

IMPACT EVALUATION OF THE LAPD COMMUNITY SAFETY PARTNERSHIP

BY SYDNEY KAHMANN^{1,a}, ERIN HARTMAN^{2,b}, JORJA LEAP^{3,c} AND
P. JEFFREY BRANTINGHAM^{4,d}

¹*Department of Statistics, University of California, Los Angeles, skahmann@ucla.edu*

²*Department of Political Science, University of California, Berkeley, ekhartman@berkeley.edu*

³*Department of Social Welfare, University of California, Los Angeles, jleap@ucla.edu*

⁴*Department of Anthropology, University of California, Los Angeles, branting@ucla.edu*

In 2011, the Los Angeles Police Department (LAPD), in conjunction with other governmental and nonprofit groups, launched the Community Safety Partnership (CSP) in several public housing developments in Los Angeles. Following a relationship-based policing model, officers were assigned to work collaboratively with community members to reduce crime and build trust. However, evaluating the causal impact of this policy intervention is difficult, given the notable differences between communities where CSP was implemented and the surrounding communities in South Los Angeles. In this paper we use a novel data set, based on the LAPD's reported crime incidents and calls-for-service, to evaluate the effectiveness of this program via augmented synthetic control models, a cutting-edge method for policy evaluation. We perform falsification analyses to evaluate the robustness of the results. In the public housing developments where it was first deployed, we find that CSP exhibited modest but statistically insignificant reductions in reported violent crime incidents, shots fired and violent crime calls-for-service, and Part I reported crime incidents. We do not find evidence of crime displacement from CSP regions to neighboring control regions.

1. Introduction. In 2011, the Los Angeles Police Department (LAPD) partnered with the Housing Authority of the City of Los Angeles and the Urban Peace Institute to launch the Community Safety Partnership (CSP) (Rice and Lee (2015)). CSP was designed to address high levels of violent crime and the corrosive effects of multigenerational violent gangs entrenched in several Los Angeles public housing developments (PHDs). In Jordan Downs, Nickerson Gardens, and Imperial Courts, located in the Watts neighborhood of South Los Angeles, rivalries between gangs in the PHDs brought near daily conflict to the communities and allowed violent and property crime to thrive.

At its core, CSP was a rethinking of how to approach the problems of violent crime and gangs built around a relationship-based policing model. Two decades of heavy-handed crime suppression had succeeded in cultivating widespread distrust of police but appeared to do little to blunt the control of gangs (Leap (2020), Rice and Lee (2015)). CSP saw arrest-based crime suppression as a last resort and rather sought to have “police officers and residents work in mutually respectful partnership to identify and prevent crime” (LA Office of the Mayor (2017)). Funded by the LAPD and the Housing Authority of the City of Los Angeles, CSP recruited and trained a special group of officers for five-year assignments in the PHDs. CSP officers were tasked with “support[ing] community and youth programs, address[ing] quality of life issues and develop[ing] programs to address and reduce violent crimes” (LAPD News Release (2015)). In effect, CSP officers were to become invested members of the community rather than simply an intervening force that shows up when there are problems.

Received September 2020; revised June 2021.

Key words and phrases. Gang violence, community policing, time-series cross-sectional data, augmented synthetic control method, policy evaluation, causal inference.

The relationship-based model was expected to impact crime and disorder while also improving trust between police and the public. The rationale was that the informal social networks linking CSP officers to community members make it easier to identify the problems that matter to community members (not just the police). Informal social networks allow police and community members to try out solutions together, in near real-time, blunting the potential that small problems metastasize into serious ones. The corollary is that solving smaller problems (and avoiding serious ones) helps to build trust in those informal social relationships and, by extension, the police-public partnership. The relationship-based model contrasts with other established forms of community policing that most often rely on intermittent community meetings to identify problems which then largely fall to police and other city agencies to solve (Eck and Maguire (2000), Skogan (2006)). CSP is, therefore, akin to just-in-time adaptive interventions in health (Nahum-Shani et al. (2018)), with the adaptive flexibility in CSP provided by the relationships formed between police and community members.

The communities participating in CSP exhibited reduced violent crime and improved relations with the police almost immediately, a fact that attracted considerable media attention (e.g., Blackstone (2014), Siegler (2013), Streeter (2014)). Though encouraging, it is critical to recognize that inferring causality is challenging under the circumstances. CSP was implemented in highly targeted communities and developed at a time when Los Angeles was experiencing an unprecedented decline in crime.¹ Crime peaked in Los Angeles in 1992, a year which saw nearly 1100 murders citywide. Los Angeles then experienced nearly two decades of falling violent and property crimes while simultaneously adding a half-million new residents. At its lowest point in 2013, there were 251 homicides citywide, a 335% reduction from the earlier peak. Given the overall decrease in crime throughout the city, it is possible that crime would have fallen in CSP areas, even if the program had not existed. Given the significance surrounding policing reform in America, it is important to carefully consider the causal impact of community-policing programs such as the CSP.

To determine whether CSP actually reduced crime, we require knowledge of what crime and disorder *would have been like* in these communities had CSP not been implemented, that is, the counterfactual. Here, using panel data, we aim to evaluate the causal impact of CSP: we have access to crime outcomes measured pre- and posttreatment across South Los Angeles communities that both participated in CSP (treated units) and did not participate in CSP (control units). We consider two common approaches to estimating causal effects using panel data, including difference-in-differences and synthetic control methods. As discussed in Section 4.2.1, the parallel trends assumption required for a difference-in-differences approach is not credible, and thus not appropriate, in this application. Therefore, we rely on an alternative approach designed for our data setting, the synthetic control method (SCM) in which there are relatively many pretreatment time periods and one (or few) treated units.

Synthetic control methods, introduced in a series of seminal papers by Abadie and Gardeazabal (2003), Abadie, Diamond and Hainmueller (2010), and Abadie, Diamond and Hainmueller (2015), in effect construct a counterfactual for the treated units using a weighted combination of “donor control units,” or control units that never receive treatment. While widely used in policy evaluations, these methods have until recently been infeasible in a number of important data settings. In particular, when SCM weights are constrained to the simplex, balancing on larger numbers of pretreatment periods and incorporating additional observable covariates, this weighting becomes less feasible (Ferman and Pinto (2019)). Additionally, when treatment is implemented for more than one unit, one must find a principled

¹Figure S1 provides Los Angeles homicide counts from 1987–2017 (Federal Bureau of Investigation (2021), Blumstein, Wallman and Farrington (2006)).

way to pool the information across the SCMs (Ben-Michael, Feller and Rothstein (2019)). We follow O'Neill et al. (2016) and Robbins, Saunders and Kilmer (2017) and fully pool the treated units by considering the average treated unit.

We use a new technique, the augmented synthetic control method (ASCM) (Ben-Michael, Feller and Rothstein (2021)), to estimate the impact of CSP. ASCM is an innovative method that extends the powerful SCM approach to a data setting like ours in which we have few treated units—many pretreatment time periods with which to fit the data and possible violations of the necessary causal identifying assumptions required by the original SCM. To address violations of the traditional SCM assumptions, such as imperfect pretreatment fit of SCM weights, the ASCM uses an outcome model for bias adjustment, simultaneously minimizing the extrapolation from the control units in a principled way using regularization via ridge regression. Thus, ASCM is a cutting-edge method that allows for robust estimation of treatment effects for policies, such as CSP, which are implemented in a small number of jurisdictions where other methods may not be credible.

Using a novel data set, based on the LAPD's reported crime incidents and calls-for-service data, we evaluate the impact of CSP on crime in Nickerson Gardens, Jordan Downs, and Imperial Courts, three public housing developments in South Los Angeles, for the six-year period between 2012–2017. Specifically, we estimate the average number of violent crime incidents, Part I crime incidents, and violent crime calls-for-service prevented biannually per public housing development. Previewing our results, we find statistically insignificant reductions across our crime measures. We estimate that, on average, CSP prevented 5.51 ($p = 0.48$) violent crimes and reduced violent crime calls-for-service by 1.26 ($p = 1$) per six-month period per public housing development between January 1, 2012 and December 31, 2017 (p -values for the joint null of no effect in any posttreatment period are provided in parentheses). Additionally, we find that, on average, CSP led to an reduction of 10.36 ($p = 0.58$) Part I reported crime incidents per six-month period per housing development during this time.

The remainder of this paper proceeds as follows. In Section 2 we describe CSP and how it sought to address well-known challenges in “community policing.” In Section 3 we describe the data used in model balancing and testing. Section 4 introduces panel data methods before presenting augmented synthetic control methods and their underlying assumptions. Section 5 turns to model evaluation using placebo tests. Section 6 presents results on the impact of CSP in Jordan Downs, Nickerson Gardens, and Imperial Courts PHDs. Section 7 discusses the implications of the results for CSP and relationship-based policing more broadly.

2. Background. The Community Safety Partnership was launched in four public housing developments in Los Angeles in late 2011 (Leap (2020)). It is one part of a comprehensive approach to violence reduction in places that have long suffered under the control of powerful street gangs and the corrosive effects of heavy-handed crime suppression tactics, persistent neglect by city officials, and concentrated social and economic disadvantage (Fagan and MacDonald (2012)). Gang prevention and intervention efforts as well as broad community engagement projects are spearheaded by the Mayor's Office of Gang Reduction and Youth Development (GRYD) (Tremblay et al. (2020)). Infrastructure improvement in PHDs is the responsibility of the Housing Authority of the City of Los Angeles. CSP is a merging of these motivations, responsible for establishing and sustaining basic security and safety in the targeted communities. The approach of CSP is to concentrate on building long-lasting relationships between police and members of the community and leveraging those relationships for collaborative problem solving (Leap (2020), Rice and Lee (2015)).

