Pre-Print Version: Journal of Experimental Criminology 2017

DOI : 10.1007/s11292-017-9308-0

Quasi-Experimental Designs for Community-Level Public Health Violence Reduction Interventions: A Case Study in the Challenges of Selecting the Counterfactual

Caterina G. Roman*

Hannah J. Klein

and

Kevin T. Wolff

*__Corresponding author contact information__:

Caterina G. Roman, PhD
Associate Professor, Temple University
Department of Criminal Justice
1115 Polett Walk, 5th Floor, Gladfelter Hall
Philadelphia, PA 19122
croman@temple.edu
215-204-1025

Acknowledgments: Data collection and evaluation of Philadelphia CeaseFire was supported by Award No. 2012-PB-FX-K004, from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. The content of this paper, however, is solely the responsibility of the authors and does not necessarily represent the official views of the U.S. Department of Justice.

Quasi-Experimental Designs for Community-Level Public Health Violence Reduction Interventions: A Case Study in the Challenges of Selecting the Counterfactual

Abstract

Objectives: We highlight the importance of documenting the step-by-step processes used for the selection of comparison areas when evaluating a community-level intervention that targets a large-scale community.

Methods: We demonstrate the proposed method using a propensity score matching framework for an impact analysis of the Cure Violence Public Health model in Philadelphia. To select comparison communities, propensity score models are run using different levels of aggregation to define the intervention site. We discuss the trade-offs made.

Results: We find wide variation in documentation and explanation in the extant literature of the methods used to select comparison communities. The size of the unit of analysis at which a community is measured complicates the decision processes, and in turn, can affect the validity of the counterfactual.

Conclusions: It is important to carefully consider the unit of analysis for measurement of comparison communities. Assessing the geographic clustering of matched communities to mirror that of the treated community holds conceptual appeal and represents a strategy to consider when evaluating community-level interventions taking place at a large scale. Regardless of the final decisions made in the selection of the counterfactual, the field could benefit from more systematic diagnostic tools that document and guide the steps and decisions along the way, and ask: "could there have been another way of doing each step, and what difference would this have made?" Overall, across community-level evaluations that utilize quasi-experimental designs, documentation of the counterfactual selection process will provide a more fine-grained understanding of causal inference.

Keywords: Counterfactual, Evaluation, Place-based, Propensity score matching, Violence

INTRODUCTION

In recent years, there has been an increasing trend toward the development of violence prevention and intervention efforts directed at the population or community level (Haegerich, Mercy and Weiss 2013). This trend, which includes understanding the social context of the at-risk behavior and hence, not only targeting efforts to the individual but also to the community, has drawn on strategies that are designed to change the culture of the harmful behavior (Holder, Treno, Saltz and Grube 1997; Mozaffarian et al. 2013). For urban communities, this is the street culture that embraces violent responses, such as shootings, as the norm.

These types of prevention strategies—in both the fields of public health and criminal justice—have theories of change that anticipate aggregate-level impacts, measured at the community level because they target the entire community, not just a specific set of individuals living in the target area (Holder et al. 1997). The theory of change encapsulates multiple components that are designed to complement each other, such as direct intervention with high risk youth, media activities, and the building of police and community-based partnerships to increase capacity for change. Public health and criminal justice scholars have been vocal about the difficulties inherent in rigorously evaluating large community-level prevention strategies and intervention programs and strategies (Farrell, Henry, Bradshaw, and Reischl 2016; Flay et al. 2005; Holder et al. 1997; Wellford, Pepper, and Petrie 2005).

Randomized controlled trials (RCTs) often are not possible in community-level evaluation for political, practical, operational, and technical reasons (Farrell et al. 2016; Flay et al. 2005). However, rigorous quasi-experimental designs (QXD) can replicate some of the strengths of an RCT by minimizing unobserved heterogeneity through statistical modeling that balances the treatment and comparison groups. Although this may incur additional limitations, statistical modeling to isolate causal effects in community-level prevention is necessary in today's evidence-based policy discourse. For instance, the U.S. Centers for Disease Control and Prevention (CDC) has recognized the dependence on quasi-experimentation across their community-level violence prevention efforts and has begun encouraging rigorous designs within quasi-experimental frameworks (CDC 2010). For their National Centers of Excellence in Youth Violence Prevention (YVPCs) portfolio (see Matjasko, Massetti, & Bacon, 2016 for more information), CDC required that grantees implement community-wide interventions with one or more appropriately matched comparison communities.

For QXDs to have rigor, the selection of the comparison group must provide a valid estimate of the counterfactual outcome for the treatment (Cook and Campbell 1979). This becomes even more important when there is only one area receiving the treatment. A method designed to help approximate the conditions of a controlled experiment and assist with causal inference is matching—which encompasses a broad category of techniques. Matching, in theory, reduces the imbalances between the treatment and comparisons conditions and helps improve efficiency and eliminate bias (Ho et al. 2007; Stuart 2010). Propensity score matching (PSM) (Rosenbaum and Rubin 1983), which has become common across criminal justice evaluation studies that assess the success of programs targeted to individuals, also is becoming more common in evaluation of programs and strategies that have outcomes at the community-level (Apel and Sweeten 2010). However, most of the extant evaluation studies using PSM as the tool for the selection of comparison communities do not discuss and defend the many critical design decisions made before the propensity score models are estimated, let alone provide detail in the selection of variables for matching. Furthermore, the geographical unit of analysis varies widely in community or place-level evaluation studies (e.g., census tract, police beat, street block, Zip code, etc.) and each geographical unit provides unique challenges in basic measurement of theoretically important covariates, yet these challenges are not always articulated.

The relative absence of academic discussion and defense of choices in these salient decision processes limits the assessment of rigor in these studies, and stymies guidance of value to future community-level research. What does an evaluator do if data on possible cofounders are not available at the geographic unit that represents the intervention area? How do evaluators obtain reliable information on whether potential comparison communities have competing interventions? When are competing interventions a threat to the validity of the comparison? When is a particular unit of analysis chosen as the measurement level more valid than another level of measurement? Given the limited body of research that discusses the decisions researchers often face and the concomitant resolutions in achieving a valid counterfactual, this paper describes and documents a multi-step matching approach for selection of comparison areas when evaluating a community-level intervention that targets a large scale neighborhood that is atypical of neighborhoods within the larger jurisdiction. We demonstrate the approach using as a case description the impact analysis of the Cure Violence Public Health model in Philadelphia. We focus on one rarely-discussed issue associated with selecting the counterfactual—the appropriateness of unit of analysis of the "community" measures and how the unit of measure might affect the validity of the counterfactual. This issue affects the characteristics used or assessed to balance the treatment and comparison groups, and as such, we also touch on potential threats to validity in this regard. We discuss the study trade-offs made in the QXD. The intent of this paper is to shed light on the complexities of community-level evaluation of violence prevention strategies with regard to selecting a valid counterfactual in order to move toward systematic discussion and documentation of the decisions inherent in these designs.

BACKGROUND

The Cure Violence Public Health Model

A well-known strategy that addresses gun violence in high violence communities is the Cure Violence public health model for violence prevention (see http://cureviolence.org). Cure Violence, known locally in Chicago as Chicago CeaseFire, is a gun violence prevention strategy that attempts to stop deadly violence before it occurs by interrupting ongoing conflicts, working with the highest risk individuals to change behavior related to violence, and changing community norms. The model is currently being replicated in 25 cities in at least eight countries. Treating gun violence as a public health problem signifies a scientific epidemiological approach where violence is preventable, and efforts to prevent violence begin by characterizing the scope or magnitude of the problem, evaluating potential risk and protective factors and developing interventions that will affect the identified risk factors and change the processes that put individuals at risk (Satcher 1995). Typically, responses to gun violence have been based in the criminal justice system, with police-based strategies and prosecution efforts taking the central stage. In the Cure Violence model, law enforcement agencies only act as a partner to provide Cure Violence staff with data on shooting locations and overall patterns of shooting, and they are encouraged to participate in the hiring panels when staff are hired.

The Cure Violence model has three components. The first component is detection and interruption of potentially violent conflicts, which is accomplished through the use of Violence Interrupters (VIs) who are trained in conflict mediation. The VIs are members of the community who are no longer active members of the street scene, but are still knowledgeable about it. These individuals mediate brewing problems before they progress in more serious forms of violence. In particular, they work to stop retaliatory shootings after there is a shooting in the community. VIs are non-system actors who obtain information about who was involved in shootings from community members and use their credibility and connection to the community to diffuse volatile situations.

