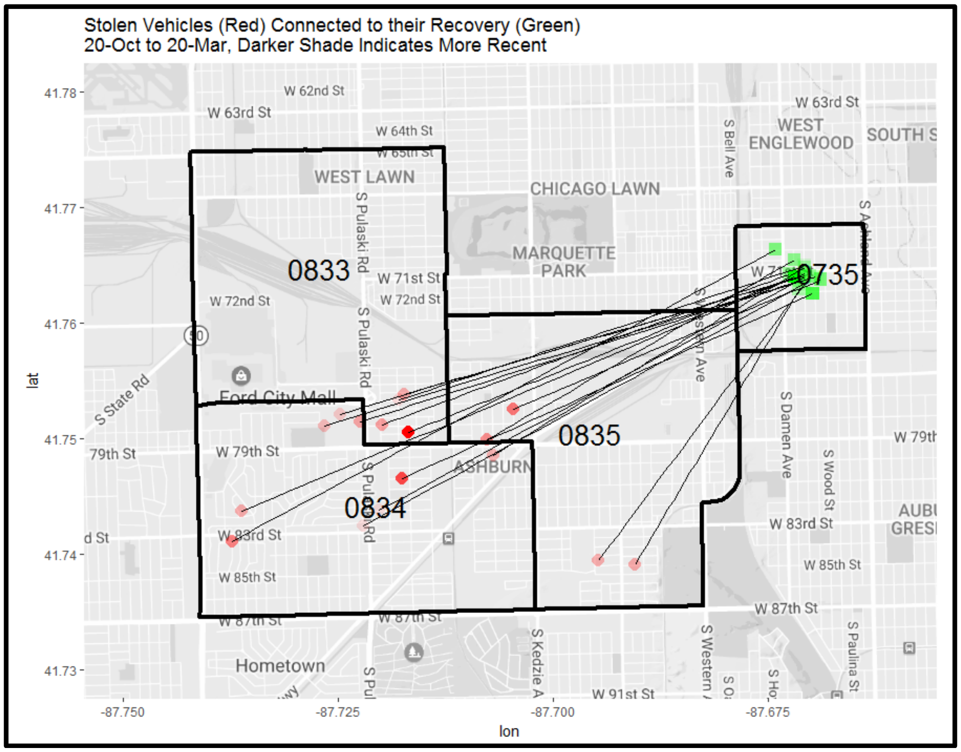
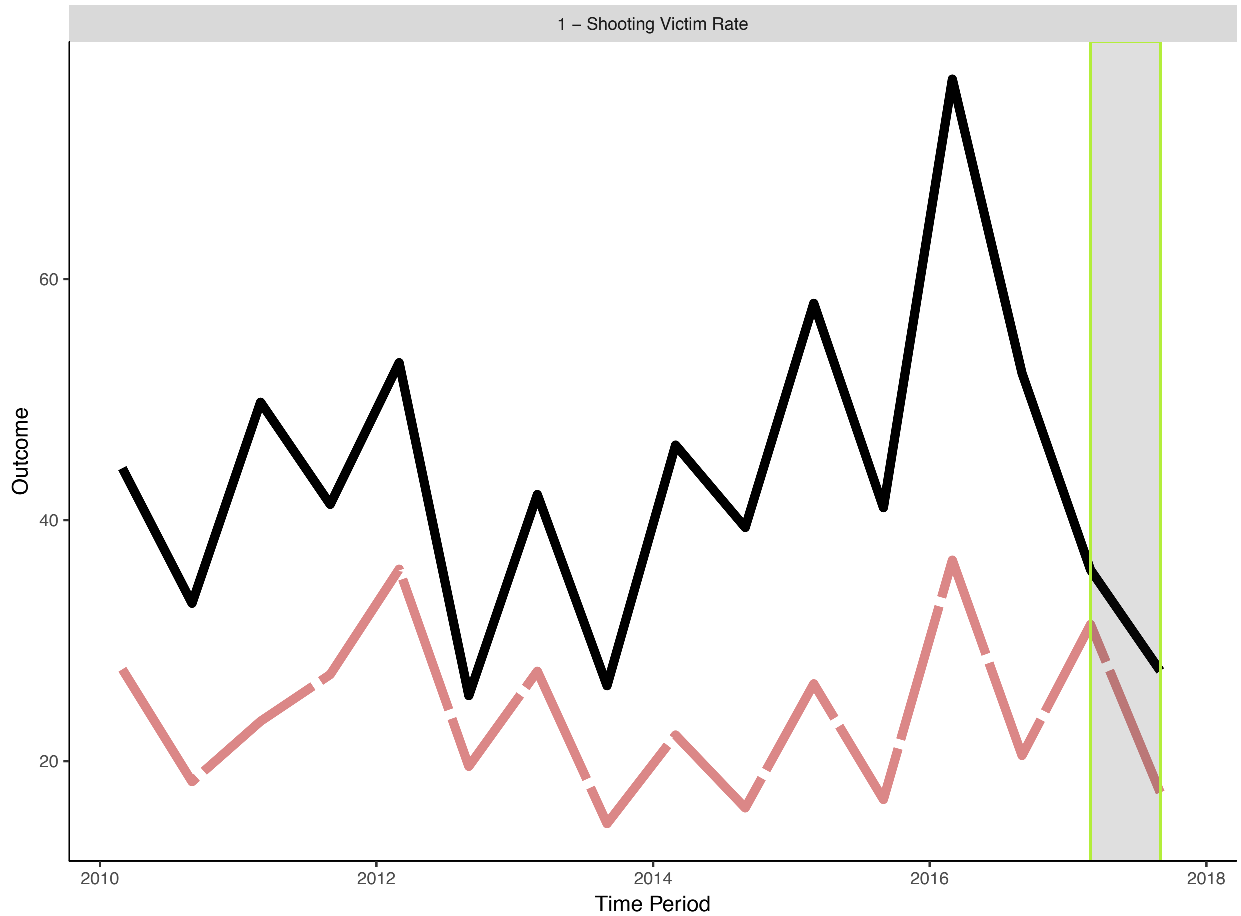
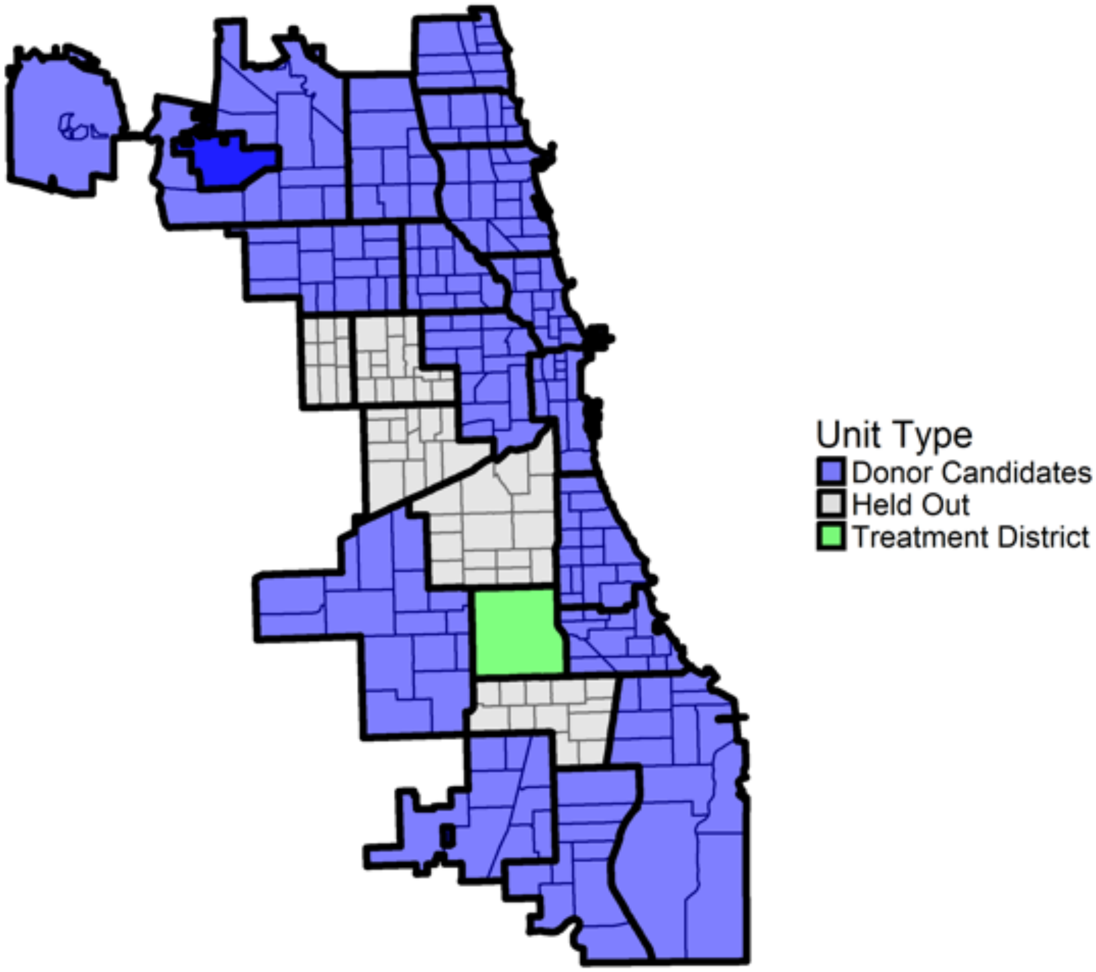


Figure 10. Shootings per capita in 7th district and ADH synthetic control, districts as donors