CSP extends a 50-year history of community policing efforts that have all sought to correct for the many deficiencies of the professionalization model that dominated much of the 20th

century (Blumstein (2018), Goldstein (1977), Goldstein (1979), Katzenbach (1967), Kelling and Moore (1989), Kelling et al. (1981), Reisig (2010), Sherman et al. (1973)). In spite of promising early steps, community policing today is perhaps best described as a general orientation adopted by policing organizations. Indeed, the U.S. Department of Justice defines community policing as “a philosophy that promotes organizational strategies which support the systematic use of partnerships and problem-solving techniques, to proactively address the immediate conditions that give rise to public safety issues such as crime, social disorder, and fear of crime” (DOJ (2009), pg.1). The most persistent criticism of community policing is that it is too amorphous (Cordner (1997), Skogan (2006)). A wide range of policing strategies and tactics may qualify as community policing. Thus, evaluation and generalization of individual community policing efforts is challenging. In Section 7 we situate our findings in the broader community policing literature.

CSP in Los Angeles was designed with these weaknesses of past community policing efforts in mind. It is a deliberate model of police recruitment, training, deployment, strategic and tactical orientation, and command oversight (Leap (2020)). CSP started with a year of planning prior to launch in 2011.² A detailed selection process was established to recruit officers with existing orientations toward problem solving. Selected officers underwent training aimed at building understanding of the interrelated cultural, demographic, and economic factors that impact public safety in CSP sites. Training was designed and delivered by the Urban Peace Institute, a community-based civilian organization. Officers were trained on techniques for defusing community-wide dangers without overrelying on traditional suppression tactics, such as arrest.

The CSP model also recognized that alternatives to suppression require trust and a network of community relationships that could be called on to solve immediate, local problems (see also Skogan (2006)). Since building reliable social networks requires both time and stability of effort, CSP established long-term deployments for officers, lasting five-years in each community. The deployments also had a separate command structure, allowing for greater autonomy and discretion of officers. Officers were provided unique incentives (promotion and pay) to reward community-engaged behaviors not captured by traditional metrics, such as crime and arrest statistics. The ultimate goal of CSP was not only to build trust in policing but also to provide the basic security and safety necessary for normal social and economic activity. While it is clear that CSP has not been immune to many of the well-known challenges facing community policing, such as ambiguity about how to balance enforcing laws against relationship building (Leap (2020)), even partial adherence to the CSP model may be expected to have an impact. Our purpose is to evaluate whether CSP succeeded in improving basic security and safety in the PHDs where it was deployed.

3. Data. We rely on three primary sources to construct our dataset: LAPD reported crime incidents data, LAPD calls-for-service data, and U.S. Census data. Below we discuss how treated units, donor control units, and outcome and covariate data are defined.

3.1. Outcomes. Our focus is on estimating the causal impact of CSP on crime and disorder. The augmented synthetic control method we rely on requires both pre- and posttreatment measures of crime as well as pretreatment covariate data. In this section, using the LAPD reported crime incidents and calls-for-service data, we describe how we define and aggregate our crime outcomes.

²In Section S2.4 we drop this year of planning (2011) from the pretreatment data and estimate “pseudo” effects of CSP. The results are comparable to the reported ATTs and, therefore, provide evidence against notable anticipation effects.

To introduce the two LAPD data sources, a crime, or a *reported crime incident*, in the LAPD's South Bureau region typically originates from a call to the police by a member of the public, that is, a *call-for-service*. However, because reported crimes also undergo a verification process, they filter out much of the noise associated with calls (Klinger and Bridges (1997)). Therefore, a given crime will only be listed once in the *reported crime incidents* data, while multiple, distinct calls pertaining to that incident may be reported in the *calls-for-service* data. Calls-for-service are thus viewed as an aggregate indicator of police demand, fear of crime, and victimization (Porter et al. (2019)). Reported crimes are viewed primarily as an indicator of victimization. We consider outcomes of both measures here.

Within these LAPD data sources we only include observations with valid geospatial coordinates for which we can attribute treatment status, that is, CSP or non-CSP locations can be determined. We exclude crime incidents and calls recorded with geospatial coordinates corresponding to police stations.³ We restrict our final dataset to the period of overlap for these two data sources, measuring reported crime incidents and calls-for-service outcomes during the period July 1, 2007 to December 31, 2017.

As discussed in Section 2, CSP was largely motivated by a desire to decrease violent crime rates in certain regions of South Los Angeles. Therefore, this work focuses on estimating the effect of CSP on several violent crime outcomes. The U.S. Federal Bureau of Investigation tracks two broad categories of crimes on an annual basis: Part I and Part II crimes. The so-called Part I crimes are serious crimes that occur with sufficient frequency and visibility to warrant statistical tracking. The Part I crimes include homicide, aggravated assault (assault with a deadly weapon), robbery, rape, burglary, car theft, and arson. It excludes crimes such as kidnapping which is serious, but rare. The so-called Part II crimes are lesser crimes that range from simple assault to vandalism.

Within this work we consider three violent crime outcomes related to this official crime severity categorization. First, we measure reported *Part I crime* incidents, excluding arson from the standard list of Part I crimes, due to irregular reporting in police data. Second, we consider two measures of nonproperty violent crime: *violent crime*, drawn from the reported crime incidents data, and *shots fired and violent crime*, drawn from the calls-for-service data.

Formally, we define our primary outcomes of interest as *violent crime* incidents, *shots fired and violent crime* calls-for-service, and *Part I crime* incidents. We categorize the reported crime data and calls-for-service data based on the LAPD Consolidated Crime Analysis Database (CCAD) code provided in the data sources. Our primary outcomes are defined as follows (CCAD code in parentheses): *violent crime* is defined as homicide (110), assault with a deadly weapon/ attempted homicide (230), and robbery (210); *shots fired and violent crime* calls-for-service is defined as shots fired (246), robbery (211), assault with a deadly weapon (245), and murder (187) calls. *Part I crime* incidents are defined as homicide (110), assault with a deadly weapon/ attempted homicide (230), robbery (210), rape (121), burglary (310), and stolen vehicle (510).⁴ Our primary analyses consider the aggregate raw count data, but we show the results are robust to evaluation using per capita outcomes, normalizing by the census population counts is discussed below. These analyses are included in Section S4.2.

Due to the sensitive nature of the data, we are not able to provide the raw data as part of the publication. Substantially similar open-source data is available at <https://data.lacity.org/>. In addition, individuals may request data for research purposes by contacting the LAPD directly. Instructions are available at https://www.lapdonline.org/inside_the_lapd/content_basic_view/9136.

³It is common practice to use a police station address when the location of the crime is unknown. This exclusion removes approximately two percent of the data.

⁴Nonviolent crime outcomes (reported crime incidents: residential burglary; calls-for-service: quality of life) are evaluated and included in Section S5.

3.1.1. *Temporal aggregation.* The data described above is reported in real-time. For our analysis we aggregate our crime data biannually (i.e., in six-month periods), which we refer to as *semesters*. Temporally, *semesters* are defined as events from January 1 to June 30 (denoted in our graphs as year.0) and from July 1 to December 31 (denoted year.5) within a given year. *Semester* is chosen as the smallest time-level specification, as it is better able to demonstrate seasonal trends than *year* and is less noisy than *quarter*-level data. Our choice of temporal aggregation leaves us with nine pretreatment measurements and 12 posttreatment. To evaluate the robustness of our choice, we consider aggregating up to three-month periods (*quarter*) and show that the *semester*-level results are substantively similar to the *quarter*-level results in Section S4.2.

3.2. *Treated units.* Our treated units are defined, geographically, by the public housing developments in LAPD's South Bureau that received CSP in November, 2011. In particular, this includes Jordan Downs, Nickerson Gardens, and Imperial Courts.⁵ We note that CSP was implemented in Jordan Downs, Nickerson Gardens, and Imperial Courts in November, 2011. Therefore, at the semester level, treatment implementation is approximated with the beginning of 2012 (i.e., 2012.0 in our figures).⁶

3.3. *Control units.* A key part of synthetic control methods is defining donor control units. As public housing developments are unique in their community structure, we aim to find naturally-occurring control units that are of similar size and lie within a similar geographic region. We use census boundaries, and their associated shapefiles, to construct control units for analysis.

Los Angeles is a large and diverse metropolis, so we first restrict the data to census tracts within the 77th Street and Southeast Divisions of LAPD's South Bureau, which are closest to the treated units, limiting to the southernmost Census tract in the region as tract 2911.10. We refer to this region as "South Los Angeles." By restricting the pool of control units to a geographic area more closely surrounding the treated units rather than using the whole of the LAPD area, we inherently address unobservable factors specific to this region (Abadie, Diamond and Hainmueller (2015)).