The second component of Cure Violence is identification and treatment of the highest risk youth and young adults, which involves intense case management for high risk community members. Community members who are recruited to become clients (known as “participants”) must meet at least four of seven criteria: gang-involved, major player in a drug or street organization, violent criminal history, recent incarceration, reputation for carrying a gun, recent victim of a shooting, and being between 16 and 25 years of age (Butts et al., 2015). Each participant has an Outreach Worker (OW) who works closely with him/her, checking in multiple times a week, with an expected three calls or visits each week. In addition to these regular contacts, OWs also provide case management in the form of helping participants access services or apply for jobs. In their capacity as OW, they are more than a case worker—they are available 24 hours a day to assist participants should they need help or support. Most attend probation/parole meetings with the participant, as well as other court engagements. OWs also are involved in conflict mediations.

The final component of the Cure Violence model is mobilization of the community to change norms. Activities associated with this component include holding rallies and marches within 72 hours of a shooting in the target community to get the neighbors to stand up and say, “stop the shootings,” and show that the neighborhood wants an end to the violence (Picard-Fritsche & Cerniglia, 2013). The components, taken together, are intended to produce community-level change, as high risk individuals and residents who may have a deeply rooted distrust of law enforcement and who buy into the “stop snitching” culture, become more aware of and trusting of Cure Violence staff, and are more likely to contact them to interrupt a potentially violent situation. The careful selection of the staff, coupled with targeting the most high-risk youth and young adults aids acceptance of the message that violence harms the community and that it can be acceptable to use alternative means of conflict resolution that do not involve gun violence.

Philadelphia CeaseFire

In the fall of 2012, Philadelphia CeaseFire, in collaboration with the City of Philadelphia, received a grant from the Office of Juvenile Justice and Delinquency Prevention, U.S. Department of Justice to replicate the Cure Violence public health model in North Philadelphia.[i] The grant, which funded two VIs and six OWs as well as one supervisor and part of a program manager’s salary, was designed with the goal of reducing fatal and nonfatal shootings in the Police Services Areas (PSAs) that comprise the 22nd Police District and one small portion (14 by 3 blocks wide) of a PSA in the 39th PD. PSAs are subunits of police districts, with two to four PSAs on average, representing each police district.[ii] The target area is described in the next section in more detail.

In order to increase the chances of success, Philadelphia sought to implement Cure Violence with high fidelity to the theoretical model. Philadelphia CeaseFire staff received 40 hours of training from the Cure Violence National Office when they were hired, as well as on site “booster” training sessions every quarter for the duration of the grant. In addition to the booster sessions, the CeaseFire Program Manager and Supervisor had bi-weekly calls with the National Office to discuss any issues or additional support needed. Program staff input their activities daily into a program database, designed and supported by the National Office. Ongoing support was provided by the National Office research team to ensure that the Philadelphia team was regularly entering performance measure data.

Neighborhood Target Area and Geographic Units of Measurement

Figure 1 shows the CeaseFire target area and sub-target areas. In partnership with the Philadelphia Police Department (PPD), the CeaseFire leadership team selected the 22nd PD and part of PSA 393 as the umbrella area for the program, due to high levels of gun violence originating from five hot spots. PPD leaders indicated the violence occurring in these particular hot spots was due in part to conflicts and retaliations emanating from street groups, and hence the area would be appropriate for an intervention targeting norms that support gun violence. In 2011 and 2012, the 22nd PD had the highest rate of shootings and homicides across all 22 districts in the city. To put the rate of shootings in perspective, although the population in this 3.5 square mile area represents 4.9% of the 2012 city population, shootings accounted for approximately 14.3% of all city shootings.

As shown in Figure 1, four of five of the shooting hot spots are situated within the northern part of the general target area, with the fifth hot spot contiguous with the southern border of PSA 222. These hot spot areas were designated as gun violence hot spots by the PPD and the map layer was provided to the CeaseFire Program Director in 2012 when target areas were being established. The CeaseFire street outreach team reported that these hot spots coincided with the street locations of active street groups in this area. Although staff intervene in conflicts where needed throughout the larger target area, they were trained to dedicate most of their time in these hot spots, and to recruit program participants from these areas. Furthermore, in the first year of the program, all staff focused their outreach and shooting responses in the PSAs on the northern side of the target area (see Figure 1: PSAs 221, 222 and part of 393). The northern PSAs were the initial focus given the vast size of the 22nd PD and the time it takes for a new Cure Violence/CeaseFire model to become known in the community. National Cure Violence Program Office staff, on visits to Philadelphia, remarked a number of times about the immense size of the full target area, and that CeaseFire staff should remain focused on smaller hot spot areas and follow-up regularly on mediations conducted in and around those hot spots. Figure 1 also shows that the overwhelming majority of CeaseFire conflict mediations conducted by staff throughout the two-year evaluation period were located within the northern PSAs and clustered around the hot spots. As such, the evaluation of CeaseFire examines the effect of the program on three different levels of geography: (1) The larger target area of four and half PSAs which represents the entire catchment area for the shooting responses and anti-shooting messaging; (2) the northern PSAs where most of the outreach and mediations were conducted; and (3) the five significantly smaller hot spot areas that represent the key focus for recruitment of participants and general canvassing by staff. Given the different levels and reach of the intervention, we followed Tita and colleagues' (2003) example to assess the intervention in the multiple areas that reflect the program dosage given the theory of change and associated program components. We believe this is particularly important for interventions such as CeaseFire that seek to change the norms not only of the individuals targeted with case management, but across the larger community of high risk individuals. To date, there remains very little known—beyond theoretical articulation—about the specific mechanisms at work in public health interventions such as Cure Violence programs or focused deterrence strategies. In Philadelphia, staff voiced their concerns to the evaluators that their work implementing the model beyond the northern PSAs and hot spots could dilute the positive impact they could have on the smaller target area.

Assessing the impact of the program for the different target areas can help provide insight about the intervention's reach, given staff activities across model components.

---Figure 1 about here -----

Notably, the size of the larger target area (#1 above) is much bigger than the average size of Cure Violence intervention neighborhoods operating across the country. The four PSAs and part of 393 have approximately 72,800 residents, of which roughly 33,000 are between the ages of 10 and 34. The northern PSAs have roughly 48,500 residents and 22,000 individuals ages 10 to 34. According to Skogan and colleagues (2008), the average size of Chicago CeaseFire neighborhoods that were part of their evaluation was 10,000 residents. The neighborhood targeted in Phoenix for the Cure Violence model (Hermoso Park) was roughly 1 by 1.5 miles and had a population of 12,000 (Fox, Katz, Choate and Hedberg 2015). Data shared by the Research and Evaluation Center at John Jay College of Criminal Justice show that of 18 sites in New York City with Cure Violence models, all are under one square mile in size, and have an average population of roughly 11,000, with a range from 3,500 people to 22,000 (J. Butts, personal communication, September 1, 2017). Size comparisons between New York and Philadelphia would put the northern PSAs in Philadelphia at more than twice the size of an average New York City Cure Violence site, and Philadelphia's full target area at more than 6 times the size of an average New York site.

The CeaseFire intervention area is also quite unique in its social and economic context. According to the 2013 five-year American Community Survey, the neighborhood struggles economically with high unemployment (22.1%), a large percentage of families living below the poverty line (44.0%), and a large percentage of female-headed households (68.0%). The target area is home to some of the poorest neighborhoods in the city, where average median incomes in

2013 ranged from $14,185 to 16,185 (in 2013 dollars) (Pew Charitable Trusts 2015). High poverty census tracts are clustered within space, creating a large section of the city that is highly segregated from more resource-rich areas. Even more significant, is that the neighborhoods represent a large number of census tracts responsible for Philadelphia having the highest *deep poverty* rate of the nation's 10 largest cities (Lubrano 2014). Deep poverty signifies people living below half of the poverty level. Life expectancy is also very low in these neighborhoods—in one of the neighborhoods in the target area, life expectancy is 69 years old compared to a high of 88 years in other parts of the city (Virginia Commonwealth University 2016).

Evaluating Community-Level Interventions

In order to evaluate Cure Violence or any other community-level intervention, it is necessary to identify a similarly-scaled community or communities to serve as a comparison group. This is a complex, yet rarely-discussed undertaking in community-level evaluation (Galster, Temkin, Walker and Sawyer 2004). There are numerous challenges in selecting a subsample of untreated communities that can affect the validity of treatment effects estimated. Some of these challenges are similar to those that affect interventions with individuals, but many are unique to community wide interventions. With the exception of official police incident reports and area-level demographic data from the census, data measuring important covariates such as attitudes, norms and general cultural context usually obtained through surveys, are rarely available. Even administrative data—measures representing factors associated with violence, such as presence of gangs (Bjerregaard and Lizotte 1995; Howell and Decker 1999), land use (Bernasco and Block 2009; Browning et al. 2010), concentrations of high-risk individuals (Berk, Barnes, Kurtz and Alhman 2009)—are generally difficult to acquire.