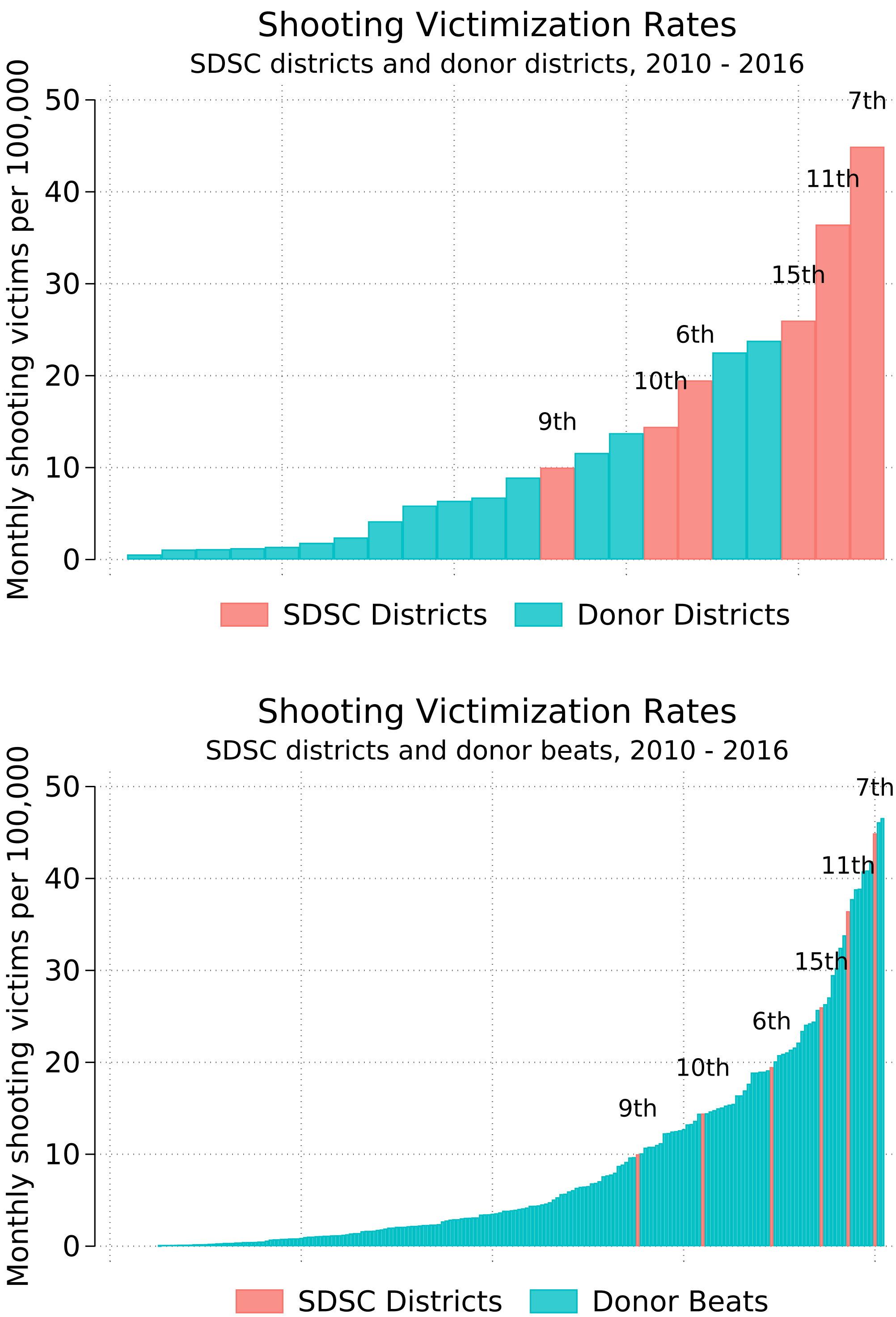
Notes: Black line denotes biannual monthly average shootings per capita in the 7th district. Red line denotes synthetic control using the ADH estimator and 16 non-SDSC districts as donors. Gray box represents post-period.

Figure 11. Chicago Police Department districts and beats



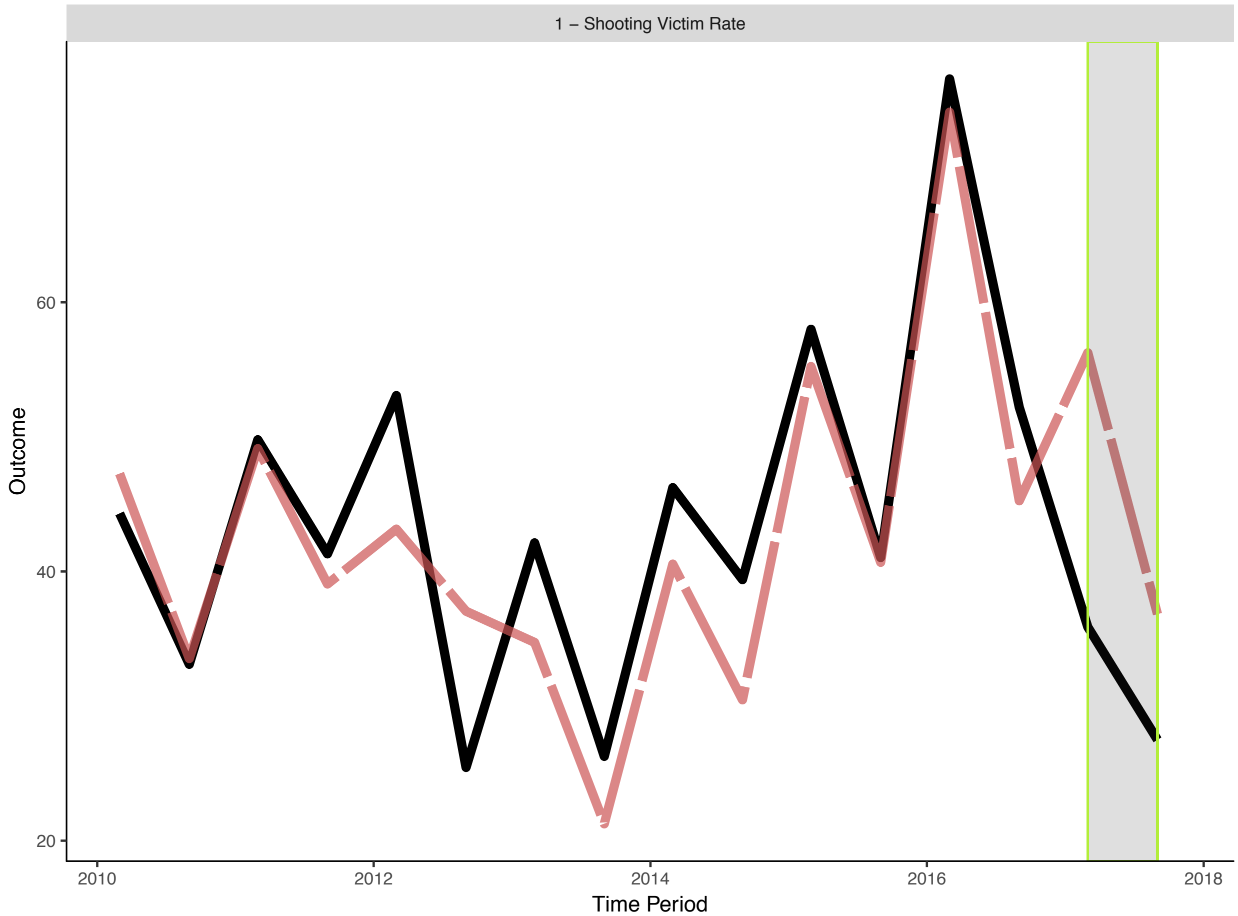
Notes: Chicago Police Department district boundaries (bold) and beat boundaries. In this example, the 7th district is shaded green; the other five SDSC districts are shaded gray and excluded from the donor pool; and the 196 beats from among the 16 non-SDSC donor districts shaded purple comprise the donor pool.