Within this restricted geographic region, donor control units are defined using census geographies. Spatially, events are grouped by census *block group* boundaries, shown for the region of study in Figure S2. In total, the pool of control units consists of 234 *block groups* where all units that eventually receive treatment are excluded. We focus on *block group*-level control units rather than *block*-level control units (the smallest census geography unit) as the former is of a similar population size to public housing units. All CSP PHDs are comprised of multiple census blocks. In our analyses, Jordan Downs and Imperial Courts each encompass seven *blocks* while Nickerson Gardens encompasses 13 *blocks*. Therefore, these PHDs are of similar geographical size to the average *block group*, or donor control unit, in our study: the average number of *blocks* within a control *block group* in our region of study is 10.41.

⁵Two additional public housing developments within LAPD's South Bureau received the CSP intervention in July 2016 (Avalon Gardens and Gonzaque Village) and one in October 2017 (Harvard Park). We do not have sufficient posttreatment data to robustly evaluate the impact; however, we use these units as a robustness check for pretreatment model fit in Section S6. CSP was also deployed in the Ramona Gardens public housing development in East Los Angeles in late 2011. Since it is outside the South Los Angeles Bureau from which we draw our donor controls, Ramona Gardens was not considered in our analyses.

⁶Under this approximation, the final pretreatment semester contains approximately two months of CSP conditions which might lead to concerns about bias due to the final pretreatment period containing some treated data. Our results are robust to assigning treatment periods to be exact (i.e., recoding incidents in November and December 2011 to have occurred during the first semester of 2012); see Section S4.2.

TABLE 1

Average violent crime counts per semester during the pretreatment period by the treated units of interest, the average Los Angeles block group, and the average control unit in the restricted, South LA region of study

Status	Region	Violent Crime Incidents	Part I Crime Incidents	Shots Fired and Violent Crime Calls
Treated	Imperial Courts	16.00	28.56	28.89
Treated	Jordan Downs	20.11	40.78	29.11
Treated	Nickerson Gardens	42.44	75.33	57.00
Control	South Los Angeles	8.31	18.20	16.43
Reference	Los Angeles	3.42	10.48	12.79

Table 1 provides average violent crime counts per semester during the pretreatment period for the treated units of study, the average unit in the control region of South LA, and, for reference, the average block group in LA more broadly. For each of the outcomes, the treated units have substantially more violent crime per semester: LA and South LA *block groups* average 3.42 and 8.31 reported *violent crime* incidents, respectively, while the treated PHDs have between 16.00 (for Imperial Courts) and 42.44 (for Nickerson Gardens) average reported *violent crime* incidents per semester. This disparity is consistent for both the reported *Part I crime* incidents and *shots fired and violent crime* calls-for-service. This outcome table shows the uniqueness of our treated units and supports our decision to limit the control area to South LA, as South LA is more similar to the treated units than LA on average.⁷

3.4. *Pretreatment covariates.* Accounting for pretreatment covariates in the construction of the weights can reduce bias in synthetic control methods (Botosaru and Ferman (2019)). Our LAPD crime data does not include additional demographic information; to account for demographic differences between the treated and donor control units, we append demographic measures from the 2010 Census. We consider the following pretreatment covariates from the census data: proportion of residents who identify as Hispanic or Latino; white, black, or “other” single-race category, total number of housing units, and variables for the distribution of residents by age and gender.⁸

Our pretreatment covariates are defined using census *block groups*.⁹ As we define our donor control units at the *block group* level, this naturally allows us to merge the census data with our crime data. However, while bigger than a census *block*, our treated PHDs do not always consist of an entire *block group* region. Therefore, to construct demographic profiles, all *blocks* within a given PHD are aggregated to construct a single *block group* unit. *Blocks* assigned to a PHD are then removed from the set of donor control *block groups*.¹⁰ For figures demonstrating the allocation of census units to PHD boundaries, see Section S1.1.

Of these included census covariates, Table 2 reports those where the treated units are most different from the South LA region from which we draw our donor control units (see Table S2

⁷We present per capita outcomes in Table S1.

⁸See Section S1.2 for the full data description.

⁹Data from the Census Bureau’s Decennial Census, which records various demographic information, are only available for *blocks* and *block groups*.

¹⁰The American Community Survey (ACS) records covariates on socioeconomic status, educational attainment, and geographical mobility. However, the ACS information is minimally recorded at the *block group* level and, therefore, would require additional assumptions to match the nonstandard PHD regions. In Section S4.1 we demonstrate the results for Jordan Downs, which comprises a single block-group, are robust to the inclusion of these ACS covariates.

TABLE 2

Demographic descriptions by the treated units of interest, the average control unit in the restricted, South LA region of study, and the average Los Angeles block group

Status	Region	Housing	Population	Hisp. Lat.	White	Black	Male Under18	Female Under18
Treated	Imperial Courts	576	1998	0.56	0.2	0.41	0.51	0.4
Treated	Jordan Downs	708	2714	0.67	0.21	0.32	0.56	0.42
Treated	Nickerson Gardens	1254	4409	0.6	0.17	0.39	0.54	0.41
Control	South Los Angeles	386	1318	0.58	0.22	0.4	0.33	0.3
Reference	Los Angeles	564	1510	0.46	0.51	0.1	0.23	0.22

for comparison of the remaining covariates). As a reference, we also present averages for broader Los Angeles. As expected, Jordan Downs, Nickerson Gardens, and Imperial Courts each have more housing units and more population than the average control unit in both Los Angeles and the South Los Angeles region of study. Nickerson Gardens is especially housing and population dense with approximately three times the population of the average LA and South LA *block group*. The treated PHDs also tend to have a greater proportion of residents under 18: the proportion of both male and female residents under 18 is nearly double that of the average control unit in LA. The racial characteristics of the treated units are similar to that of South LA, but unique from the average LA *block group*, supporting our decision to restrict donor control units to South LA.

4. Methodology. Synthetic control methods (SCMs) solved an important problem for applied policy researchers by providing a method for estimating the causal effect with one (or few) treated unit(s), relying on a version of a selection-on-observables assumption. In essence, synthetic control methods construct a counterfactual for the treated unit using a weighted combination of the control units, known as a “synthetic control.” The weights are constructed to make the synthetic control’s pretreatment outcomes match those of the treated unit, and the effect is estimated as the average difference in the posttreatment period between the treated unit and the synthetic control unit.¹¹ However, SCMs are no panacea, and the original authors only suggest the use of synthetic controls with (near) exact pretreatment fit on the pretreatment outcomes and when many pretreatment periods are available. Despite these constraints, SCM is “arguably the most important innovation in the policy evaluation literature in the last 15 years” (Athey and Imbens (2017)).

Recent advances in SCMs aim to relax the perfect pretreatment fit for pretreatment outcomes and covariates. We note extrapolation is a concern in our data, as the crime outcomes in our treated units are higher than many of the donor control units; see Figure S4. Finding a feasible set of weights becomes more difficult with higher dimensional data, particularly with a growing number of pretreatment outcomes and covariates; thus, many papers have suggested regularized approaches to constructing the weights (e.g., Ben-Michael, Feller and Rothstein (2021), Doudchenko and Imbens (2016), Robbins, Saunders and Kilmer (2017), Abadie and L’Hour (2020)). These relax the balance constraints, possibly allowing for extrapolation (such as Doudchenko and Imbens (2016) and Ben-Michael, Feller and Rothstein (2021)). There are papers that aim to weaken the linear-factor model assumption required for identification (Hazlett and Xu (2018)). Another branch of research has shown the relationship between SCM and difference-in-differences and uses this to incorporate a type of

¹¹Matching methods are related to this approach, where, with one-to-one matching, control units are given a weight of zero or one.

model-based adjustment to SCM (e.g., Arkhangelsky et al. (2018), Doudchenko and Imbens (2016), Ferman and Pinto (2019), Ben-Michael, Feller and Rothstein (2021)). We employ one such approach, augmented synthetic control methods, detailed in Section 4.4, which combines regularized weights with a model-based adjustment to improve inference when exact pretreatment fit is infeasible (Ben-Michael, Feller and Rothstein (2021)). ASCM incorporates many of the model-based adjustment approaches (including Doudchenko and Imbens (2016), Ferman and Pinto (2019), Arkhangelsky et al. (2018)) as special cases.

Finally, our data setting does not fit in the canonical SCM framework in that we have three, rather than one, treated unit. Many recent methods have extended both difference-in-differences and SCM to multiple treated units with, potentially, staggered adoption. Ben-Michael, Feller and Rothstein (2019) show that researchers can partially or fully pool the information in the treated units. We take a “fully pooled” approach suggested by O’Neill et al. (2016) and Robbins, Saunders and Kilmer (2017), in which treated units are first averaged before constructing the SCM weights and conducting the augmentation step, in effect averaging the results of separate ASCM estimates for each unit.

4.1. *Notation and estimands.* We now introduce notation and details, following those outlined in Ben-Michael, Feller and Rothstein (2021), for the augmented synthetic control method. We have $i = 1, \dots, N$ units observed for $t = 1, \dots, T$ time periods with N and T fixed. We consider D_i as an indicator that unit i is treated at time $T_0 < T$, where units with $D_i = 0$ never receive treatment. For simplicity of notation and following conventional notation, we assume that only one unit, $i = 1$, is treated. As discussed above, we have three treated units treated at time T_0 in which we use full-pooling, in effect averaging the three treated units to define unit $i = 1$. We let $N_0 = N - 1$ be the number of possible donor control units. We define the pretreatment time period as $T \leq T_0$, and the posttreatment time period as $T > T_0$.