Other issues include that interventions might be placed first into communities that openly welcome the program, or it could be that the community ranks the highest with regard to need for the intervention, as is the case with Philadelphia CeaseFire. These “features” of the treatment community reduce the likelihood that a similarly structured community, to be used as a comparison, can be located. Communities are influenced by an array of macro-, meso- and micro-level forces that may limit the extent to which an analogous set of communities exists (Braveman, Kumanyika, Fielding & LaVeist 2011; Cohen, Davis, Realini 2016). The size of the treated community also can affect the likelihood of finding a valid comparison (Farrell et al. 2016). Large intervention areas that have specific risk factors associated with the spatial context—such as concentrated poverty and segregation, or spatially autocorrelated high crime levels – generate a specific setting for behavior that is difficult to replicate. At its most basic level, the credibility of any QXD rests on whether the untreated community is comparable to the treated community in every way, with the exception of the intervention (Meyer 1995).

As Pawson and Tilley (1993) argue, an understanding of the mechanisms of the intervention and the community context must be part of the evaluation research design—and, in particular, involved in the process of obtaining a valid counterfactual. In the case of CeaseFire and many similar interventions that attempt to change the street culture of violence and retaliation, there are likely important contextual factors at play in creating the violence in the targeted communities. Furthermore, recent research has shown that community crime prediction models are greatly improved when contextual factors are added to models using the past period’s crime (Taylor, Ratcliffe, Perenzin 2015). Importantly, however, the study also showed that, for homicide (a crime type particularly relevant to the current study), models using only a measure of race outperformed models using demographics plus past period’s crime.

An added level of complexity arises with community-level evaluation because “a community” is not a single unit of measure like an individual, but instead can be represented or measured by a wide variety of geographic units. A community might be a Zip code, a census tract, a school catchment boundary, a police beat, or any other administrative aggregate relevant to the intervention. Targeted communities can also be aggregations of those boundary units, such as the case with CeaseFire, where the larger intervention area represents four and a half contiguous PSAs. Ideally, the intervention would be studied at the geographic unit that defines the intervention, but this is not always possible. The size of the treatment community may constrain choices of the unit of measure, and create problems when the community does not correspond to an officially designated unit. Evaluators must also be aware of co-occurring crime-reduction interventions that are beyond “business as usual” operating in potential comparison communities. This is not an easy task; it is one that takes in-depth knowledge of the local community, which usually means evaluators must have extensive community-based and government contacts, with frequent and ongoing discussion to fully understand what existing interventions or policy changes could pose a threat to the selection of comparison areas.

Table 1 provides a list of published community-level evaluations of crime reduction initiatives that have a unit of analysis larger than a street block and have utilized a QXD with some type of matching in the selection of comparison communities. The columns in the table list the unit of analysis for the target area, specify the matching procedures (including binocular or eyeball matching) and criteria used, and indicate whether competing interventions were considered or mentioned in the publication. We include this table to illustrate the great variation in comparison community selection procedures across all aspects of the procedures.

Of the 17 studies described, 11 (65%) did not use PSM to select comparison units. If we exclude the studies where the intervention was implemented citywide (4 studies), 9 of 13 studies (69%) did not employ PSM. Across the studies that employed PSM, most matched on crime/violence in the pre-intervention year and a few basic community contextual variables from census data, but the use of census variables varied widely. With the exception of one of the Spergel studies and Tita and colleagues' study matching specifically on gang-crime, none of the studies matched on characteristics associated with the density of criminal-justice involved individuals or gangs—characteristics that all of the interventions seek to address. The majority of the studies did not directly examine whether competing interventions existed in comparison areas. Although not shown in the table, with the exception of the city-level studies and discussions about treatment diffusion, the overwhelming majority of the studies did not indicate any challenges with regard to selecting the counterfactual. Does this indicate there were few or no challenges? We do not believe so. The dearth of open and systematic dialogue about challenges hinders the pursuit of rigorous procedures. The extant research provides little direction for continued quasi-experimentation evaluating promising violence reduction models at a time when QXDs are an inevitable part of evaluation research.

In this paper we move decisions on selecting the counterfactual from behind the scenes to a transparent discussion of the complex issues facing researchers evaluating a large-scale public health intervention to reduce violence. Developing rigorous methods in this area and documenting them are important to improve their application in the evaluation of community-level strategies. This paper is not concerned with the particular matching methods used to define closeness of matches, nor does it address the worthiness of a particular design model, such as the interrupted time series model, over other approaches; these topics have been addressed elsewhere in the literature (see for example, Galster and colleagues 2004; Penfold and Zhang 2013).

METHODS

This section details data sources and measures, and discusses the methodological challenges faced and the resulting approach followed to finalize comparison group selection. We delineate tradeoffs made in deriving a counterfactual that most approximates the treated community. Then we briefly discuss the interrupted time series estimation models employed to assess the impact of the treatment.

Going into the evaluation, we planned to use a pre-test/post-test nonequivalent control group design, relying on propensity score matching to identify comparison communities at the PSA level, with the option of matching comparison communities at the level smaller than the PSA—the census tract. The city has 65 PSAs, of which five are part of the treated community. At the time that CeaseFire started, we did not know that there would be competing interventions taking place in other high violence PSAs, let alone that a new state-funded contract would soon be awarded to put additional CeaseFire outreach workers to work in hot spots in other areas of the city.

Data and Measures

As CeaseFire is a prevention model to reduce gun violence, particularly shootings, the outcomes of interest are fatal and non-fatal criminal shootings. Criminal shootings exclude officer shootings and self-inflicted shootings and are counted at the “victim” level (i.e., one perpetrator who shoots three people in the same incident equals three shootings). Address-level data for all criminal shootings were received from the Philadelphia Police Department for the period January 2003 through March 2015. The dependent variables for the following analyses are modeled as rates (per 10,000 residents).

Variables selected for the PSM treatment status model were based on theoretically important neighborhood-level variables that have been shown in the literature to be associated with levels of violence—and in particular those variables that are relevant to the type of violence CeaseFire attempts to address. We determined that it was important to include other community characteristics variables in addition to the pre-intervention level of violence for four main reasons: (1) the street gang landscape and racial/ethnic makeup of neighborhoods is greatly associated with the nature of and motivations behind homicides and nonfatal shootings in Philadelphia (Roman, Link, Mayes & Hyatt 2015; Taylor et al. 2015); (2) research has shown that levels of poverty/SES, and segregation shape the geographic extent and clustering of high violence locations (Morenoff and Sampson 1997; Pratt and Cullen 2005), (3) these variables add explanatory power over and above the previous time period's violence in models predicting violence in Philadelphia—in models predicting violence in Philadelphia using the potential confounders added to the treatment status model, the pseudo r^2 is 0.326, compared to 0.053 when just using the previous period's violence. and (4) research shows that the longer the list of potential confounders of selection into treatment, the better for the model, because when the long list of variables are shown to be in balance, one can be more confident in the validity of treatment effect estimates from the final impact models (Haviland and Nagin 2007). Furthermore, the ranking of some of the remaining PSAs with regard to gun violence rates fluctuated from year to year, making it unlikely only matching on rates of gun violence would yield valid comparison communities, particularly at the PSA level. Nine variables were used in the treatment selection models. With the exception of the US Census data, the data for all variables were geo-located, providing the research team with the ability to map the data into various units of aggregation (e.g., census tract, block group, PSA, etc.), as needed.

- The rate of shootings and robbery with a gun for the pre-intervention year (2012) is computed as the count of shootings and gun robberies divided by the resident population. We also considered and compared the output when matching on the gun violence rate averaged over a three-year pre-intervention period. Using a three-year average of gun violence produced units that did not greatly reduce the bias in important community characteristics such as the number of probationers/parolees and number of gangs in an area. The three-year average of gun violence measure also did not perform well when matching at the smallest geographic level—the hot spot. For this model, the "percent bias" statistic was not under 20—the usual rule of thumb for adequate amounts of bias. It is also important to note that matching using a three-year violence average produced neighborhoods that differed somewhat geographically (e.g., 10 to 50 percent of the matches were different depending which of the three target areas was used), but the final results assessing the impact of the intervention did not change substantively—in fact, the results were basically identical. Given the robustness of our findings to this testing, we chose to use the one-year rate of gun violence (2012) to be consistent with the selection of pre-intervention crime variables used in the published community-level evaluation (Braga, Hureau, and Papachristos 2011; Tita, Riley, Ridgeway, et al. 2003). However, given the differences in geographic locations of comparison neighborhoods found when we compared balancing variables, we believe that for future studies, researchers should carefully consider and document which measures accurately capture the essence of the target area. As stated in the introduction—the counterfactual should be similar in every way (to the extent possible) to the treatment area, minus the intervention.