Figure 12. Distributions of monthly shooting victims per capita



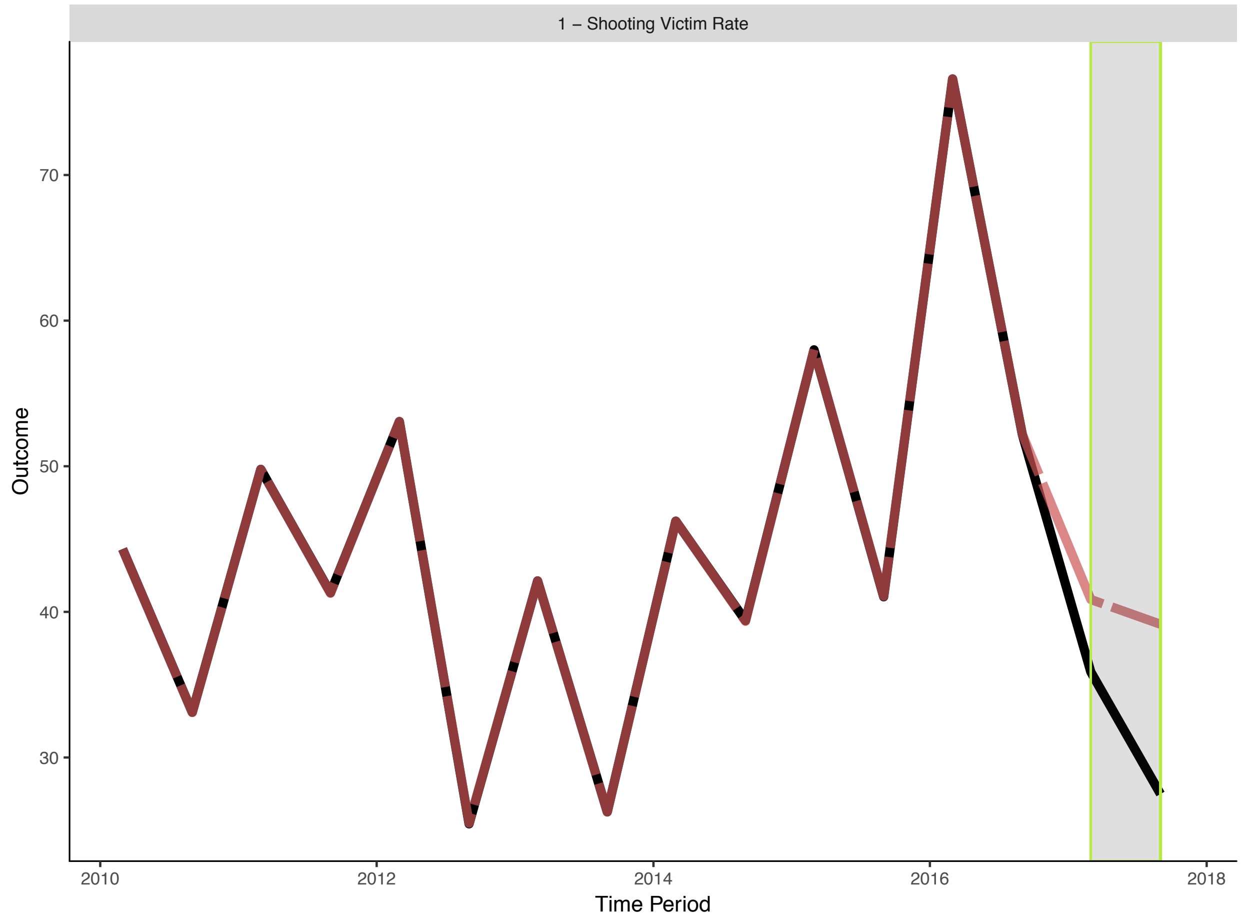
Notes: Figure shows distributions of average monthly shooting victims per 100,000 in the pre-treatment period for SDSC districts and donor districts (top panel) or donor beats (bottom panel).

Figure 13. Shootings per capita in 7th district and ADH synthetic control, beats as donors



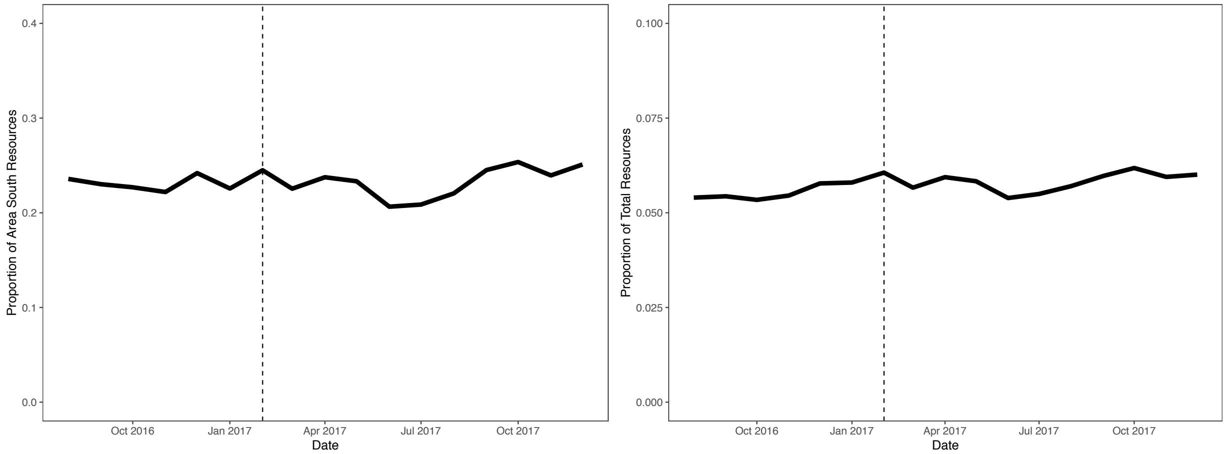
Notes: Black line denotes biannual monthly average shootings per capita in the 7th district. Red line denotes synthetic control using the ADH estimator and 196 non-SDSC beats as donors. Gray box represents post-period.

Figure 14. Shootings per capita in 7th district and DI synthetic control, beats as donors

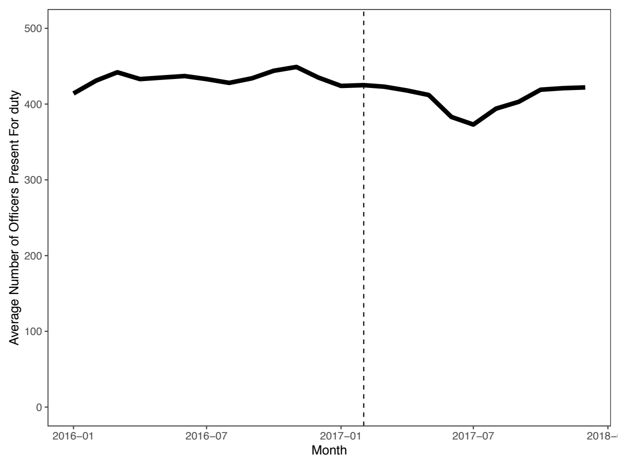


Notes: Black line denotes biannual monthly average shootings per capita in the 7th district. Red line denotes synthetic control using the DI estimator and 196 non-SDSC beats as donors. Gray box represents post-period.

Figure 15. Officer working in the 7th district



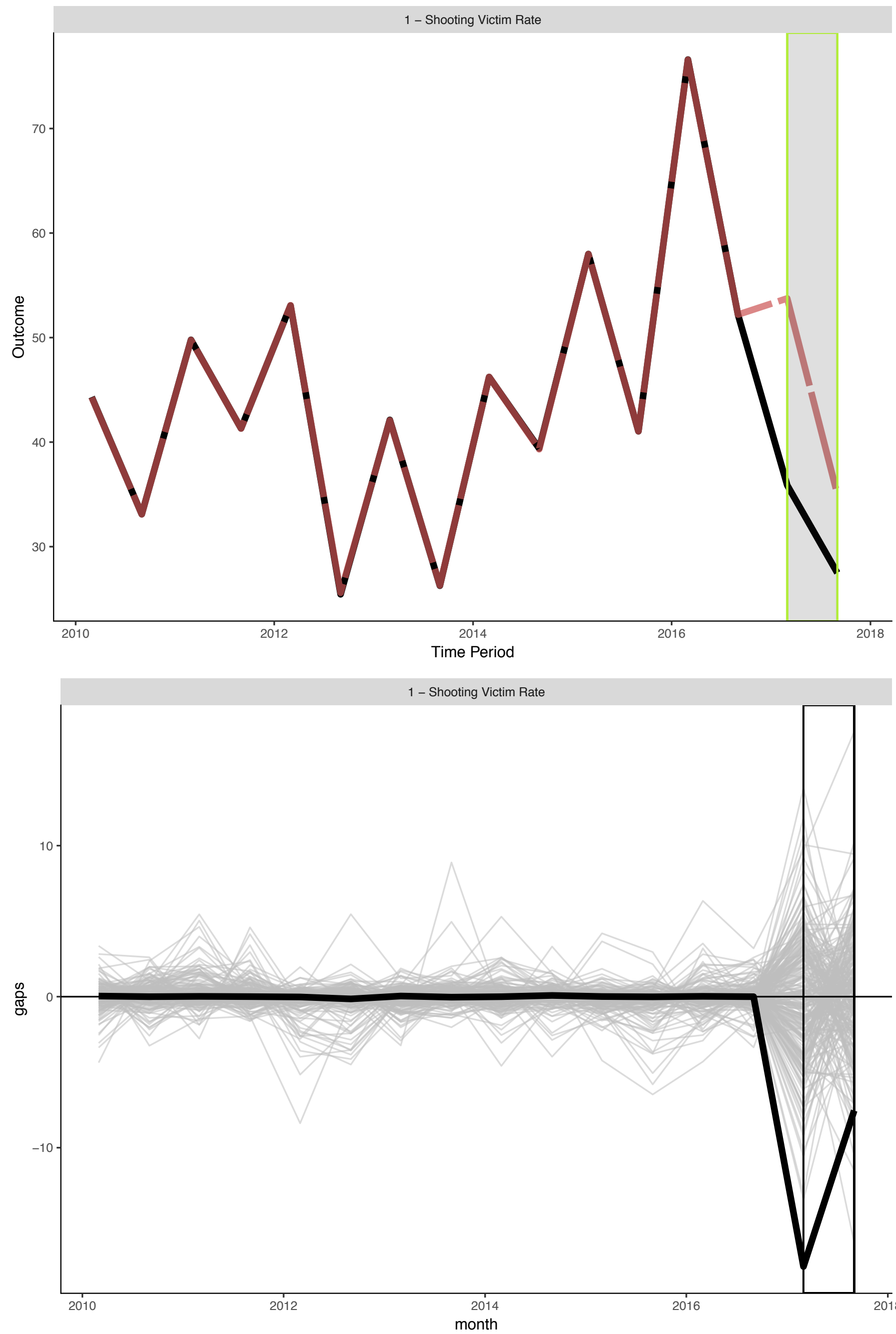
Panel A. Share of all officers in Area South (left) and citywide (right) in the 7th district, GPS data



Panel B. Number of officers present for duty in the 7th district, attendance roster data