We define the potential outcomes for unit i at time-period t as $Y_{it}(1)$ for treatment and $Y_{it}(0)$ for control (Rubin (1974)). We rely on the common stable unit treatment value assumption (SUTVA) which rules out interference between units and assumes treatment and control is commonly defined among all units (Rubin (1978)).¹² We define pretreatment outcomes as \mathbf{X} , where \mathbf{X}_1 refers to the pretreatment outcomes for the treated unit $i = 1$ and \mathbf{X}_0 refers to the matrix of pretreatment outcomes for donor control units. We also consider a small set of pretreatment covariates $\mathbf{Z}_i \in \mathbb{R}^K$ from the appended census data, with \mathbf{Z}_1 and \mathbf{Z}_0 defined analogously. The observed outcomes in our dataset are thus

$$Y_{it} = \begin{cases} Y_{it}(0) & \text{if } D_i = 0 \text{ or } t \leq T_0, \\ Y_{it}(1) & \text{if } D_i = 1 \text{ and } t > T_0. \end{cases}$$

We define the potential outcomes under control as generated by a fixed component, m_{it} , and an additive idiosyncratic component ϵ_{it} drawn from some distribution $P(\cdot)$ with mean zero, making $Y_{it}(0) = m_{it} + \epsilon_{it}$. We then define the treated potential outcome as $Y_{it}(1) = Y_{it}(0) + \tau_{it}$ where the τ_{it} are the individual level treatment effects for a given time period t . The fundamental problem of causal inference is that the potential outcomes for unit i and time t are never jointly observable, making τ_{it} unobservable (Holland (1986)). Yet this quantity is of primary interest in policy problems, such as the CSP evaluation, where we wish to know the impact of CSP for, specifically, the units who received treatment. To evaluate the CSP policy, our primary quantity of interest, then, is the average treatment effect on the treated

¹²Spillover effects are a genuine concern for crime deterrence interventions. We further evaluate potential violations of this assumption, particularly the impact of interference, in Sections S2.3 and S2.3.1, and find no statistically significant evidence of spillover effects.

(ATT), a policy-relevant estimand that captures the impact of the CSP for the PHDs that participated in the program. We can also evaluate the average treatment effect for the treated units at a given time period, called the ATT_t ,

$$(1) \quad ATT_t = \mathbb{E}[Y_{it}(1) - Y_{it}(0) | D_i = 1], \quad T_0 < t \leq T;$$

$$(2) \quad ATT = \mathbb{E}[ATT_t], \quad T_0 < t \leq T.$$

Substantively, the ATT_t (equation (1)) is the average difference between the crime rates in treated regions posttreatment and *what the crime rate would have been* in those regions during the posttreatment years had CSP not been implemented for a given posttreatment time period $t > T_0$. Equation (2), the average of the ATT_t over all posttreatment time periods, captures the overall impact of CSP during our study period.

In [Ben-Michael, Feller and Rothstein \(2021\)](#), the authors follow the extant literature and consider two models for m_{it} , including an outcome that is linear in pretreatment outcomes and covariates and a linear factor model. They consider bias bounds of ASCM under these models. Causal identification of the estimands above relies on an ignorability assumption, namely, that the treatment assignment D_i is ignorable, given model m_{it} , and formalized as

$$(3) \quad \mathbb{E}[D_i \epsilon_{iT}] = \mathbb{E}_{\epsilon_T}[(1 - D_i) \epsilon_{iT}] = \mathbb{E}_{\epsilon_T}[\epsilon_{iT}] = 0 \quad \forall T > T_0,$$

which states that the noise terms in the posttreatment time-periods are mean-zero and uncorrelated, meaning that the noise terms do not systematically differ for treated and control units. This allows us to estimate the counterfactual synthetic control as a weighted combination of the donor control units, with weights estimated to balance the pretreatment outcomes and covariates. The bias bounds further impose independence across time and units with common variance.

4.2. Difference-in-differences. A common approach to estimating causal effects with panel data is difference-in-differences, which leverage a “parallel trends” assumption, that the treated units follow a parallel, but mean shifted, path to the control units in the absence of treatment to address confounding of treatment assignment. Difference-in-differences was originally derived under a single pre- and posttreatment time period ([Card and Krueger \(1993\)](#), [Angrist and Pischke \(2008\)](#)), but recent literature has extended these results to allow for more pre- and posttreatment time periods with optional staggered treatment implementation (e.g., [Athey and Imbens \(2018\)](#), [Callaway and Sant’Anna \(2020\)](#), [Egami and Yamauchi \(2021\)](#)) and shown the relationship to common two-way linear fixed effects estimators in time-series cross-sectional analysis (e.g., [Imai, Kim and Wang \(2020\)](#)). With numerous pretreatment observations, researchers can rely on the extended parallel trends assumption, which provides a testable implication (i.e., a placebo test), namely that parallel trends holds among the pretreatment time periods ([Callaway and Sant’Anna \(2020\)](#), [Egami and Yamauchi \(2021\)](#)).

4.2.1. Failure of parallel trends. We leverage our multiple pretreatment time periods to evaluate the plausibility of the parallel trends assumption in our application. The right-hand panels of [Figure 1](#) presents the time trends in the treated and control units for each outcome. As is visually evident, the treated and control units do not follow similar pretreatment trends, with crime measures spiking then steeply declining in our treated PHDs and crime more gently declining in the control region. We conduct a statistical test of this assumption, using the double difference-in-differences estimator of [Egami and Yamauchi \(2021\)](#), which tests the parallel trends-in-trends assumption, a further relaxation of the extended parallel trends assumption allowing for linear time-varying unmeasured confounding.

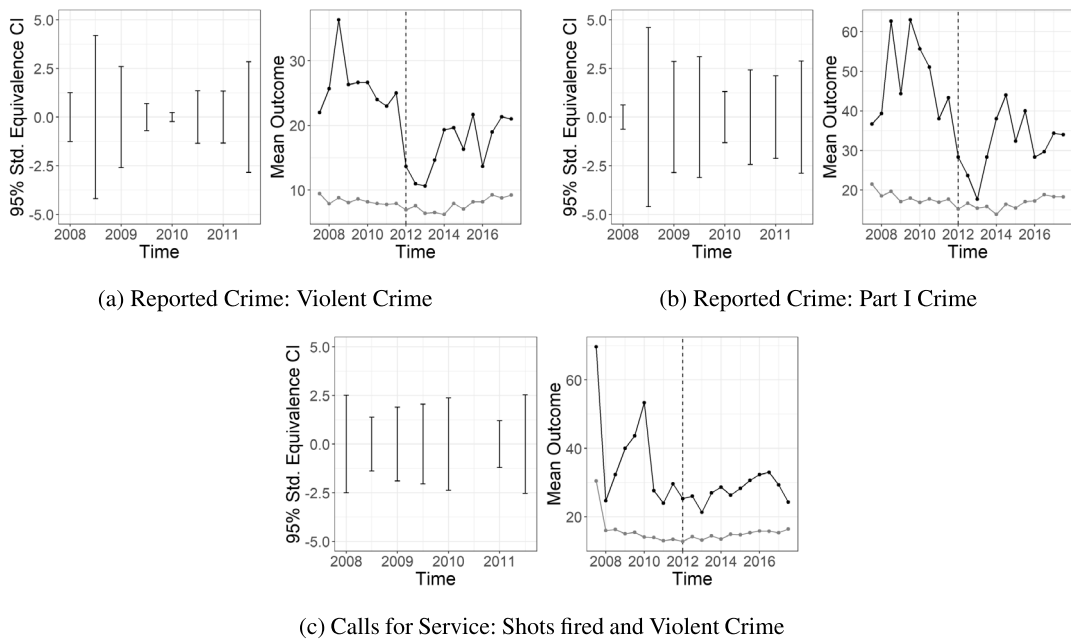


FIG. 1. To assess the plausibility of the parallel trends assumption, we plot the equivalence confidence intervals, in standard deviations from baseline control group mean, across pretreatment time (left) and the mean outcome across time (right) for the treated units (black) and the control units (gray). Parallel trends are not plausible for these outcomes.

The left-hand panels of Figure 1 provides statistical evidence of strong deviations from even this weaker version of parallel trends. Using the double difference-in-differences estimator, the figure presents standardized equivalence confidence intervals (Hartman and Hidalgo (2018)) of the effect at each pretreatment time-period. The parallel trends-in-trends assumption implies these estimates should be close to zero, and, as discussed in Egami and Yamauchi (2021), the equivalence confidence interval should be narrow to provide credible evidence for parallel trends. Our data exhibit large ranges around zero, with many time periods exhibiting differences of one or more standard deviations above the baseline control mean, indicating a lack of evidence for the plausibility of the parallel trends assumption. One alternative is to rely on a conditional parallel trends assumption (see Callaway and Sant’Anna (2020)) which allows for the inclusion of covariates to meet the parallel trends assumption; however, given the large deviations from parallel trends, our relatively limited set of additional observed covariates, and our limited number of treated units, we instead rely on the alternative identification strategy provided by SCMs.