- Policing activity is a scale derived from obtaining the z-score of the rate of car and pedestrian stops by the PPD for 2012. Z-scores less than -0.5 received a score of 0 on the scale, -0.5 to 0.499 received a score of 1, 0.5 to 1 receives a 2, and z-scores greater than 1 received a score of 3. These data were provided to a colleague under a strict data agreement with PPD. The colleague then mapped the data onto the respective geographic unit and provided the authors with aggregate counts (with permission from PPD).
- Count of street gangs is derived from the geo-located set spaces of all street gangs in Philadelphia as of 2012. These data were obtained from law enforcement focus group meetings designed to map street gang membership across the city of Philadelphia.
- Count of active probationers/parolees is the aggregate count of the home location of probationers/parolees in 2009-2010. These data, obtained from the Philadelphia Adult Probation and Parole Department, were provided to a colleague under a strict data agreement. The colleague then mapped the data onto the respective geographic unit and provided the authors with aggregate counts (with permission from APPD).
- The remaining variables were derived from the American Community Survey data for 2007 to 2011: Concentrated disadvantage is the sum of z-scores for public assistance, unemployment, poverty, and female-headed-households divided by four; residential stability is sum of z-scores for the percentage of homeowners residing in home for last five years and the percentage of households that are owner-occupied divided by two; percentage of population that identifies as any part Black; percentage of the population that is Hispanic, and total population.

Propensity Score Matching

PSA as the Unit of Analysis

As mentioned earlier, the original intent was to specify the treatment status model for the propensity score matching at the PSA level because this unit represents the sub-boundaries in which the CeaseFire staff worked. We recognize that the PSA level is not ideal for the unit of measure for matching because the intervention, as a whole, is situated in the larger police district. However, the limited number of police districts, which vary greatly by in levels of violence and demographic characteristics, made it unfeasible to identify matches at this large level of aggregation. At the PSA level, we began by running the psmatch2 routine with no replacement for exact matching in Stata 14.0, because exact matching, or 1 to 1, is considered more ideal than matching with replacement (Imai, King and Stuart 2008; Nagin, Cullen and Jonson 2009). Because there were only 60 PSAs available as a pool for potential comparison PSAs (Philadelphia has 65 PSAs in total), we did not exclude any PSAs from the first pass on the matching routine. In addition to excluding areas where competing interventions occurred, researchers often exclude areas contiguous to the treatment because some place-based interventions may experience diffusion of the intervention, in that the program seeps into surrounding geographic areas given the nature of strategy (e.g., the outreach workers may diffuse conflicts right outside the boundaries of the treated area). We knew that, depending on outcome of the treatment status model, we would have to exclude contiguous PSAs because OWs had conducted conflict mediations in some of these areas, and exclude the PSAs in the 24th and 25th Police District where the CeaseFire expansion took place. We potentially would also need to exclude 8 PSAs in three Police Districts in South Philadelphia where the Focused Deterrence gun violence reduction initiative was taking place simultaneously.

To test the appropriateness of matches derived from PSM, post tests were conducted in Stata using the *pstest* command. Examination of the reduction in bias from the matches selected showed that the routine was not successful, in that it did not reduce the standardized bias across all variables (see Table 2). The standardized bias statistics represent the mean difference as a percentage of the average standard deviation between the treated and the untreated (Rosenbaum and Rubin 1985). In fact, bias on the matches was greater than the bias for the unmatched areas, and significant differences remained between the treatment and comparison PSAs. This was not a surprising finding in that the CeaseFire PSAs are extreme outliers on the rate of gun violence and concentrated disadvantage. Matching with replacement was also attempted, but this too did not yield an adequate group of comparison PSAs.

---- Table 2 about here ----

Census Tracts as the Unit of Analysis

Next, we turned to the next smaller unit of analysis—the census tract—as the matching unit. The CeaseFire treatment area is comprised of 23 census tracts, representing six percent of the 384 census tracts in Philadelphia. Before matching was performed, we excluded the tracts that were contiguous to the 23 tracts representing the treatment. A one-to-one matching routine was used. As Table 3 highlights, the standardized bias statistic is relatively low for most of the covariates. For the 2012 rate of gun violence, the standardized bias between the treated and control census tracts is larger than 20 percent, a threshold used to determine that the bias reduction was adequate (Caliendo and Kopeing, 2008). In addition to the rate of gun violence, the standardized bias for concentrated disadvantage is larger than 20, indicating that the matched comparison census tracts still varied greatly in disadvantage from the census tracts that received CeaseFire. These differences are graphically depicted in Figure 2a. Additionally, when examining the geographic location of matches, as shown in Figure 3, it is notable that the tracts selected are spread out over the entire city and there are very few areas where selected comparison tracts are clustered together into larger geographic area that could mimic the size of the CeaseFire treatment community. This is problematic as one would be comparing a large 3.5 square mile area comprised of 23 tracts grouped together to 23 different, non-contiguous "neighborhoods" spread throughout the city. Given that the expectation in the treatment effect models will be that the outcome would "reach" the larger PSA level for the CeaseFire area, using the matched tracts could lead to underestimation of the treatment effect since one would be comparing the success of achieving violence reduction in a set of clustered tracts to a set of untreated tracts not clustered.

---Table 3 about here ---

---Figure 2 about here ---

---Figure 3 about here ---

In search of balancing variables that would likely result in tracts with some clustering, we re-ran the matching routine to include a balancing variable to represent the spatial autocorrelation of concentrated disadvantage. The GeoDa software (version 1.6.7) was used to calculate the lag of concentrated disadvantage with queen contiguity chosen to calculate the weights matrix. We did this under the assumption that the concentration of poverty and extent of economic segregation contributes to the chronically high levels of violence that have plagued the CeaseFire intervention area. However, the addition of this variable did not result in the selection of tracts with a much higher degree of clustering (i.e., selected comparison tracts contiguous to other selected tracts).[iii]

Census Block Groups as the Unit of Analysis

We then theorized that matching on a unit smaller than the census tract may provide output more representative of the larger geography of the CeaseFire area. This is partially due to the modifiable area unit problem (MAUP)—that any observed aggregated values will vary depending on the unit to which the data is aggregated. Essentially, different impressions of the underlying pattern we are trying to replicate perhaps could be created by using alternative aggregations. With this in mind, we ran a matching routine at the census block group level. The CeaseFire treatment area represents 72 block groups out of 1,336 block groups in Philadelphia, or 5.3% of block groups in the city. We excluded the areas contiguous to the treatment area, but to understand the possible issues with competing interventions, we initially left in the block groups that were receiving similar or competing interventions. Interestingly, the matching routine did not select any block groups that had competing interventions (even the block groups in the 24th and 25th Police Districts, which had very high levels of violence in 2012). Block groups in this area were likely not good matches because the neighborhoods comprising them have a very different demographic makeup and gang landscape compared to the CeaseFire target area.

Matching at the block group level yielded respectable matches with regard to balance. As Table 4 shows, characteristics of the treated units were not significantly different from the untreated matches at the block group level, with only percent Hispanic nearing statistical significance (p=.10). Figure 4 indicates that the reduction in bias improved when compared to the output from matching at the census tract level (Figure 2b). The map of the geographic locations of the matched block groups (Figure 4) also shows that the matched comparison block groups tended to cluster together more than the matched census tracts. There are also vast differences between the locations of the matched units when comparing the census tract map (Figure 3) to the block group map (Figure 4). Examining these differences and the clustering, and recognizing that reliance on the smaller unit of aggregation might be the better unit to capture the processes that are associated with hot spots of shootings (which led to the selection of the treatment areas), we decided to utilize the block group level matching output to select comparison areas. When overlaying the PSA boundaries, there were five PSAs comprised of at least five block groups output in the matching. To select the final comparison communities, we then selected those five PSAs (161, 162, 182, 192, and 353) to represent the counterfactual. Essentially, the block group matches identified the PSAs which had the largest number of matched block groups, and then we used all block groups within those 5 PSAs as the counterfactual. The selected PSAs were also carefully vetted with the Philadelphia Police Department and city leaders to determine whether any simultaneous interventions were taking place contemporaneously with CeaseFire; it was determined that there were no competing interventions in the five PSAs.