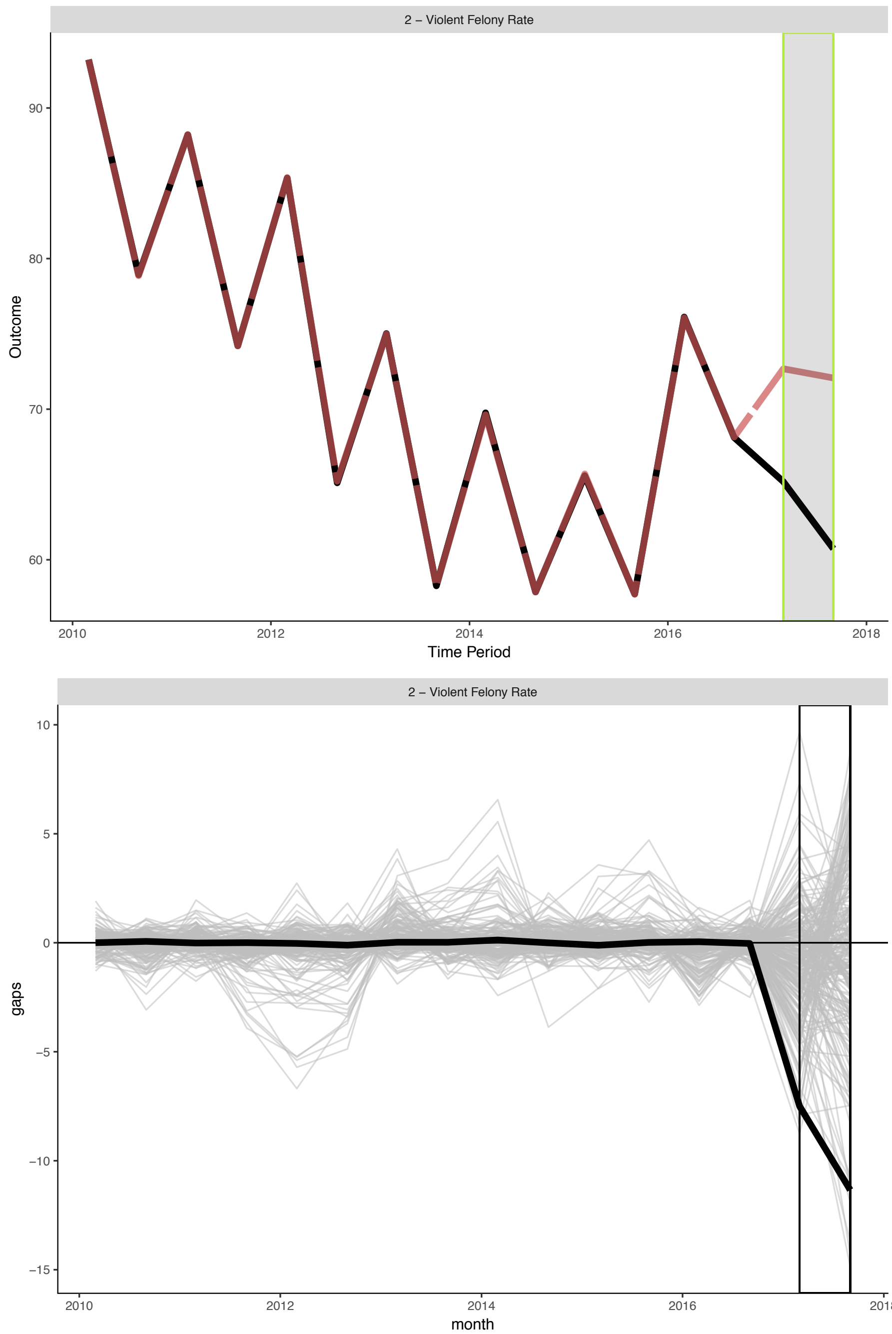
Notes: Figures in Panel A show the share of officers (at the Area level and citywide) operating in the 7th district each month, based on data from GPS units located in their vehicles. As a robustness check, the figure in Panel B relies on data from attendance rosters to record the number of unique CPD employees reporting to work in the 7th district.

Figure 16. Synthetic control results for the 7th district: shooting victims



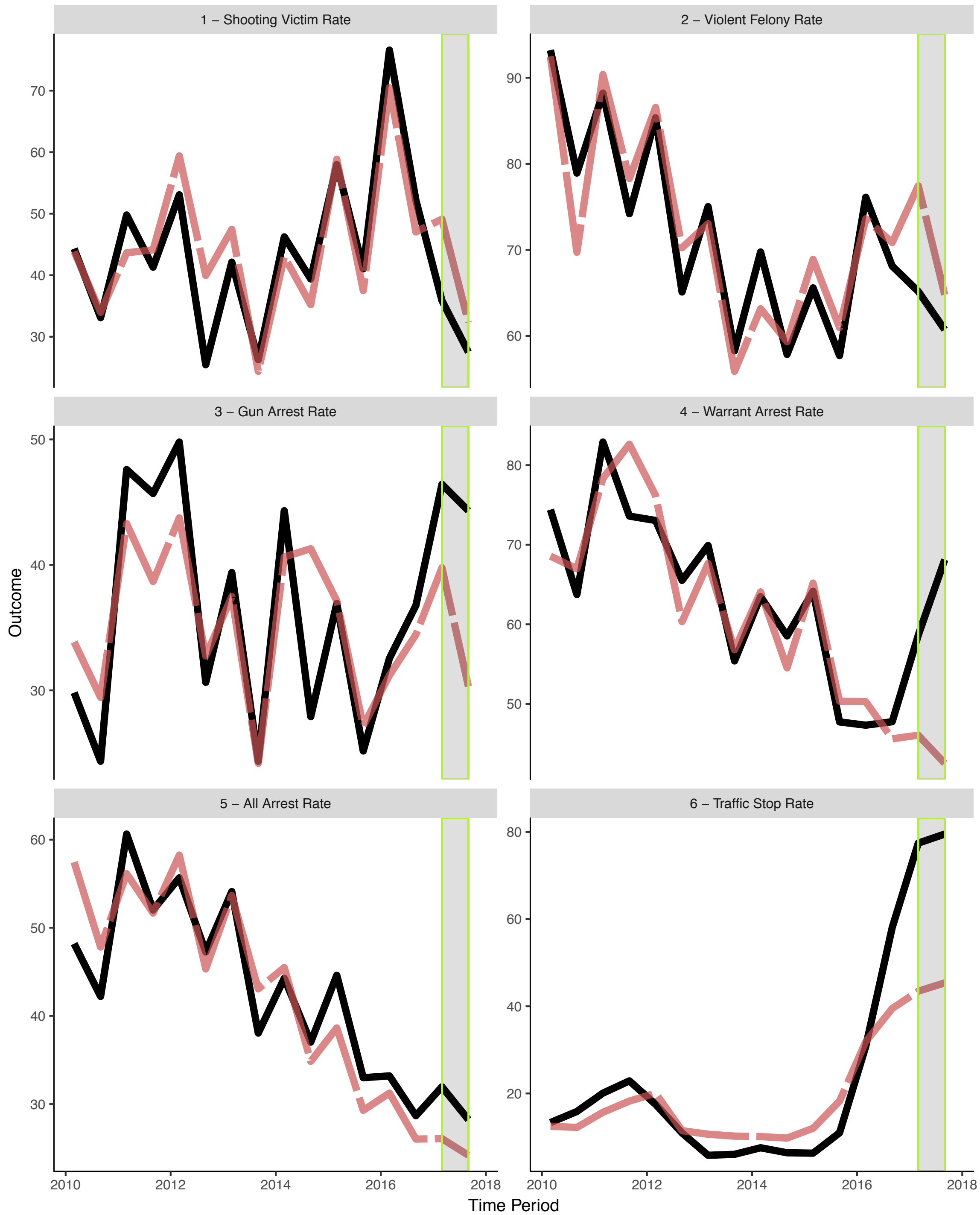
Notes: Synthetic control for the 7th district, using modified DI estimator and beats as donors, in red. Placebo test results plot differences between the actual and synthetic control unit, for the 7th district (bold line) and the resampled donor districts (light gray lines).

Figure 17. Synthetic control results for the 7th district: Part I violent felonies