4.3. Synthetic control method weights. We begin by describing estimation using synthetic control methods (SCMs), which estimate equations (1) and (2) by constructing a counterfactual for the treated unit as a convex combination of the donor control units, where the weights are chosen such that the weighted average of the controls closely approximates the pretreatment outcomes and covariates for the treated units. Specifically, we follow Ben-Michael, Feller and Rothstein (2021) and choose SCM to solve the following constrained optimization problem:¹³

$$(4) \quad \min_{w^{\text{SCM}} \in \Delta^{N_0}} \theta_x \| \mathbf{X}_1 - \mathbf{X}_0' w^{\text{SCM}} \|_2^2 + \theta_z \| \mathbf{Z}_1 - \mathbf{Z}_0' w^{\text{SCM}} \|_2^2 + \zeta \sum_{i:D_i=0} f(w_i^{\text{SCM}}),$$

¹³An importance matrix of the pretreatment outcomes and covariates can be incorporated; see Ben-Michael, Feller and Rothstein (2021).

where the weights w_i^{SCM} are constrained to be in the simplex $\Delta_0^N = \{w^{\text{SCM}} \in \mathbb{R}^{N_0} \mid w_i^{\text{SCM}} \geq 0 \forall i, \sum_i w_i^{\text{SCM}} = 1\}$ and are normalized to sum to 1. θ_x and θ_z serve as a type importance measure. Following the original authors, we consider pretreatment outcomes and covariates of equal importance, with $\theta_x = \theta_z = 1$. The “synthetic control” counterfactual is estimated as the weighted average of the posttreatment outcomes of the control units, $\widehat{Y}_{1t}^{\text{SCM}}(0) = \sum_{i:D_i=0} w_i^{\text{SCM}} Y_{it}$ for time period $t > T_0$, with weights w^{SCM} . Thus, we can estimate the $\text{ATT}_t, t > T_0$, as the difference in the treated unit and the weighted donor control units, and the ATT as the average of the ATT_t over all $t, t > T_0$,

$$(5) \quad \widehat{\text{ATT}}_t = Y_{1t} - \widehat{Y}_{1t}^{\text{SCM}}(0) \quad \text{for } t > T_0,$$

$$(6) \quad \widehat{\text{ATT}} = \frac{1}{T - T_0} \sum_{t:t>T_0} (Y_{1t} - \widehat{Y}_{1t}^{\text{SCM}}(0)).$$

The ζ in equation (4) is a hyperparameter proposed by [Abadie, Diamond and Hainmueller \(2015\)](#), which penalizes a measure of dispersion of the weights, defined by $f(w_i^{\text{SCM}})$. We return to the choice of ζ and $f(w_i^{\text{SCM}})$ below, ultimately relying on ridge regression for regularization which penalizes the variance of the weights with $f(w_i^{\text{SCM}}) = (w_i^{\text{SCM}} - \frac{1}{N_0})^2$. The constraint that the weights must be positive implies that the treated unit must be within the convex hull of the donor control units, a requirement that is harder to meet as the dimensionality of the problem increases, such as with more pretreatment time periods or covariates ([Ferman and Pinto \(2019\)](#)). If weights are allowed to be negative, the estimator extrapolates beyond the convex hull. While the original suggestion of [Abadie, Diamond and Hainmueller \(2010\)](#) was to avoid using SCM when excellent pretreatment fit on the pretreatment outcomes and covariates cannot be achieved, ASCM relaxes the constraints to allow for principled extrapolation beyond the convex hull.

4.3.1. *Synthetic control weights with intercept shift.* One approach to addressing the failure of the parallel trends assumption for difference-in-differences is to combine the synthetic control method with an intercept shift ([Doudchenko and Imbens \(2016\)](#), [Ferman and Pinto \(2019\)](#)). This allows for unobservable unit-level differences, unaddressed by the synthetic control weights, by differencing the unit-level pretreatment average for each unit. Let $\bar{Y}_i = \frac{1}{T_0} \sum_{t:t \leq T_0} Y_{it}$ be the pretreatment average for unit i . This results in the following estimator for the ATT_t :

$$(7) \quad \widehat{\text{ATT}}_t = Y_{1t} - \bar{Y}_1 - \sum_{i:D_i=0} w_i^{\text{SCM}} (Y_{it} - \bar{Y}_i) \quad \text{for } t > T_0.$$

When the weights are uniform, this estimator is equivalent to the standard difference-in-differences estimator. This “SCM with intercept-shift” model is a special case of the augmented synthetic control method that we describe below.

4.4. *Augmented synthetic control methods (ASCM).* When excellent pretreatment fit on the pretreatment outcomes and covariates is not achievable, given the constraints on the weights in equation (4), [Ben-Michael, Feller and Rothstein \(2021\)](#) suggest an augmented synthetic control. ASCM relaxes the constraint that the weights lie within the simplex, allowing for negative weights on donor control units—a form of extrapolation. Akin to doubly robust estimation ([Bang and Robins \(2005\)](#)), ASCM introduces an augmentation step that first estimates then adjusts for the potential bias in the SCM estimate, due to residual imbalance in the pretreatment outcomes and covariates. The generic form of the ASCM estimator

for the potential outcome under control is

$$(8) \quad \widehat{Y}_{it}^{ASCM}(0) = \underbrace{\sum_{i:D_i=0} w_i^{SCM} Y_{it}}_{\text{SCM Estimator}} + \underbrace{\left(\widehat{m}_{1t}(\mathbf{X}, \mathbf{Z}) - \sum_{i:D_i=0} w_i^{SCM} \widehat{m}_{it}(\mathbf{X}, \mathbf{Z}) \right)}_{\text{Augmentation (Bias Adjustment)}}$$

for $t > T_0$, and \widehat{ATT}_t and \widehat{ATT} are estimated analogously to SCM, replacing $\widehat{Y}_{1t}^{SCM}(0)$ with $\widehat{Y}_{1t}^{ASCM}(0)$. Notice the difference from SCM is that ASCM includes a model-based bias adjustment as the difference between a model-based prediction for the treated unit at time t , given pretreatment outcomes and covariates ($\widehat{m}_{1t}(\mathbf{X}, \mathbf{Z})$) and the SCM weighted model-based prediction for the donor control units ($\sum_{i:D_i=0} w_i^{SCM} \widehat{m}_{it}(\mathbf{X}, \mathbf{Z})$) at time t . Following Ben-Michael, Feller and Rothstein (2021), we rely on ridge regression (Hoerl and Kennard (1970)) for our augmentation step, with $\widehat{m}(\mathbf{X}, \mathbf{Z}) = \widehat{\eta}_0 + \mathbf{X}'\widehat{\eta}_x + \mathbf{Z}'\widehat{\eta}_z$, which is fit with ridge regression to solve

$$(9) \quad \min_{\eta_0, \eta_x, \eta_z} = \frac{1}{2} \sum_{i:D_i=0} (Y_i - (\eta_0 + \mathbf{X}'\eta_x + \mathbf{Z}'\eta_z))^2 + \lambda_x \|\eta_x\|_2^2 + \lambda_z \|\eta_z\|_2^2.$$

As noted above, we consider pretreatment outcomes and covariates of equal importance, with $\theta_x = \theta_z = 1$, and we regularize with a common parameter $\lambda_x = \lambda_z = \lambda^{ridge}$. Under a linear data-generating process, ridge ASCM, as described above, is a type of regularized SCM weight, as described in equation (4), with a ridge penalty. Pretreatment outcomes and covariates that are more predictive of the outcome will have larger coefficients and, therefore, be more important in balancing, with larger imbalances leading to larger augmentation adjustments. Under a linear model m_{it} , ridge ASCM penalizes divergence from the SCM weights, where λ^{ridge} controls the level of extrapolation.

We follow the suggestion of Ben-Michael, Feller and Rothstein (2021) and select λ^{ridge} using cross-validation with a leave-one-out time-period for $t < T_0$ (i.e., the pretreatment period) to find the λ^{ridge} that minimizes MSE. We also consider an additional scenario in which we achieve exact balance on lagged outcomes ($\lambda_z = 0$) and conduct ridge ASCM after residualizing the pretreatment outcomes on the pretreatment covariates. Importantly, the hyperparameter λ allows ridge ASCM to navigate the bias-variance trade-off, when SCM weights are feasible and controls extrapolation in a principled way, while deviating as little as possible from the SCM weights (thus minimizing extrapolation). Ben-Michael, Feller and Rothstein (2021) show that ASCM controls bias under both a linear and an (unobserved) linear factor model, extending the well-known results of Abadie, Diamond and Hainmueller (2010) for the original SCM weights.

5. Model evaluation. Evaluating the credibility of SCM methods is difficult, particularly in light of the fact that we are estimating the treatment effect for a very small number of units, in this case only three PHDs. Traditionally, the most common way to evaluate the fit of synthetic control methods has been to use placebo tests (Abadie, Diamond and Hainmueller (2015)), which rely on the idea that we should not find evidence of treatment effects where none should exist, that is, before treatment has been implemented or among the control units. Estimation of nonzero placebo effects would undermine the credibility of the final results. The more evidence that passes the scrutiny of the placebo tests, the more credible the resulting analysis.

These placebo tests, suggested by the literature (Abadie, Diamond and Hainmueller (2010), Abadie, Diamond and Hainmueller (2015), Heckman and Hotz (1989)), serve as the primary method of evaluating the appropriateness of SCM methods. In particular, using *in-time placebos* estimating placebo (i.e., null) effects for the pretreatment period to assess the

model fit, we assess the fit of the ASCM for each outcome separately. Taken holistically, these checks are used to evaluate the credibility and confidence in our estimates.