---Table 4 about here ----

---Figure 4 about here ----

Although not shown, selecting the PSAs from block group matches is not a perfect tradeoff—the characteristics of these PSAs are not fully balanced across the confounders when examining the block group characteristics for treated and untreated matched areas (72 treatment block groups compared to the 167 blocks that make up the 5 matched PSAs)—the rate of gun violence and concentrated disadvantage remains, on average, significantly higher in the treated units. However, given the place-based nature of the intervention, we erred on the side of geographic size, under the belief that we should attempt to mirror the size (as much as possible) of the larger CeaseFire treatment area because geographic units in close proximity (or better yet, contiguous) tend to be more alike, in many aspects that the available data cannot capture (Cohen, Inagami, and Finch 2005; Taylor 1995). Our other choice would have given us non-contiguous block groups (with an average population of 4,850) from a wider range of PSAs distributed throughout the entire city of Philadelphia.

We followed the same framework as outlined above to select the comparison communities for the northern treatment area. We used block groups as the unit of measure (the two and one half PSAs roughly equate to 46 block groups) to select those PSAs that had at least five block groups clustered within a PSA—this yielded 2 PSAs (161 and 162) as the comparison communities. For the hot spot areas, because the size of the hot spots were smaller than the size of a PSA, we directly selected the block groups that comprised the hot spot areas and the block groups that were matched as the units on which to run the treatment effect models. For both of the matched comparison units, an examination of the bias statistics indicated that balance between the treatment and comparison units was achieved. These post-match bias reduction statistics are shown in Tables 5 and 6. There were no significant differences in characteristics across treatment and control units. Four characteristics in the matching for the Northern PSAs treatment area had bias statistics right around 20 (gun violence rate, gang count, pedestrian/car stops, and total population), which is the general cutoff for good matches. Importantly, however, the PSAs in which these comparison block groups fall (PSAs 161 and 162) were most proximal to CeaseFire PSAs in the rankings for shootings in 2011 and 2012. For both the Northern PSA and hot spot levels of geography, we did not have to pre-exclude block groups that had competing interventions, because, similar to the analysis for the larger target area, none were output as best matches in the PSM treatment status model.

ARIMA Time Series Modeling

To assess the impact of the treatment, we employed AutoRegressive Integrated Moving Average (ARIMA) models which account for temporal dependencies of time series data. Using Stata 13.0 software, we relied on a three part strategy to estimate ARIMA models: (1) each series was transformed, if needed, to reduce any bias or trend that might have been found in the changing time series; (2) an appropriate transfer function was chosen to the assess the impact in the series; and (3) diagnostic checks were performed to check for temporal autocorrelation within the residuals. The first, and arguably the most important step in constructing a time-series model is to assess stationarity and examine temporal autocorrelation (Chatfield, 2004). Based on initial unit root tests run on the pre-intervention trend, and the examination of the autocorrelation function (ACF) and partial autocorrelation function (PACF) results, it is very unlikely that violence in Philadelphia during sample period should be modeled using first and seasonal differences, since there is no evidence of unit roots. Second, zero-order immediate and permanent transfer functions were utilized, testing the hypothesis that CeaseFire, fully implemented by April 30, 2013, would result in an immediate and continued impact over time (the post-implementation period consists of 24 months). Types of ARIMA processes—autoregressive (AR) and moving average (MA) parameters—were estimated with the pre-policy change data series based on the examination of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) in order to identify the best fitting model. A final model with a chosen AR term and an appropriate MA term was then selected, and the intervention component was added to the full model.

Looking at the goodness of fit measures for the ARIMA time series procedure (i.e. the AIC and BIC scores), the best fit models were as follows: for the larger CeaseFire area, the best fit model was (1,0,0) while for the larger comparison area the best fitting specification was (2,0,0). For the CeaseFire Northern PSA target area the most appropriate model was also (1,0,0) and the best fit model for the comparison was (0,0,0). Finally, residuals were checked for normality and independency using diagnostics measures, the Ljung–Box Q tests were used to test for residual autocorrelation at various lags up to 12 months (Ljung and Box, 1978).

TREATMENT EFFECT FINDINGS

The ARIMA time series models estimating the treatment effect of CeaseFire on the monthly rate of shootings (per 10,000 population) are presented in Table 7 for: (1) the full treatment area and associated matched comparison areas (i.e., the PSAs that were selected using matched blocked groups), (2) the Northern PSAs that represent the area of focus for the CeaseFire workers and the corresponding matched comparison areas (i.e., PSAs selected using matched block groups), and (3) the hot spot areas and the matched comparison hot spot areas (i.e., block groups). Each of the three sets of CeaseFire areas exhibited a significant reduction in shootings after the implementation of CeaseFire. However, for both the larger treatment area and the northern PSAs, the comparison areas also witnessed a significant decrease in shootings. Only for the gun crime hot spots does it appear that the impact was unique to CeaseFire, as the matched comparison areas did not exhibit a significant reduction in the rate of shootings after CeaseFire began.

It is interesting that across the different treatment areas modeled to test the intervention, only in the smallest unit of analysis—the hot spots—was there a clear cut finding of an effect that the paired comparison areas did not witness. This could be due to the nature of the intervention itself, in that the CeaseFire staff conducted most of the conflict mediations and outreach to high risk and gang youth in these hot spot areas and in relation to high risk individuals who have street gang territories that align with the hot spots, and that any wider "spread" of the intervention or norm change theorized in the Cure Violence model was not achieved. In addition, we know that the comparison communities for the gun crime hot spots were nicely matched (see Table 6); we did not have to make tradeoffs in aspects of validity—we did not aggregate up to PSAs which would have included some block groups that were not directly matched in the PSM models. For the hot spot treatment effect models we can be more confident that we ruled out differences in unobserved community characteristics that could have caused us to understate program effects. The CeaseFire community chosen in Philadelphia has characteristics that make these communities perhaps less likely to want the treatment (e.g., program fatigue, enduring street groups with family members and residents who support "code of the street" norms), and hence, respond to it, but also more in need of the treatment—features which could lead to understating the effects if the treatment and comparison communities are not well-matched.

---Table 7 about here ---

DISCUSSION AND CONCLUSION

The Cure Violence Public Health model of gun violence prevention offers communities a promising strategy that, given the theory of change and evidence from some past evaluations of the model (Skogan et al. 2008; Webster et al. 2012), has potential to effect community-level change. In a perfect world, evaluation methods should rise to the occasion. Comparison communities are useful sources of information about expected outcomes for treated communities, but only when they closely resemble the pre-treatment attributes of those communities. The current study demonstrates that geographically large treatment areas reduce the likelihood that PSM methods result in valid comparison groups, as evidenced by unbalanced sets of units across treatment and comparison groups when the PSM model was conducted at large units of aggregation. We propose an alternative solution by conducting the matching using smaller units—the census block group, then looking for clusters of matched block groups that, together, form larger community areas—PSAs. The study showed that the comparison units output when the matching was performed at the block-group level were balanced. We used this information to inform the selection of PSAs as the final comparison group for two of three target areas. The PSAs chosen as the counterfactual for the largest treatment area and the northern treatment ranked almost as high on gun violence rates as those representing the target area. Across the three different sizes of target areas studied, the treatment and control groups were well matched on both number of gangs and number of probationers/parolees—two measures not typically used in the extant studies (and not used in the studies in Table 1), but we believe well worth considering for delivering a valid counterfactual to an intervention that targets gun violence associated with gangs and the street code of violence.

We recognize that there will be fewer potential matches as the number of community characteristics one uses for the matching increases, but the current study erred on the side of inclusion, using nine relevant measures in the PSM treatment status model. The extant communities and crime literature supports the assertion that contextual variables play a critical role in conditioning violence, and coupled with the logic of the intervention at hand, suggest inclusion of these measures. Furthermore, past studies indicate that it is more harmful to exclude potential confounders than to include variables that are not associated with treatment assignment (Stuart 2010). Given paper length constraints, coupled with the extent of analyses needed to describe processes related to our primary focus on target area size and unit of analysis, we did not take on additional testing to examine the robustness of our findings against other sets of variables that could be utilized in the matching process (e.g., matching only on pre-intervention levels of violence). This is a noted limitation of our study. We did however, examine if our findings held up when using a three-year average of gun violence instead of a one-year measure of violence. Although the treatment effect findings remained the same, the differences by geography (i.e., location) in output of comparison block groups should be noted. Future research should examine the potential issue of how hidden bias not accounted for in the treatment status model might affect findings (Loughran, Wilson, Nagin and Piquero 2015).