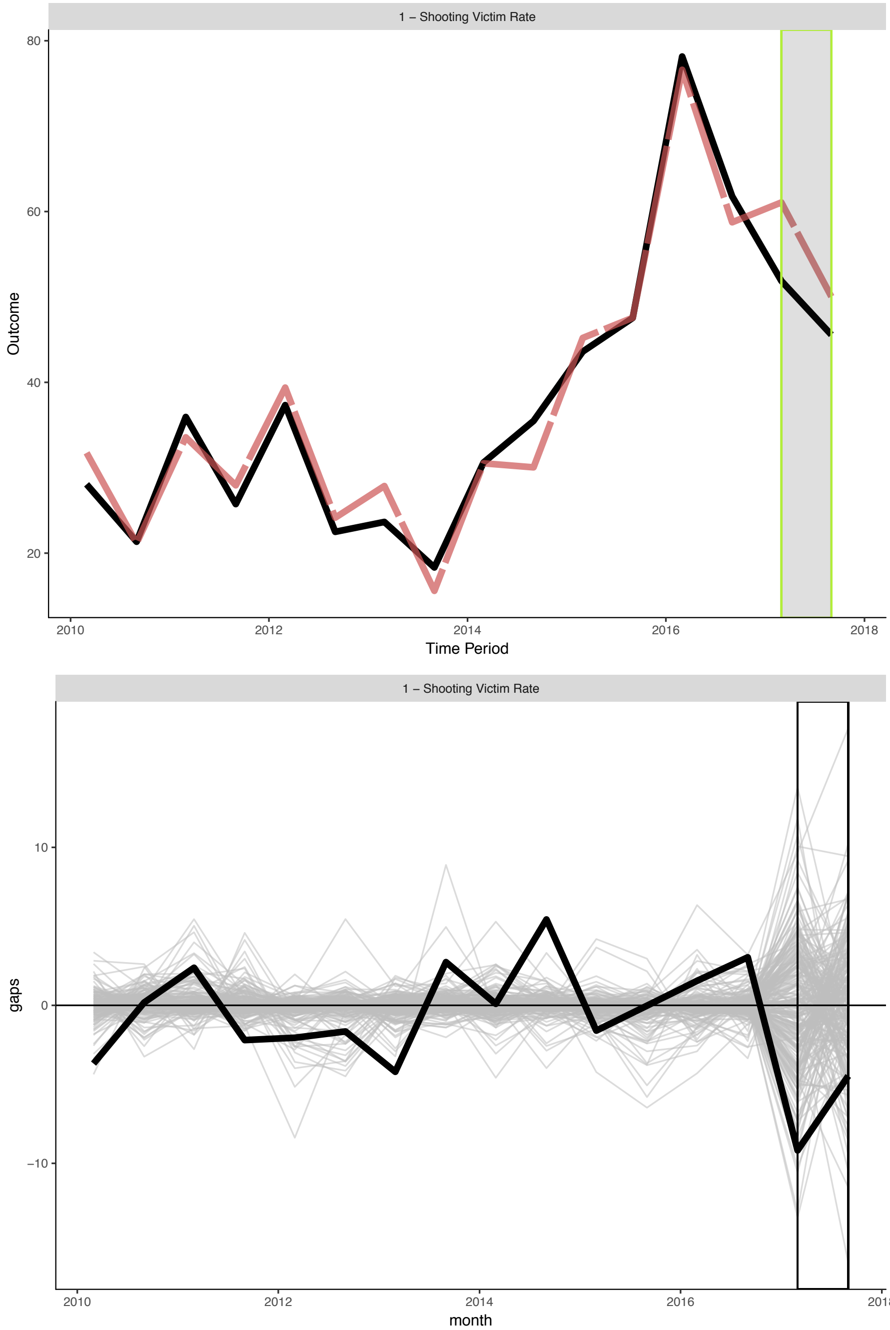
Notes: Synthetic control for the 7th district, using modified DI estimator and beats as donors, in red. Placebo test results plot differences between the actual and synthetic control unit, for the 7th district (bold line) and the resampled donor districts (light gray lines).

Figure 18. Synthetic control results for the 7th district: primary and secondary outcomes



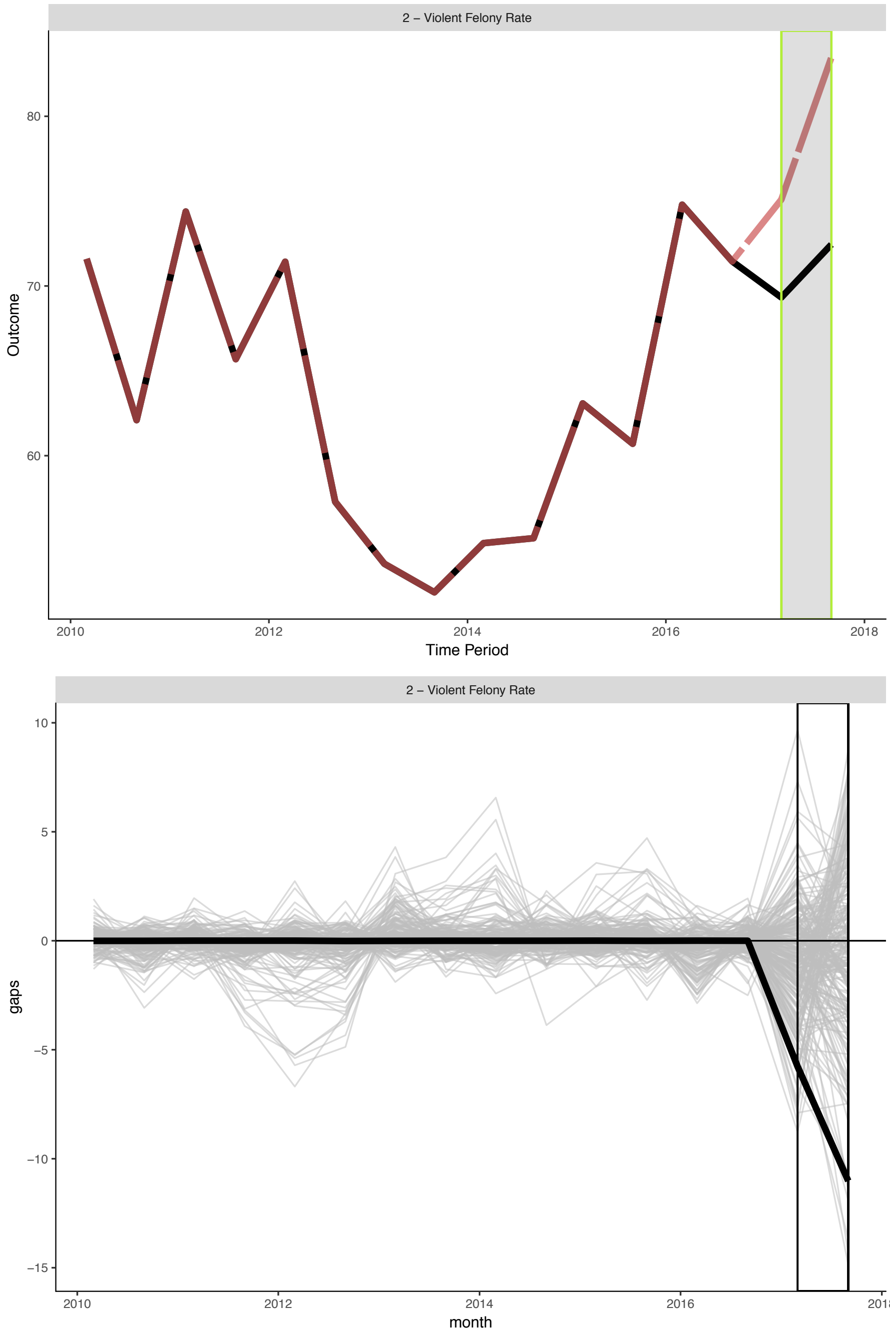
Notes: Synthetic control for the 7th district, using modified DI estimator fit to five outcomes (shooting victims per capita, Part I violent felony incidents per capita, gun arrests per capita, warrant arrests per capita, overall arrests per capita, traffic stops per capita) and beats as donors, in red.

Figure 19. Synthetic control results for the 11th district: shooting victims



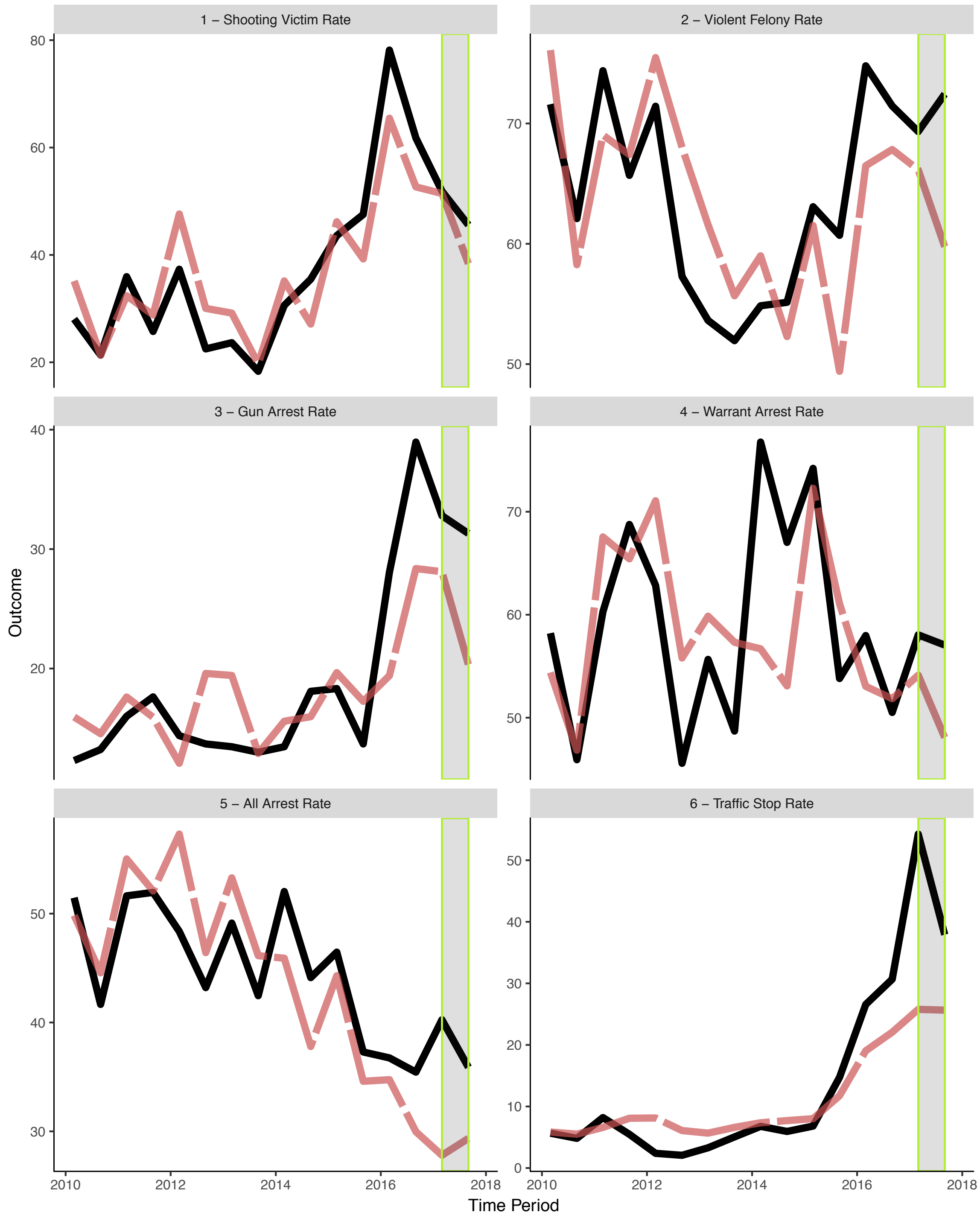
Notes: Synthetic control for the 11th district, using modified DI estimator and beats as donors, in red. Placebo test results plot differences between the actual and synthetic control unit, for the 11th district (bold line) and the resampled donor districts (light gray lines).