Before evaluating the impact of CSP in the posttreatment periods, we assess the ability of the ASCM to balance the trajectory of the pretreatment outcome for the treated units and the synthetic control. During the pretreatment period, CSP was not implemented and, therefore, the ASCM models should not detect a treatment effect for any outcome. Nonzero effects would indicate possible remaining confounding, calling into question causal inferences on that outcome. Any observed imbalance could also indicate the potential scale of bias in the estimated impact of CSP.

To ensure the placebo ASCM models are only evaluated on the pretreatment period, we split the pretreatment period into training and testing sets, using roughly a 2/3 : 1/3 rule, with pseudo-implementation dates as 2010.5 and 2011. Therefore, these methods have 6 : 3 and 7 : 2 training to testing periods, respectively, with which to implement the Ridge ASCM with covariates model. While the results of this paper focus on the Jordan Downs, Nickerson Gardens, and Imperial Courts PHDs, the model specification approach described here severely limits the time frame for these regions. Therefore, we provide additional placebo checks, using the in-time placebos on two additional PHDs with later implementation dates in Section S6 of the Supplementary Material (Kahmann et al. (2022)).

The results for these in-time, model evaluation placebos are shown in Table 3. The placebo effect for *violent crime* is estimated as 4.88 ($p = 0.71$) and 5.74 ($p = 0.45$), for the 2010.5 and 2011 pseudo-implementation dates, respectively. For reference, the pretreatment semester average for this outcome in Jordan Downs, Nickerson Gardens, and Imperial Courts is 26.19 incidents. Therefore, the ratio of the placebo estimate to the pretreatment average is approximately 19% and 22%, respectively. Given only six and seven periods were used to train these pretreatment models, respectively, this is a positive result suggesting the ASCM is able to capture unobserved confounding in the outcome and selection into treatment.

Part I crime incidents have an estimated pseudo-ATT of 0.03 ($p = 0.74$) and -5.84 ($p = 0.88$) for the two implementation dates, respectively, approximately 0% and 12% of the pretreatment semester average of 48.22. The *shots fired and violent crime* calls-for-service estimates are -3.49 ($p = 0.87$) and -2.48 ($p = 0.79$) for the 2010.5 and 2011 treatment dates, approximately 9% and 6%, respectively, of the pretreatment semester average of 38.33. We consider the proportionally small estimated bias in these outcomes a good indicator of our ability to construct appropriate synthetic control models for the *Part I crime* incidents and *shots fired and violent crime* calls-for-service outcomes.

6. Results. In this section we evaluate the treatment effect of CSP on the reported *violent crime* incidents, reported *Part I crime* incidents, and *shots fired and violent crime* calls-for-service outcomes across the Jordan Downs, Nickerson Gardens, and Imperial Courts PHDs. Following Ben-Michael, Feller and Rothstein (2021), we report results across a suite of models. We consider: (1) standard SCM balancing on pretreatment outcomes and covariates,

TABLE 3

In-time Placebos: *The estimated ATTs are provided for the pseudo-treatment implementation dates of 2010.5 and 2011. For comparison, the average counts of each violent crime outcome per semester are shown. The p-values for the joint null of no effect in any posttreatment period are provided in parentheses*

Outcome	Pre-T Average	2010.5	2011
Violent Crime Incidents	26.19	4.88 (0.71)	5.74 (0.45)
Part I Crime Incidents	48.22	0.03 (0.74)	-5.84 (0.88)
Shots Fired and Violent Crime Calls	38.33	-3.49 (0.87)	-2.48 (0.79)

TABLE 4

Estimated impact of CSP for the Jordan Downs, Nickerson Gardens, and Imperial Courts PHDs. ATT estimates are provided with conformal inference p -values for the joint null that the effect is zero in every posttreatment period in parentheses across our four models

Outcome	SCM	SCM + FE	ASCM Ridge	Residualized ASCM Ridge
Violent Crime Incidents	-3.97 (0.96)	-8.86 (0.59)	-5.51 (0.48)	-5.04 (0.52)
Part I Crime Incidents	-7.51 (0.91)	-16.26 (0.77)	-10.36 (0.58)	-11.26 (0.93)
Shots Fired and Violent Crime Calls	2.02 (0.99)	-6.36 (0.99)	-1.26 (1)	-3.73 (0.92)

(2) an SCM with intercept-shift model balancing on pretreatment outcomes and covariates after removing unit-level pretreatment means, (3) Ridge ASCM with covariates, balancing on pretreatment outcomes and pretreatment covariates, and (4) Ridge ASCM with residualized outcomes, balancing on pretreatment outcomes residualized against the pretreatment covariates. 95% pointwise confidence intervals for each time period in the plots are computed via conformal inference (Chernozhukov, Wüthrich and Zhu (2021)).

From Section 5 we find evidence that these results pass several placebo tests, bolstering confidence in the fit of the ASCM models for these outcomes of interest. As presented in Table 4, CSP has an estimated impact of, approximately, 5.3 fewer reported *violent crime* incidents, with estimates of -5.51 ($p = 0.48$) and -5.04 ($p = 0.52$) for the unresidualized and residualized ridge ASCM covariate models, respectively, per semester per public housing development during the posttreatment period. Here and subsequently, we present permutation-based conformal inference p -values for the joint null hypothesis of no effect in any posttreatment period in parentheses.¹⁴ The *shots fired and violent crime* calls-for-service outcome suggests an insignificant reduction of approximately 2.5, with estimates of -1.26 ($p = 1$) and -3.73 ($p = 0.92$) for the unresidualized and residualized covariate models, respectively. Compared to the pretreatment semester averages, these PHDs experienced an average decrease of 19–21% and 3–10% in reported *violent crime* incidents and *shots fired and violent crime* calls-for-service per semester, respectively.

Additionally, CSP has an estimated reduction of approximately 10.8 reported *Part I crime* incidents per semester per housing development during the posttreatment period, with estimates of -10.36 ($p = 0.58$) and -11.26 ($p = 0.93$) for the unresidualized and residualized covariate models, respectively. Considering these covariate models, the ATT estimate corresponds to a decrease of 21–23% compared to the pretreatment semester average.

Figure 2 provides the ATT_t estimates across time across the suite of models for each of our outcomes. Consistent with our results above, our point estimates generally indicate reductions in violent crimes and calls-for-service, but they are statistically insignificant.

We find that the results are robust to the type of SCM or ASCM model used. The estimates for SCM and Ridge ASCM without covariates are very similar for each outcome. For *violent crime* incidents (Figure 2(a)) and *Part I crime* incidents (Figure 2(b)), we see substantial improvement in model fit from including the augmentation step, as evidenced by the pretreatment trends moving closer to zero. We still see some evidence of pretreatment imbalance in the *shots fired and violent crime calls* (Figure 2(c)), even with the inclusion of the augmentation step, warranting some caution when interpreting results.

¹⁴Abadie, Diamond and Hainmueller (2010) propose a permutation-based inference technique estimating the placebo effect for each control unit. While we rely on the conformal inference approach, we note that the p -values for our *violent crime* and *Part I crime* estimates are significant (both $p = 0$), while the *shots fired and violent crime calls* estimate is insignificant ($p = 0.97$) using this technique; see Section S3.2 for details.