The finding that there were vast differences in matched comparison units with regard to geographic location across Philadelphia (comparing Figures 3 and 4) between census tract matching and block group matching has implications for studies that have options for the unit of analysis. Evaluations of community-based violence reduction strategies operating at larger than the block group level may have multiple options. Choosing the unit because of convenience should not guide a matching strategy. A discussion of options and documentation of decisions should become standard in publications. In the introduction to this paper we asked: “When is a particular unit of analysis more valid than another?” Although there is no definitive answer to this question, we can likely agree that conducting multiple sets of tests and assessing robustness of findings to the various tests and procedures used could help make a strong case for the validity of the QXD.

The review provided in Table 1 found that there is little published research on methods to guide the process of selecting the counterfactual in community-level evaluation, and no consensus on how matching should be performed or evaluated. If we transfer the lessons from evaluation studies focused on individual-level change, the field can generally agree, however, that better matching equates to better quality of resulting inferences. When there are enough units to conduct PSM, detailed attention should be given to the treatment status model. Evaluators should ask: "Could there have been another way of doing each step and what difference would this have made?" The steps include: (1) choosing the unit of analysis, (2) consideration of potential confounders, (3) collecting data that adequately operationalizes them, (4) assessing geographic location of matches; and (5) addressing issues with regard to potential competing interventions.

It is important to note that the current analysis was limited to geographic areas within the city of Philadelphia. An alternative approach to identifying the counterfactual could have included additional urban geographies within the larger metropolitan area as potential matches. Doing so may have increased our chances of finding suitable matches at larger levels of aggregation based on our key matching characteristics. However, including proximate areas such as the city of Camden (a historically disadvantaged and violent place) would have introduced several significant barriers. First, obtaining a rich set of matching variables would have been extremely difficult. Although data at the jurisdiction-level is made available from the FBI's Uniform Crime Reporting program, address-level or block-level data are often painstaking to obtain. Smaller cities often lack the resources to prepare such data in a timely matter and obtaining sensitive data on violence which is geographically identified could be met with hesitation. It would be incumbent on the evaluators to verify, to the extent possible, that the outcome measures collected for different jurisdictions, are collected in the same way, and more importantly, measuring the same underlying construct. Second, even evaluators have access to data for outlying cities or jurisdictions, assuring that any areas matched from the surrounding jurisdictions are an appropriate counterfactual would present additional challenges due to possible differences in crime policy and law enforcement strategies and tactics present between jurisdictions. If, for example, Camden had been working to combat crime and violence using other techniques, it would be necessary to account for those between-area differences in a meaningful way to ensure that business as usual was, indeed, business as usual. This concern presents an additional challenge for researchers, especially those assessing programs in smaller jurisdictions (i.e., with fewer potential matching units) or programs that are implemented jurisdiction-wide.

In this study, we provide and defend the reasons behind a variety of decisions made in these steps, with emphasis on the importance of the size of the unit of analysis and the geographic locations of matched comparison units. We acknowledge that there are other considerations not fully explored, such as the outcome of the PSM if we had not stuck with one-to-one matching, or the utilization of newer techniques such as the use of synthetic controls to reduce bias (Saunders, Lundberg, Braga et al. 2015). But, given space constraints we did not discuss them; we hope that future studies will. Indeed, documentation of these steps and processes will provide a more fine-grained understanding of causal inference within community-level evaluations that utilize quasi-experimental designs.

References

Abbott, A. (2001). *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.

Apel, R.J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. In A.R. Piquero & D. Weisburd, (Eds.), *Handbook of Quantitative Criminology* (pp. 543–562). New York: Springer.

Berk, R., Sherman, L., Barnes, G., Kurtz, E., & Ahlman, L. (2009). Forecasting murder within a population of probationers and parolees: A high stakes application of statistical learning. *Journal of Royal Statistical Society*, 172(1), 191-211.

Bernascco, W. & Block, R. (2009). Where offenders choose to attack: A discrete choice model of robberies in Chicago. *Criminology*, 47, 91-130.

Bjerregaard, B. & Lizotte, A.J. (1995). Gun ownership and gang membership. *The Journal of Criminal Law and Criminology,* 86(1), 37-58.

Boyle, D.J., Lanterman, J.L., Pascarella, J.E., & Cheng, C-C. (2010). The impact of Newark's Operation Ceasefire on Trauma Center Gunshot Wound Admissions. *Justice Research and Policy*, 12(2), 105-123.

Braga, A.A. (2008). Pulling levers focused deterrence strategies and the prevention of gun homicide. *Journal of Criminal Justice*, 36(4), 332-343.

Braga, A.A., Hureau, D.M., & Papachristos, A.V. (2011). An ex post facto evaluation framework for place-based police interventions. *Evaluation Review*, 35(6), 592-626.

Braveman, P. Kumanyika, S. Fielding, J. & LaVeist, T. (2011), Health disparities and health equity: The issue is justice. *American Journal of Public Health*, 101(Suppl 1), S149–S155.

Browning, C.R., Byron, R.A., Calder, C.A., Krivo, L.J. &, Kwan, M-P. (2010). Commercial density, residential concentration, and crime: Land use patterns and violence in neighborhood context. *Journal of research in Crime and Delinquency*, 47(3): 329-357.

Butts, J.A., Roman, C.G., Bostwick, L., & Porter, J.R. (2015). Cure Violence: A public health model to reduce gun violence. *Annual Review of Public Health*, 36, 39-53.

Caliendo, M. & Kopeining, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31-72.

Centers for Disease Control and Prevention [CDC] (2010). Cooperative agreement program for the national academic centers of excellence in youth violence prevention. http://www.grants.gov/search/search.do?mode=VIEW&oppId=52740.

Chatfield, C. (2004). *The analysis of time series: An introduction.* New York: CRC Press.

Cohen, L., Davis, R., & Realini, A. (2016). Communities are not all created equal: Strategies to prevent violence affecting youth in the United States. *Journal of Public Health Policy*, 37, S81–S94.

Cohen, D. A., Inagami, S., & Finch, B. (2008). The built environment and collective efficacy. *Health & Place*, 14(2), 198–208. http://doi.org/10.1016/j.healthplace.2007.06.001.

Cook, T.D. & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis for field settings*. Chicago, IL: Rand McNally.

Corsaro, N., Brunson, R.K., & McGarrell, E. (2013). Problem-oriented policing and open-air drug markets examining the Rockford Pulling Levers Deterrence Strategy. *Crime and Delinquency*, 59(7), 1085-1107.

Corsaro, N., & McGarrell, E. (2010). An evaluation of the Nashville Drug Market Initiative (DMI) Pulling Levers Strategy. East Lansing, MI: Michigan State University, School of Criminal Justice.

Corsaro, N. & Engel, R.S. (2015). Most challenging of contexts: Assessing the impact of focused deterrence on serious violence in New Orleans. *Criminology and Public Policy*, 14, 471-505.
Engel, R.S., Tillyer, M.S., & Corsaro, N. (2013). Reducing gang violence using focused Deterrence: Evaluating the Cincinnati Initiative to Reduce Violence (CIRV). *Justice Quarterly*, 30, 403-439.

Farrell, A. D., Henry, D., Bradshaw, C., & Reischl, T. (2016). Designs for evaluating the community-level impact of comprehensive prevention programs: Examples from the CDC Centers of Excellence in Youth Violence Prevention. *The Journal of Primary Prevention*, 37, 165–188. http://doi.org/10.1007/s10935-016-0425-8

Flay, B. R., Biglan, A., Boruch, R. F., Gonza´lez Castro, F., Gottfredson, & D., Kellam, S. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science*, 6, 151–175. doi:10.1007/s11121-005-5553-y.

Fox, A. M., Katz, C.M., Choate, D. E., & Hedberg, E. C. (2015). Evaluation of the Phoenix, TRUCE Project: A replication of Chicago CeaseFire. *Justice Quarterly*, 32, 85-115.

Galster, G., Temkin, K., Walker, C., & Sawyer, N. (2004). Measuring the impacts of community development initiatives. *Evaluation Review*, 28 (6), 1–38.

Haegerich T.M., Mercy, J., & Weiss, B. (2013). What is the role of public health in gang-membership prevention? In: Simon T., Ritter N., Mahendra R., editors. Changing Course: Preventing Gang Membership. Washington, D.C.: U.S Department of Justice, Office of Justice Programs, and the Centers for Disease Control and Prevention.