Figure 20. Synthetic control results for the 11th district: Part I violent felonies



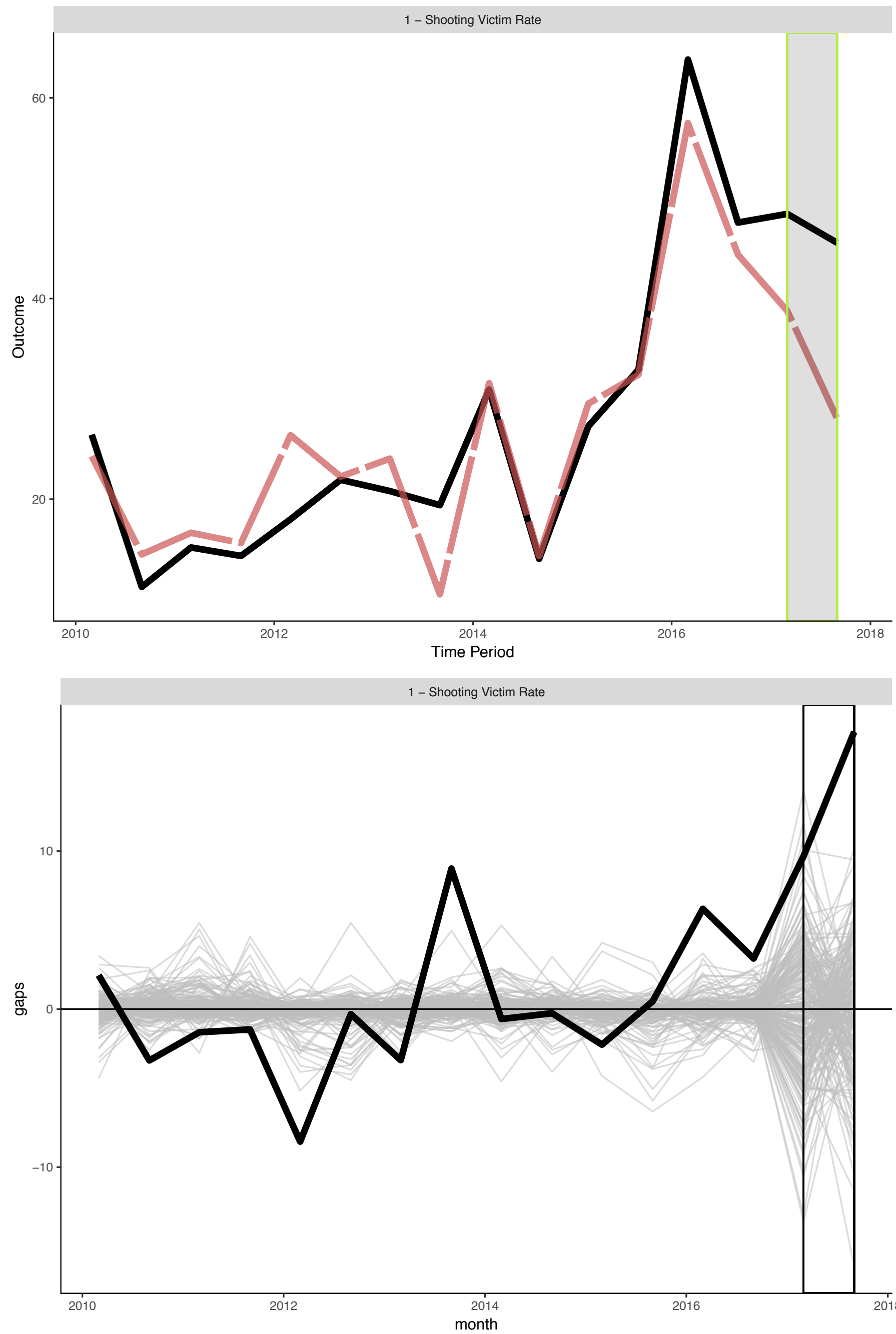
Notes: Synthetic control for the 11th district, using modified DI estimator and beats as donors, in red. Placebo test results plot differences between the actual and synthetic control unit, for the 11th district (bold line) and the resampled donor districts (light gray lines).

Figure 21. Synthetic control results for the 11th district: primary and secondary outcomes



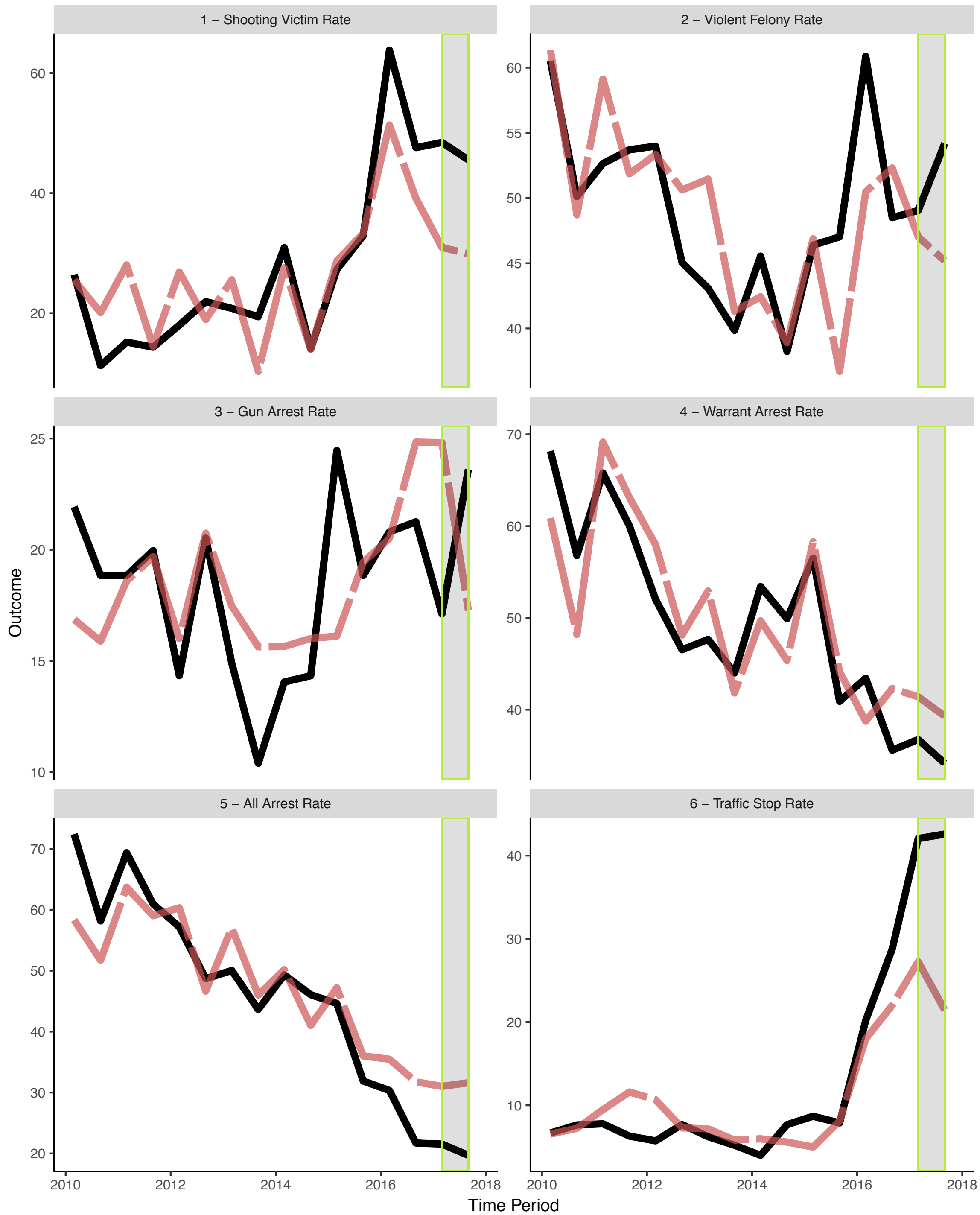
Notes: Synthetic control for the 11th district, using modified DI estimator fit to five outcomes (shooting victims per capita, Part I violent felony incidents per capita, gun arrests per capita, warrant arrests per capita, overall arrests per capita, traffic stops per capita) and beats as donors, in red.