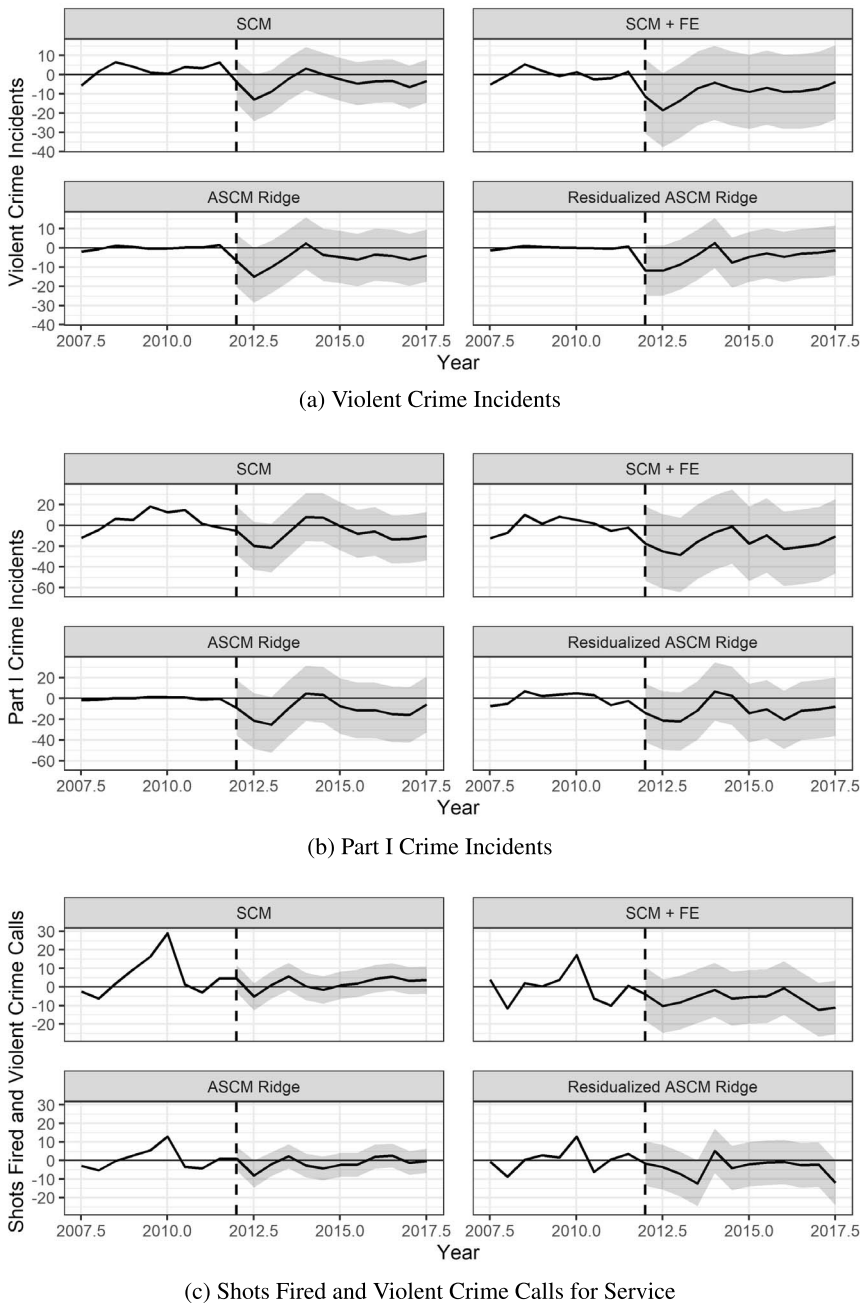


FIG. 2. The over time $ATT_t (t > T_0)$ estimates for for each outcome for each model, including (1) standard SCM; (2) SCM with fixed effects model; (3) Ridge ASCM; (4) Residualized Ridge ASCM. 95% point-wise conformal inference confidence intervals provided in gray.

In Section S3.1 we evaluate the similarities in the weights and residual covariate imbalances across model specification for each outcome. We note that the SCM and SCM with intercept shift models ascribe weights of zero or near-zero to most controls, with only a handful of donor units having nonzero weights. The ridge ASCM models have far less sparse weights with many between -0.10 to 0.10 on these same donor units (Figure S9). This extrapolation, combined with the augmentation step, allows the ASCM to achieve better covariate balance

and pretreatment fit than SCM and SCM with intercept shift models (Figure S8), boosting confidence in the ridge ASCM model estimates.

7. Discussion and conclusions. This paper seeks to quantify the effect of the Community Safety Partnership (CSP) on reported crime incidents and calls-for-service in Jordan Downs, Nickerson Gardens, and Imperial Courts, three public housing developments in South Los Angeles. These housing developments experienced high levels of violent crime, entrenched multigenerational gangs, and policing focused on crime suppression before treatment implementation in 2011. Using the augmented synthetic control method, which allows us to construct counterfactuals for communities in which CSP was implemented, observed reductions in violent crime measures were not statistically significant. We estimate that CSP led to an average reduction of 5.51 ($p = 0.48$) fewer violent crimes and 1.26 ($p = 1$) fewer violent crime calls-for-service per semester per housing development between January 1, 2012 and December 31, 2017. The estimates reflect reductions of 21% and 3% in reported crime incidents and calls-for-service, respectively, compared to preintervention means. CSP led to an average decrease of 10.36 ($p = 0.58$) Part I crime incidents per semester per housing development during this same period, corresponding to an average decrease of 21% in the Part I crime rate. Confidence bounds for these estimates are large, perhaps unsurprisingly, due to the nature of the synthetic control framework which has few treated units as well as the inherent noisiness of the crime data. Furthermore, preliminary analyses suggest CSP did not simply displace crime from Jordan Downs to neighboring regions, as discussed in Sections S2.3 and S2.3.1.

We report results using raw counts of crime and disorder outcomes. However, the population density of the PHDs we study is above that of the average control unit, as shown in Table 2. We also analyze per-capita outcomes, defined as crime counts per 1000 residents. The census population vector is recorded by *block*, so we can construct an exact estimate of population, in terms of perfectly matched spatial boundaries for both the treated and control units. Ultimately, as seen in Section S4.2, the per-capita results are substantively similar to our analyses.

Our results are consistent with many in the broader literature. Community policing programs variously incorporate police organizational change, community engagement, and problem-solving efforts (Eck and Maguire (2000), Skogan (2006)). While the focal problems are supposed to be those that the community identifies as most important to them (e.g., abandoned cars, public intoxication), the responsibility for solving those problems has generally fallen on the police with (or without) help from partner agencies within the city. The heterogeneity in implementations has made evaluation of community policing difficult. Contextualizing our results, we note that many studies have found limited evidence that community policing reduces crime (Cordner (1997), Eck and Maguire (2000), MacDonald (2002)), and, as with our study, a number of studies have found statistically insignificant reductions in crimes (see the systematic review in Blair et al. (2021)). A recent experimental study of community policing across six countries in the Southern Hemisphere found largely similar results (Blair et al. (2021)). Partial implementation and declining model fidelity over time contribute to observed null effects (MacDonald (2002)). Overall, the point estimates for the impact of CSP appear at a similar level to other intervention strategies, including both crime suppression, such as gang injunctions (Grogger (2002), Ridgeway et al. (2019)), focused deterrence approaches (Braga et al. (2001), Kennedy (1997), Kennedy, Piehl and Braga (1996)), and community-led interventions, such as GRYD (Brantingham, Tita and Herz (2021), Park et al. (2021)). Evaluations of focused deterrence programs implemented in cities across the United States have documented reductions in crime and delinquency both larger than and

comparable to that observed with CSP.; see [Braga and Weisburd \(2012\)](#) for a comprehensive meta-study of focused deterrence strategies which finds statistically significant overall effects.

There are several limitations to our work that are worth noting. First, CSP existed alongside other police and community-based approaches to crime and disorder. Gangs based in Nickerson Gardens and Jordan Downs were subject to gang injunctions, starting in 2003 and 2005, respectively. Gangs based in Imperial Courts were not subject to an injunction, though the injunction in Nickerson Gardens covered Imperial Courts geographically. The two injunctions were in place continuously over the period of observation and were only curtailed in 2018. The impact of injunctions are, therefore, expected to be part of both the pretreatment and treatment periods at Nickerson Gardens and Jordan Downs and may have also influenced Imperial Courts. The GRYD program, also in place during our period of study, covered the entire geographic region (the LAPD's South Bureau) containing both our treated and donor control units. While these bundled interventions do not affect our underlying causal identification strategy, since they were in existence throughout the entire time period of our study, they may impact our interpretation of the treatment provided by CSP, especially if the impact of such community policing programs are dependent on the context of existing policing interventions.

Second, it is important to recognize that our results here do not speak to the other major component of CSP which was to restore trust and build lasting relationships between police and the communities they serve ([Leap \(2020\)](#), [Rice and Lee \(2015\)](#)). Clearly ensuring the safety and security of community members is a necessary component of such a process, but there is more involved than simply low crime numbers. Future work will need to integrate evidence from the qualitative impact of CSP on people's lives as the counterpart to this quantitative story.

Acknowledgments. The authors would like to thank LAPD Chief of Police Michel Moore, the CSP Research and Evaluation Advisory Committee, and the community residents and institutional partners that make CSP possible. We acknowledge Connie Rice as a cocreator of the CSP public safety model and a visionary regarding relationship-based policing. The fourth author serves on the board of PredPol.

We also thank Chad Hazlett, the UCLA Causality Reading Group, Avi Feller, and Eli Ben-Michael for thoughtful comments and discussion. We additionally thank Eli Ben-Michael for providing plot code from the [Ben-Michael, Feller and Rothstein \(2021\)](#) paper.

Funding. This research was funded by the Balmer Group, The California Endowment, Caruso, Cindy Miscikowski, The Smidt Foundation, the Weingart Foundation, and an anonymous donor.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1650604 and DGE-2034835. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

SUPPLEMENTARY MATERIAL

Impact evaluation of the LAPD community safety partnership (DOI: [10.1214/21-AOAS1543SUPP](https://doi.org/10.1214/21-AOAS1543SUPP); .pdf). This supplementary document includes additional data descriptions, an extended discussion of identification assumptions and their validity in this context, nonviolent crime outcome models, additional robustness analyses, model specification checks, and preliminary analyses for Avalon Gardens and Gonzague Village.