Ho, D., Imai, K., King, G. & Stuart, E.A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, 15, 199–236. http://gking.harvard.edu/files/abs/matchp-abs.shtml.

Holder, H.D, Treno, A.J., Saltz, R.F., & Grube, J.W. (1997). Recommendations and experiences for evaluation of community-level prevention programs. *Evaluation Review*, 21, 268-278.

Howell, J.C. and Decker, S.H. (1999). The youth gangs, drugs, and violence connection. Washington, DC: OJJDP Juvenile Justice Bulletin.

Imai K., King G., & Stuart E.A. (2008). Misunderstandings among experimentalists and observationalists in causal inference. *Journal of the Royal Statistical Society Series A*, 171(2), 481–502.

Joyce, N. (2016). The Philadelphia Police Department: Moving into the 21st Century, 2008-2015. Philadelphia, PA: Philadelphia Police Department. Available: https://www.phillypolice.com/assets/directives/MovingThePhiladelphiaPoliceDepartmentIntoThe21stCentury.pdf. Accessed January 13, 2017.

Ljung, G.M. & Box, G.E.P. (1978). Measure of lack of fit in time-series models. *Biometrika*, 65, 297–303.

Loughran, T. A., Wilson, T., Nagin, D. S., & Piquero, A. R. (2015). Evolutionary regression? Assessing the problem of hidden biases in criminal justice applications using propensity scores. *Journal of Experimental Criminology*, 11(4), 631-652.

Lubrano, A. (2014, September 25). Philadelphia rates highest among top 10 cities for deep poverty. *The Philadelphia Inquirer*. Available: http://www.philly.com/philly/news/20140925_Phila_s_deep_poverty_rate_highest_of_nation_s_10_most_populous_cities.html

Matjasko, J. L., Massetti, G. M., & Bacon, S. (2016). Implementing and evaluating comprehensive evidence-based approaches to youth violence prevention: Partnering to create communities where youth are safe from violence. *The Journal of Primary Prevention*, 37, 109-119.

McGarrell, E.F., Chermak, S., Wilson, J.M., & Corsaro, N. (2006). Reducing homicide through a “lever-pulling” strategy. *Justice Quarterly*, 23(2), 214-231.

Meyer, B.D., (1995). Natural and quasi-experiments in economics. *Journal of Business and Economic Statistics*, 13(2), 151-162.

Morenoff, J.D. & Sampson, R.J. (1997). Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970-1990. *Social Forces*, 76, 31-64.

Mozaffarian D., Hemenway D., & Ludwig, D.S. (2013). Curbing Gun Violence: lessons from public health successes. *Journal of the American Medical Association*, 309(6), 551-552. doi:10.1001/jama.2013.38

Nagin, D.S., Cullen, F.T. & Jonson, C.L. (2009). Imprisonment and reoffending. In M. Tonry (Ed.), *Crime and Justice*: *A Review of Research, Volume 38* (pp. 115-200). Chicago: University of Chicago Press.

Papachristos, A.V., Meares, T.L., & Fagan, J. (2007). Attention felons: Evaluating Project Safe Neighborhoods in Chicago. *Journal of Empirical Legal Studies,* 4(2), 223-272.

Pawson, R., & Tilley, N. (1994). What works in evaluation research? *British Journal of Criminology,* 34(3), 291–306

Penfold, R.B., & Zhang, F. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. *Academic Pediatrics,* 13(6), S38-S44.

Pew (2015). Philadelphia 2015: State of the city. Philadelphia, PA: Pew Trusts. Retrieved from http://www.pewtrusts.org/~/media/assets/2015/11/2015-state-of-the-city-report_web_v2.pdf

Picard-Fritsche, S. & Cerniglia, L. (2010). Testing a public health approach to gun violence: An evaluation of Crown Heights Save Our Streets, a replication of the Cure Violence Model. New York: Center for Court Innovation.

Roman, C., Link, N., Mayes, L., & Hyatt, J. (2015). Gang-involvement in shootings: Preliminary analysis of gang-involved shootings in Philadelphia, 2009-2015. Report to the Focused Deterrence Executive Committee. Philadelphia, PA: Temple University.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33–38.

Satcher, D. (1995). Violence as a public health issue. *Bulletin of the New York Academy of Medicine*, 72(1), 46-56.

Saunders, J., Lundberg, R., Braga, A.A., Ridgeway, G., Miles, J. (2015). A synthetic control approach to evaluating place-based crime interventions. *Journal of Quantitative Criminology*, 31, 413-434. doi:10.1007/s10940-014-9226-5

Skogan, W.G., Hartnett, S.M., Bump, N., & Dubois, J. (2008). Evaluation of CeaseFire Chicago. Report for Office of Justice Programs, US Department of Justice.

Spergel, I.A., Wa, K.M., & Sosa, R.V. (2005). Evaluation of the Mesa Gang Intervention Program (MGIP). Chicago, IL: School of Social Service Administration, The University of Chicago.

Spergel, I.A., Wa, K.M., and Sosa, R.V. (2005). Evaluation of the Riverside Comprehensive Community-Wide Approach to Gang Prevention. Chicago, IL: School of Social Service Administration, The University of Chicago.

Stuart, Elizabeth A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25(1), 1–21.

Taylor, R. B. (1995). The impact of crime on communities. *The Annals of the American Academy of Political and Social Science,* 539(1), 28-45.

Taylor, R.B., Ratcliffe, J.H., & Perenzin, A. (2015). Can we predict long-term community crime problems? The estimation of ecological continuity to model risk heterogeneity. *Journal of Research in Crime and Delinquency*, 52(3), 635-657.

Tita, G.E., Riley, K.J., Ridgeway, G., Grammich, C., Abrahamse, A., & Greenwood, P.W. (2003). *Reducing Gun Violence: Results from an Intervention in East Los Angeles*. Santa Monica, CA: RAND Corporation.

Virginia Commonwealth University (2016). Philadelphia life expectancy methodology and data table. Richmond, VA: Virginia Commonwealth University, Center on Society and Health. Accessed March 2, 2016. http://societyhealth.vcu.edu/media/society-health/pdf/LE-Map-Philly-Methods.pdf

Webster, D.W., Whitehall, J.M., Vernick, J.S., & Parker, E.M. (2012). Evaluation of Baltimore's Safe Streets Program: Effects on attitudes, participants' experiences, and gun violence. Baltimore, MD: Johns Hopkins Bloomberg School of Public Health.

Wellford, C., Pepper, J. & Petrie, C. (2005). *Firearms and violence: A critical review*. Washington, DC: National Science Academies Press.

Wilson, J.M., S. Chermak, & McGarrell, E.F. (2010). Community-based violence prevention: An assessment of Pittsburgh's One Vision One Life Program. Santa Monica, CA: RAND.

Figure 1. Philadelphia CeaseFire Intervention Areas and Locations of Conflict Mediations