Figure 22. Synthetic control results for the 15th district: shooting victims



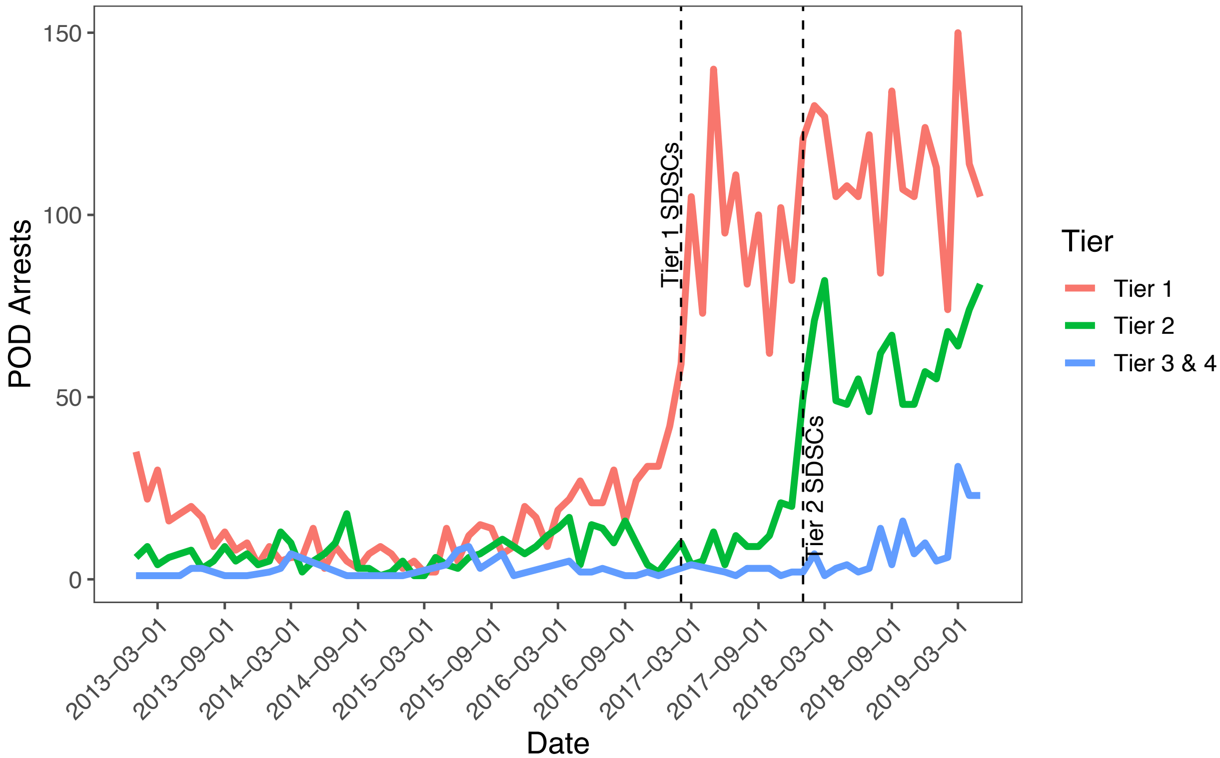
Notes: Synthetic control for the 15th district, using modified DI estimator and beats as donors, in red. Placebo test results plot differences between the actual and synthetic control unit, for the 15th district (bold line) and the resampled donor districts (light gray lines).

Figure 23. Synthetic control results for the 15th district: primary and secondary outcomes



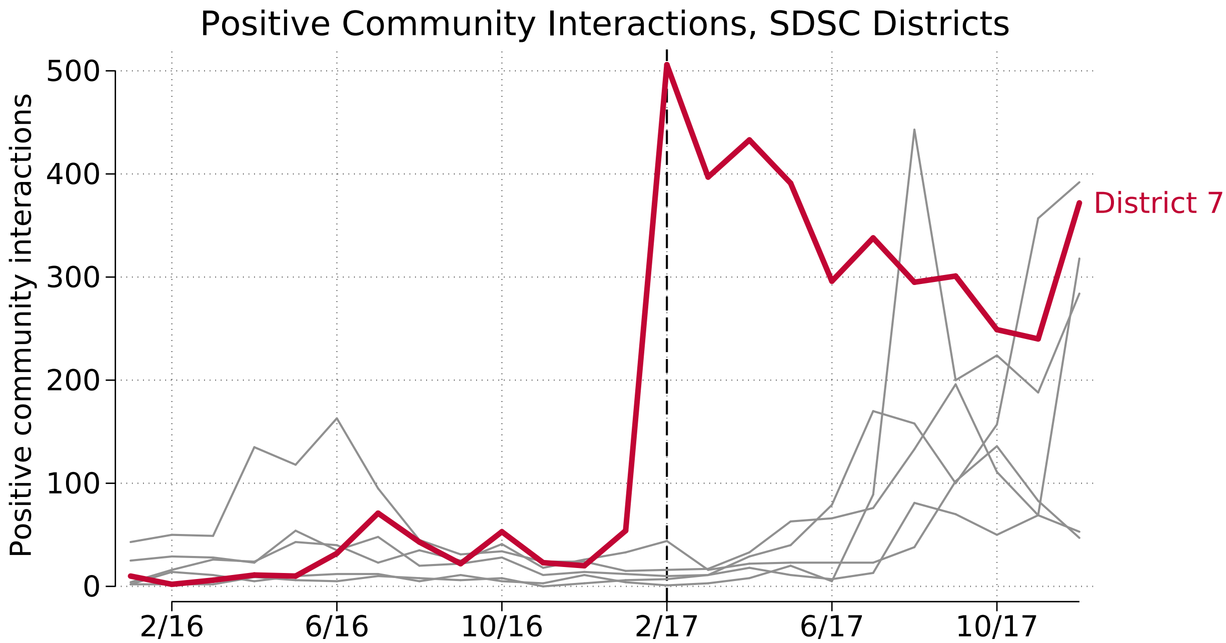
Notes: Synthetic control for the 15th district, using modified DI estimator fit to five outcomes (shooting victims per capita, Part I violent felony incidents per capita, gun arrests per capita, warrant arrests per capita, overall arrests per capita, traffic stops per capita) and beats as donors, in red.

Figure 24. POD camera-initiated arrests



Notes: Arrests initiated by officers monitoring POD cameras, by district Tier. SDSCs were first introduced in early 2017 in the six Tier 1 districts, followed by the seven Tier 2 districts in early 2018.

Figure 25. Positive community interactions: 7th district vs. other SDSC districts



Notes: Monthly volume of positive community interaction (PCI) calls by officers in Tier 1 districts.

Table 1. Synthetic control results for the 7th district

Outcome	Effect Estimate	*p*-value	*q*-value: FWER Control	*q*-value: FDR Control
	(1)	(2)	(3)	(4)
Synthetic controls estimated separately				
Shooting Victims	-30%	0.000	0.000	0.000
Part I Violent Felonies	-12%	0.004	0.045	0.050
Synthetic control estimated jointly				
Shooting Victims	-21%	0.008	0.268	0.244
Part I Violent Felonies	-12%	0.008	0.268	0.244
Gun Arrests	27%	0.102	1.000	1.000
Warrant Arrests	36%	0.041	1.000	0.672
Overall Arrests	20%	0.142	1.000	1.000
Traffic Stops	78%	0.000	0.000	0.000

Notes: Synthetic control treatment effects using modified DI estimator and beats as donors. FWER control *q*-values calculated using the Holm-Bonferroni correction. FDR control *q*-values calculated using the Benjamini-Yekutieli correction.

Table 2. Synthetic control results for the 11th district

Outcome	Effect Estimate	p-value	q-value: FWER Control	q-value: FDR Control
	(1)	(2)	(3)	(4)
Synthetic controls estimated separately				
Shooting Victims	-13%	0.049	0.390	0.363
Part I Violent Felonies	-10%	0.020	0.183	0.189
Synthetic control estimated jointly				
Shooting Victims	6%	0.187	1.000	1.000
Part I Violent Felonies	11%	0.024	0.732	0.458
Gun Arrests	28%	0.240	1.000	1.000
Warrant Arrests	11%	0.289	1.000	1.000
Overall Arrests	37%	0.024	0.732	0.458
Traffic Stops	90%	0.004	0.138	0.204

Notes: Synthetic control treatment effects using modified DI estimator and beats as donors. FWER control q-values calculated using the Holm-Bonferroni correction. FDR control q-values calculated using the Benjamini-Yekutieli correction.

Table 3. Synthetic control results for the 6th district

Outcome	Effect Estimate	*p*-value	*q*-value FWER Control	*q*-value FDR Control
	(1)	(2)	(3)	(4)
Synthetic controls estimated separately				
Shooting Victims	2%	0.585	1.000	1.000
Part I Violent Felonies	1%	0.663	1.000	1.000
Synthetic control estimated jointly				
Shooting Victims	12%	0.431	1.000	1.000
Part I Violent Felonies	-1%	0.813	1.000	1.000
Gun Arrests	7%	0.280	1.000	1.000
Warrant Arrests	0%	0.508	1.000	1.000
Overall Arrests	13%	0.252	1.000	1.000
Traffic Stops	30%	0.118	1.000	1.000

Notes: Synthetic control treatment effects using modified DI estimator and beats as donors. FWER control *q*-values calculated using the Holm-Bonferroni correction. FDR control *q*-values calculated using the Benjamini-Yekutieli correction.

Table 4. Synthetic control results for the 9th district

Outcome	Effect Estimate	*p*-value	*q*-value FWER Control	*q*-value FDR Control
	(1)	(2)	(3)	(4)
Synthetic controls estimated separately				
Shooting Victims	-9%	0.321	1.000	1.000
Part I Violent Felonies	-23%	0.077	0.463	0.411
Synthetic control estimated jointly				
Shooting Victims	-19%	0.370	1.000	1.000
Part I Violent Felonies	-6%	0.805	1.000	1.000
Gun Arrests	-42%	0.402	1.000	1.000
Warrant Arrests	11%	0.715	1.000	1.000
Overall Arrests	28%	0.325	1.000	1.000
Traffic Stops	-37%	0.167	1.000	1.000

Notes: Synthetic control treatment effects using modified DI estimator and beats as donors. FWER control *q*-values calculated using the Holm-Bonferroni correction. FDR control *q*-values calculated using the Benjamini-Yekutieli correction.