REFERENCES

- ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *J. Amer. Statist. Assoc.* **105** 493–505. MR2759929 <https://doi.org/10.1198/jasa.2009.ap08746>
- ABADIE, A., DIAMOND, A. and HAINMUELLER, J. (2015). Comparative politics and the synthetic control method. *Amer. J. Polit. Sci.* **59** 495–510.
- ABADIE, A. and GARDEAZABAL, J. (2003). The economic costs of conflict: A case study of the Basque country. *Am. Econ. Rev.* **93** 113–132.
- ABADIE, A. and L' HOUR, J. (2020). A penalized synthetic control estimator for disaggregated data.
- ANGRIST, J. D. and PISCHKE, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, Princeton.
- ARKHANGELSKY, D., ATHEY, S., HIRSHBERG, D. A., IMBENS, G. W. and WAGER, S. (2018). Synthetic difference in differences. ArXiv.
- ATHEY, S. and IMBENS, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *J. Econ. Perspect.* **31** 3–32.
- ATHEY, S. and IMBENS, G. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. ArXiv.
- BANG, H. and ROBINS, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61** 962–972. MR2216189 <https://doi.org/10.1111/j.1541-0420.2005.00377.x>
- BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2019). Synthetic controls with staggered adoption. ArXiv.
- BEN-MICHAEL, E., FELLER, A. and ROTHSTEIN, J. (2021). The augmented synthetic control method. *J. Amer. Statist. Assoc.* 1–34.
- BLACKSTONE, J. (2014). A change of tune for LAPD as community policing makes gains.
- BLAIR, G., WEINSTEIN, J., CHRISTIA, F., ARIAS, E., BADRAN, E., BLAIR, R. A., CHEEMA, A., FAROOQUI, A., FETZER, T. et al. (2021). Does community policing build trust in police and reduce crime? Evidence from six coordinated field experiments in the global South. Submitted.
- BLUMSTEIN, A. (2018). Science and technology and the president's crime commission: Past and future. *Criminol. Public Policy* **17** 271–282.
- BLUMSTEIN, A., WALLMAN, J. and FARRINGTON, D. (2006). *The Crime Drop in America*. Cambridge Univ. Press, Cambridge.
- BOTOSARU, I. and FERMAN, B. (2019). On the role of covariates in the synthetic control method. *Econom. J.* **22** 117–130. MR4021116 <https://doi.org/10.1093/ectj/utz001>
- BRAGA, A. A. and WEISBURD, D. L. (2012). The effects of focused deterrence strategies on crime: A systematic review and meta-analysis of the empirical evidence. *J. Res. Crime Delinq.* **49** 323–358.
- BRAGA, A. A., KENNEDY, D. M., WARING, E. J. and PIEHL, A. M. (2001). Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's operation ceasefire. *J. Res. Crime Delinq.* **38** 195–225.
- BRANTINGHAM, P. J., TITA, G. and HERZ, D. (2021). The impact of the city of Los Angeles mayor's office of Gang Reduction and Youth Development (GRYD) comprehensive strategy on crime in the city of Los Angeles. *Justice Eval. J.* 1–20.
- CALLAWAY, B. and SANT' ANNA, P. H. C. (2020). Difference-in-differences with multiple time periods. *J. Econometrics*.
- CARD, D. and KRUEGER, A. B. (1993). Minimum wages and employment: A case study of the fast food industry in New Jersey and Pennsylvania Technical Report National Bureau of Economic Research.
- CHERNOZHUKOV, V., WÜTHRICH, K. and ZHU, Y. (2021). An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls.
- CORDNER, G. W. (1997). Community policing: Elements and effects. *Crit. Issues Policing: Contemp. Read.* **5** 401–418.
- DOJ, U. (2009). Community policing defined. *Office Community Oriented Policing Serv.*
- DOUDCHENKO, N. and IMBENS, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. ArXiv.
- ECK, J. E. and MAGUIRE, E. R. (2000). Have changes in policing reduced violent crime? An assessment of the evidence. *Crime Drop Amer.* **207** 207–265.
- EGAMI, N. and YAMAUCHI, S. (2021). Using multiple pre-treatment periods to improve difference-in-differences and staggered adoption design. ArXiv.
- FAGAN, J. and MACDONALD, J. (2012). Policing, crime and legitimacy in New York and Los Angeles: The social and political contexts of two historic crime declines. Columbia Public Law Research Paper 12-315.
- FERMAN, B. and PINTO, C. (2019). Synthetic controls with imperfect pre-treatment fit. ArXiv.
- GOLDSTEIN, H. (1977). *Policing a Free Society*. Ballinger Pub. Co., Cambridge, MA.
- GOLDSTEIN, H. (1979). Improving policing: A problem-oriented approach. *Crime Delinq.* **25** 236–258.

- GROGGER, J. (2002). The effects of civil gang injunctions on reported violent crime: Evidence from Los Angeles county. *J. Law Econ.* **45** 69–90.
- HARTMAN, E. and HIDALGO, F. D. (2018). An equivalence approach to balance and placebo tests. *Amer. J. Polit. Sci.* **62** 1000–1013.
- HAZLETT, C. and XU, Y. (2018). Trajectory balancing: A general reweighting approach to causal inference with time-series cross-sectional data.
- HECKMAN, J. J. and HOTZ, V. J. (1989). Choosing among alternative nonexperimental methods for estimating the impact of social programs: The case of manpower training. *J. Amer. Statist. Assoc.* **84** 862–874.
- HOERL, A. E. and KENNARD, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.
- HOLLAND, P. W. (1986). Statistics and causal inference. *J. Amer. Statist. Assoc.* **81** 945–970. [MR0867618](https://doi.org/10.2307/2287000)
- IMAI, K., KIM, I. and WANG, E. (2020). Matching methods for causal inference with time-series cross-sectional data. Unpublished Manuscript.
- KAHMANN, S., HARTMAN, E., LEAP, J. and BRANTINGHAM, P. J. (2022). Supplement to “Impact Evaluation of the LAPD Community Safety Partnership.” <https://doi.org/10.1214/21-AOAS1543SUPP>
- KATZENBACH, N. D. (1967). *The Challenge of Crime in a Free Society: A Report by the President's Commission on Law Enforcement and Administration of Justice*. United States Government Printing Office.
- KELLING, G. L. and MOORE, M. H. (1989). *The Evolving Strategy of Policing* **4**. US Department of Justice, Office of Justice Programs, National Institute of Justice.
- KELLING, G. L., PATE, A., FERRARA, A., UTNE, M. and BROWN, C. E. (1981). *The Newark Foot Patrol Experiment*. Police Foundation, Washington, DC, 94–96.
- KENNEDY, D. M. (1997). Pulling levers: Chronic offenders, high-crime settings, and a theory of prevention. *Valparaiso Univ. Law Rev.* **31** 449–484.
- KENNEDY, D., PIEHL, A. and BRAGA, A. (1996). Youth violence in Boston: Gun markets, serious youth offenders, and a use reduction strategy. *Law Contemp. Probl.* **59** 147–183.
- KLINGER, D. A. and BRIDGES, G. S. (1997). Measurement error in calls-for-service as an indicator of crime. *Criminology* **35** 705–726.
- LEAP, J. (2020). *Evaluation of the LAPD Community Safety Partnership*. UCLA Luskin School of Public Affairs, Los Angeles.
- MACDONALD, J. M. (2002). The effectiveness of community policing in reducing urban violence. *Crime Delinq.* **48** 592–618.
- NAHUM-SHANI, I., SMITH, S. N., SPRING, B. J., COLLINS, L. M., WITKIEWITZ, K., TEWARI, A. and MURPHY, S. A. (2018). Just-in-time adaptive interventions (JITIs) in mobile health: Key components and design principles for ongoing health behavior support. *Annals Behav. Med.* **52** 446–462.
- O'NEILL, S., KREIF, N., GRIEVE, R., SUTTON, M. and SEKHON, J. S. (2016). Estimating causal effects: Considering three alternatives to difference-in-differences estimation. *Health Serv. Outcomes Res. Methodol.* **16** 1–21. <https://doi.org/10.1007/s10742-016-0146-8>
- FEDERAL BUREAU OF INVESTIGATION (2021). Uniform Crime Reporting Program Data: Supplementary Homicide Reports, 1987–2017. Inter-university Consortium for Political and Social Research.
- LA OFFICE OF THE MAYOR (2017). Mayor Garcetti announces new expansion of Community Safety Partnership.
- PARK, J., SCHOENBERG, F. P., BERTOZZI, A. L. and BRANTINGHAM, P. J. (2021). Investigating clustering and violence interruption in gang-related violent crime data using spatial-temporal point processes with covariates. *J. Amer. Statist. Assoc.* 1–14.
- PORTER, L. C., CURTIS, A., JEFFERIS, E. and MITCHELL, S. (2019). Where's the crime? Exploring divergences between call data and perceptions of local crime. *Br. J. Criminol.* **60** 444–467.
- REISIG, M. D. (2010). Community and problem-oriented policing. *Crime Justice* **39** 1–53.
- LAPD NEWS RELEASE (2015). LAPD's Community Safety Partnership program NR15021SF.
- RICE, C. and LEE, S. K. (2015). *Relationship-Based Policing Achieving Safety in Watts*. Advancement Project, Washington, DC.
- RIDGEWAY, G., GROGGER, J., MOYER, R. A. and MACDONALD, J. M. (2019). Effect of gang injunctions on crime: A study of Los Angeles from 1988–2014. *J. Quant. Criminol.* **35** 517–541.
- ROBBINS, M. W., SAUNDERS, J. and KILMER, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. *J. Amer. Statist. Assoc.* **112** 109–126. [MR3646556 https://doi.org/10.1080/01621459.2016.1213634](https://doi.org/10.1080/01621459.2016.1213634)
- RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66** 688.
- RUBIN, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *Ann. Statist.* **6** 34–58. [MR0472152](https://doi.org/10.2307/2346178)
- SHERMAN, L. W., MILTON, C., KELLY, T. V. and MACBRIDE, T. F. (1973). *Team Policing: Seven Case Studies*. Police Foundation, Washington, DC.

SIEGLER, K. (2013). After years of violence, L.A.'s Watts sees crime subside.

SKOGAN, W. G. (2006). *Police and Community in Chicago: A Tale of Three Cities*. Oxford Univ. Press, Oxford.

STREETER, K. (2014). In Jordan Downs housing project, police are forging a new relationship.

TREMBLAY, A., HERZ, D., ZACHERY, R. and KRAUS, M. (2020). The Los Angeles Mayor's Office of Gang Reduction and Youth Development Comprehensive Strategy. GRYD Research Brief No. 1.