Table 5. Synthetic control results for the 10th district

Outcome	Effect Estimate	p-value	q-value FWER Control	q-value FDR Control
	(1)	(2)	(3)	(4)
Synthetic controls estimated separately				
Shooting Victims	38%	0.065	0.455	0.404
Part I Violent Felonies	-2%	0.837	1.000	1.000
Synthetic control estimated jointly				
Shooting Victims	26%	0.106	1.000	1.000
Part I Violent Felonies	0%	0.996	1.000	1.000
Gun Arrests	58%	0.154	1.000	1.000
Warrant Arrests	-5%	0.744	1.000	1.000
Overall Arrests	9%	0.390	1.000	1.000
Traffic Stops	-30%	0.215	1.000	1.000

Notes: Synthetic control treatment effects using modified DI estimator and beats as donors. FWER control q-values calculated using the Holm-Bonferroni correction. FDR control q-values calculated using the Benjamini-Yekutieli correction.

Table 6. Synthetic control results for the 15th district

Outcome	Effect Estimate	*p*-value	*q*-value FWER Control	*q*-value FDR Control
	(1)	(2)	(3)	(4)
		Synthetic controls estimated separately		
Shooting Victims	36%	0.004	0.045	0.050
Part I Violent Felonies	-3%	0.232	1.000	1.000
		Synthetic control estimated jointly		
Shooting Victims	52%	0.000	0.000	0.000
Part I Violent Felonies	11%	0.106	1.000	1.000
Gun Arrests	-12%	0.260	1.000	1.000
Warrant Arrests	-14%	0.293	1.000	1.000
Overall Arrests	-31%	0.045	1.000	0.672
Traffic Stops	67%	0.012	0.378	0.305

Notes: Synthetic control treatment effects using modified DI estimator and beats as donors. FWER control *q*-values calculated using the Holm-Bonferroni correction. FDR control *q*-values calculated using the Benjamini-Yekutieli correction.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American statistical Association* 105(490): 493-505.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2015. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59(2): 495-510.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The economic costs of conflict: A case study of the Basque Country." *American Economic Review* 93(1): 113-132.

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind. 2017. "What can we learn from experiments? Understanding the threats to the scalability of experimental results." *American Economic Review* 107(5): 282-286.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2019. "Synthetic Difference in Differences." National Bureau of Economic Research Working Paper Series No. 25532.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2018. "Matrix completion methods for causal panel data models." National Bureau of Economic Research Working Paper Series No. 25132.

Banerjee, Abhijit, Rukmini Banerji, James Berry, Esther Duflo, Harini Kannan, Shobhini Mukerji, Marc Shotland, and Michael Walton. 2017. "From proof of concept to scalable policies: challenges and solutions, with an application." *Journal of Economic Perspectives* 31(4): 73-102.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2018. "The augmented synthetic control method." arXiv 1811.04170.

Benjamini, Yoav, and Daniel Yekutieli. 2001. "The control of the false discovery rate in multiple testing under dependency." *Annals of Statistics* 29: 1165-1188.

Bertrand, Marianne, Monica Bhatt, Christopher Blattman, Sara Heller, and Max Kapustin. "Reducing Gun Violence in Chicago: Experimental Results from the READI Program."

Billmeier, Andreas, and Tommaso Nannicini. 2013. "Assessing economic liberalization episodes: A synthetic control approach." *Review of Economics and Statistics* 95.3: 983-1001.

Blattman, Christopher, Donald P. Green, Daniel Ortega, and Santiago Tobon. 2019. "Place Based Interventions at Scale: The Direct and Spillover Effects of Policing and City Services on Crime." SSRN.

Bloom, Nicholas, and John Van Reenen. 2007. "Measuring and Explaining Management Practices Across Firms and Countries." *Quarterly Journal of Economics* 122(4): 1351-1408.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics* 128(1): 1-51.

Braga, Anthony, Andrew V. Papachristos, and David Hureau. 2012. "Hot spots policing effects on crime." *Campbell Systematic Reviews* 8(8): 1-96.

Bratton, William J., and Jon Murad. 2018. "Precision Policing: A Strategy for the Challenges of 21st Century Law Enforcement." In *Urban Policy 2018*. Manhattan Institute.

Cassell, Paul G., and Richard Fowles. 2018. "What Caused the 2016 Chicago Homicide Spike: an Empirical Examination of the ACLU Effect and the Role of Stop and Frisks in Preventing Gun Violence." *U. Ill. L. Rev.*: 1581.

Chalfin, Aaron, and Justin McCrary. 2018. "Are US cities underpoliced? Theory and evidence." *Review of Economics and Statistics* 100(1): 167-186.

Cook, Philip J., and Jens Ludwig. 2000. *Gun Violence: The Real Costs*. Oxford University Press on Demand.

Davis, Jonathan, Jonathan Guryan, Kelly Hallberg, and Jens Ludwig. 2017. "The economics of scale-up." National Bureau of Economic Research Working Paper Series No. 23925. No. w23925

Di Tella, Rafael, and Ernesto Schargrodsky. 2004. "Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack." *American Economic Review* 94(1): 115-133.

Doudchenko, Nikolay, and Guido W. Imbens. 2017. "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." arXiv 1610.07748.

Draca, Mirko, Stephen Machin, and Robert Witt. 2011. "Panic on the Streets of London: Police, Crime, and the July 2005 Terror Attacks." *American Economic Review* 101(5): 2157-2181.

Durlauf, Steven N., and Daniel S. Nagin. 2011. "Imprisonment and crime: Can both be reduced?" *Criminology & Public Policy* 10(1): 13-54.

Evans, William N., and Emily G. Owens. 2007. "COPS and Crime." *Journal of Public Economics* 91(1-2): 181-201.

Goodman-Bacon, Andrew. 2019. “Difference-in-Differences with Variation in Treatment Timing.” National Bureau of Economic Research Working Paper Series No. 25018.

Heaton, Paul. 2010. “Understanding the Effects of Antiprofiling Policies.” *The Journal of Law and Economics* 53(1): 29-64.

Hochberg, Yosef. 1988. “A Sharper Bonferroni Procedure for Multiple Tests of Significance.” *Biometrika* 75(4): 800-802.

Holm, Sture. 1979. “A Simple Sequentially Rejective Multiple Test Procedure.” *Scandinavian Journal of Statistics* 6(2): 65-70.

Imbens, Guido W., and Donald B. Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Kapustin, Max, Jens Ludwig, Marc Punkay, Kimberley Smith, Lauren Speigel, and David Welgus. 2017. “Gun Violence in Chicago, 2016.” University of Chicago Crime Lab.

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. 2018. “Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates.” Working Paper.

Klick, Jonathan, and Alexander Tabarrok. 2005. “Using terror alert levels to estimate the effect of police on crime.” *The Journal of Law and Economics* 48(1): 267-279.

Levitt, Steven D. 1997. “Using electoral cycles in police hiring to estimate the effects of police on crime.” *American Economic Review* 87(3): 270-290.

Levitt, Steven D. 2002. “Using electoral cycles in police hiring to estimate the effects of police on crime: Reply.” *American Economic Review* 92(4): 1244-1250.

Machin, Stephen, and Olivier Marie. 2011. “Crime and police resources: The street crime initiative.” *Journal of the European Economic Association* 9(4): 678-701.

Mello, Steven. 2019. “More COPS, less crime.” *Journal of Public Economics* 172: 174-200.

Muralidharan, Karthik, and Paul Niehaus. 2017. “Experimentation at scale.” *Journal of Economic Perspectives* 31(4): 103-24.

National Academies of Sciences, Engineering, and Medicine. 2018. *Proactive Policing: Effects on Crime and Communities*. Washington, DC: The National Academies Press.

Owens, Emily G. 2013. “1 COPS and Cuffs.” *Lessons from the Economics of Crime: What Reduces Offending?*: 17.

Papachristos, Andrew V., Tracey L. Meares, and Jeffrey Fagan. 2007. "Attention felons: Evaluating project safe neighborhoods in Chicago." *Journal of Empirical Legal Studies* 4(2): 223-272.

Powell, David. 2018. "Imperfect Synthetic Controls: Did the Massachusetts Health Care Reform Save Lives?" RAND Working Paper.

Robbins, Michael W., Jessica Saunders, and Beau Kilmer. 2017. "A framework for synthetic control methods with high-dimensional, micro-level data: evaluating a neighborhood-specific crime intervention." *Journal of the American Statistical Association* 112(517): 109-126.

Rubin, Donald B. 1974. "Estimating causal effects of treatments in randomized and nonrandomized studies." *Journal of Educational Psychology* 66(5): 688.

Saunders, Jessica, Priscillia Hunt, and John S. Hollywood. 2016. "Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot." *Journal of Experimental Criminology* 12(3): 347-371.

Schochet, Peter Z. 2009. "Statistical power for regression discontinuity designs in education evaluations." *Journal of Educational and Behavioral Statistics* 34(2): 238-266.

Shi, Lan. 2009. "The limit of oversight in policing: Evidence from the 2001 Cincinnati riot." *Journal of Public Economics* 93(1-2): 99-113.

Sullivan, Christopher M., and Zachary P. O'Keeffe. 2017. "Evidence that curtailing proactive policing can reduce major crime." *Nature human behaviour* 1(10): 730.

Farrington, David P., and Brandon C. Welsh. 2008. *Saving children from a life of crime: Early risk factors and effective interventions*. Oxford University Press.

Weisburd, David, and John E. Eck. 2004. "What Can Police Do to Reduce Crime, Disorder, and Fear?" *Annals of the American Academy of Political and Social Science* 593(1): 42-65.

Wing, Coady, and Thomas D. Cook. 2013. "Strengthening the regression discontinuity design using additional design elements: A within-study comparison." *Journal of Policy Analysis and Management* 32(4): 853-877.

Wolfgang, Marvin, Robert M. Figlio, and Thorstin Sellin. 1972. *Delinquency in a Birth Cohort*. Chicago: University of Chicago Press.