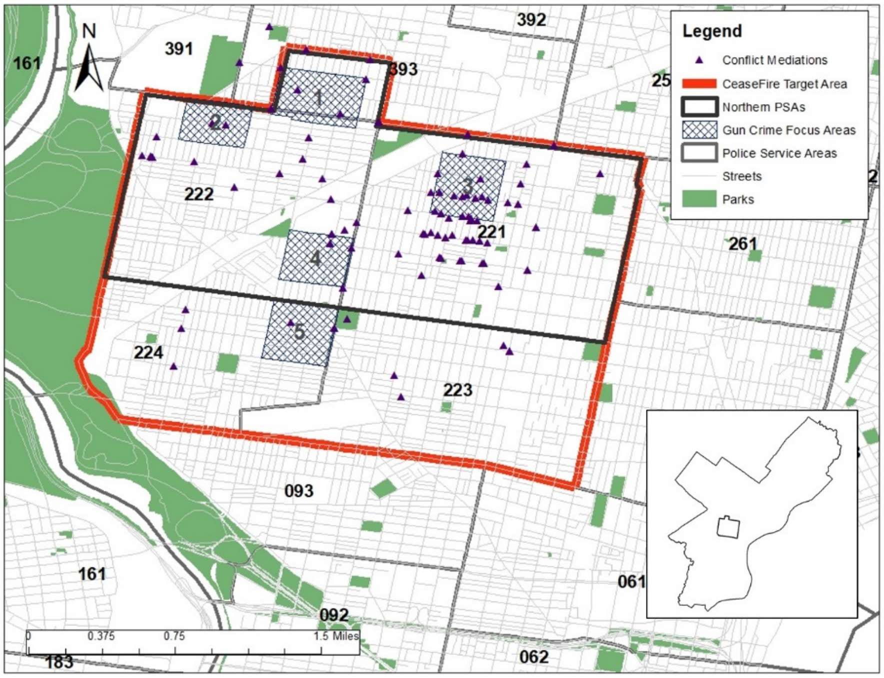

Table 1. Studies of Community-Level Violence Reduction Interventions that Utilize a Quasi-Experimental Design with Counterfactual						
Citation	Intervention// Outcome of Interest	Used PSM?	Target Area and Unit of Analysis and Utilization of Comparison Group	Matching Criteria	Did Study Discuss Threat of Competing Inter-ventions?	Stat. Significant Reductions in Outcomes in Treatment Area?
Boyle, D.J., Lanterman, J.L., Pascarella, J.E., & Cheng, C-C. (2010)	Cure Violence//guns hot wounds	No	Block groups with intervention compared to same number of block groups in same city (Newark, NJ)	Matched on gunshot wound rate but used t-tests to check the differences between population, rate, Hispanic, median income, vacant housing, poverty, and age	No	No
Braga, A.A. (2008)	Focused Deterrence//gun homicides	No	City-level (Stockton, CA) compared against 8 other cities in California	Compared on similar populations and general similarly of geographic area of the state	Yes, no other competing interventions existed	Yes
Braga, A.A., Hureau, D.M., and Papachristos, A.V. (2011)	POP//violent crime counts	Yes	Hot spots broken down into 776 treated street segments and intersections and 2,472 matched untreated areas	Matched on 2006 (pre-treatment) violent crime counts, concentration of social disadvantage in surrounding block groups, and type of street unit	No	Yes
Corsaro, N., Brunson, R.K., McGarrell, E. (2013)	Focused Deterrence Drug Market Initiative//drug, property, nuisance and violent crimes	No	One neighborhood (Rockford, IL) compared to remained of city	No matching criteria but show demographic differences for 6 census variables aggregated from block group level	No	Significant reduction in drug, property and nuisance offenses, but not for violent crimes
Corsaro, N. & Engel, R.S. (2015)	Focused Deterrence-violence reduction//homicide by type,	No	Citywide in New Orleans compared to 6 high trajectory homicide cities	Selected by examining similarities on homicide trends	Yes	Yes

Citation	Intervention// Outcome of Interest	Used PSM?	Target Area and Unit of Analysis and Utilization of Comparison Group	Matching Criteria	Did Study Discuss Threat of Competing Inter-ventions?	Stat. Significant Reductions in Outcomes in Treatment Area?
	and gun assaults					
Corsaro, N., and McGarrell, E. (2010)	Focused Deterrence Drug Market Initiative//drug and property crimes	No	One neighborhood (Nashville, TN)	The neighborhood (neighborhood is the unit) is compared to an adjoining neighborhood and the remainder of Davidson County, TN as a whole	No	Yes
Engel, R.S., Tillyer, M.S., Corsaro, N. (2013)	Focused Deterrence Violence Reduction//gang-member involved homicides and violent firearm incidents	No	Citywide in Cincinnati, OH; did not use a comparison area as counterfactual, instead used comparison *outcome* of non-gang member involved homicides.	N.A.	N.A.	Yes
Fox, A.M., Katz, C.M., Choate, D.E., & Hedberg, E.C. (2015)	Cure Violence//shootings, assaults and all violent crimes combined	No	One neighborhood (Hermoso Park) in Phoenix, AZ compared to 3 clusters of block groups	Used block groups as the unit; first constructed disadvantage index to score block groups, then took most disadvantaged and used census, police, and hospital data identify three sets of five contiguous block groups (the size of the target area) that best matched target area.	Yes-indicated no competing interventions in a footnote	Significant *increase* in shootings, but a decrease in assaults and all violent incidents combined.
McGarrell, E.F., Chermak, S., Wilson, J.M., & Corsaro, N. (2006)	Focused Deterrence//homicide	No	Matched the city of Indianapolis (Focused Deterrence) to comparison cities that ran across Highway	Comparison cities were chosen that were along Highway 64	No	Yes

Table 1. Studies of Community-Level Violence Reduction Interventions that Utilize a Quasi-Experimental Design with Counterfactual

Table 1. Studies of Community-Level Violence Reduction Interventions that Utilize a Quasi-Experimental Design with Counterfactual						
Citation	Intervention// Outcome of Interest	Used PSM?	Target Area and Unit of Analysis and Utilization of Comparison Group	Matching Criteria	Did Study Discuss Threat of Competing Inter-ventions?	Stat. Significant Reductions in Outcomes in Treatment Area?
			64: Cincinnati, Cleveland, Columbus, Kansas City MO, Louisville, and Pittsburgh	Used time-series analyses to determine how the homicide patterns changed to control for city differences		
Papachristos, A.V. Meares, T.L. & Fagan, J. (2007)	Focused Deterrence//gun violence measured as homicides and aggravated assaults	Yes	24 police beats comprised the target area in Chicago, IL (1 sq. mile area each) compared to the control area of 30 beats	PSM using rates of homicide & gun violence; Then eyeballed matches for balance on census variables.	Mentioned as a limitation (didn't directly address it)	Yes (homicides, gun homicides and aggravated assaults) No reduction in gang homicides
Picard-Fritsche, S. & Cerniglia, L. (2010)	Cure Violence model – Save Our Streets//shootings	No	1 Brooklyn police precinct compared to 3 similar precincts in Brooklyn and borough as a whole	Eyeballed similarities on a few demographic characteristics and gun violence in pre-intervention year.	Yes- conducted research to address possibility	Yes- reduction in shootings compared to increase in comparison areas
Skogan, W.G., Hartnett, S.M. Bump, N. & Dubois, J. (2008)	Cure Violence//gun crime: shots fired and gun homicides	No	Chicago, IL police beats with each target area having approximately 3 matched beats	Comparison areas chosen based on similar demographic features (census variables)	No	Mixed findings across different treatment areas and outcomes.
Spergel, I.A., Wa, K.M., & Sosa, R.V. (2005	Comprehensive Gang Model//violent crimes	No	Mesa, AZ school attendance area that received the intervention compared to three	Comparison school attendance areas chosen based on data from a city health and human	No	Yes (violent crimes)

Citation	Intervention// Outcome of Interest	Used PSM?	Target Area and Unit of Analysis and Utilization of Comparison Group	Matching Criteria	Did Study Discuss Threat of Competing Inter-ventions?	Stat. Significant Reductions in Outcomes in Treatment Area?
			other school attendance areas in Mesa	services report on discipline and demographics		
Spergel, I.A., Wa, K.M., & Sosa, R.V. (2005).	Comprehensive Gang Model//violence and gun crimes	No	Community within Riverside, CA receiving the program compared to other neighborhoods	Comparison area chosen based on amount of gang crime	No	Yes (less-serious violence)
Tita, G.E.; Riley, K.J., Ridgeway, G.; Grammich, C., Abrahamse, A., & Greenwood, P.W. (2003)	Hollenback Initiative (Focused Deterrence)//violent crime, gang crime, gun crime	Yes	Block groups across Los Angeles, CA; Three levels of matches: rest of the neighborhood that hadn't received program; five police districts that received program vs. remainder of the whole community; and then census block groups	Matched one-to-one on violence and gang-related crimes, per capita income, poverty, occupied housing units rented, population that moved within 5 years, population density, high school graduation rate, and percentage of population 15 to 24 years old	Discussion of choosing comparison areas touches on possibility of competing interventions	Yes
Webster, D.W., Whitehall, J.M., Vernick, J.S., & Parker, E.M. (2012)	Cure Violence//homicides and shootings	No	Baltimore, MD Police precincts that received the intervention compared to other police precincts	Compared police precincts that received program to other precincts that were in top quartile for the number of homicides and nonfatal shootings during the 3 years before program was implemented	No	Mixed findings across different neighborhoods
Wilson, J.M., Chermak, S., McGarrell, E.F. (2010)	Modeled after Cure Violence//homicides, aggravated assaults, gun assaults	Yes	3 program neighborhoods (not tied to police-defined boundaries or census aggregations) in Pittsburgh were compared to matched neighborhoods combined to form a simulated counterfactual neighborhood without	Matched on homicide, aggravated assault, and gun assault rates in pre-intervention year and a host of demographic characteristics from census data; also ran treatment effects models with neighborhoods that	Briefly mentioned that a focused deterrence strategy was not occurring at the same time	No reductions in outcomes examined

Table 1. Studies of Community-Level Violence Reduction Interventions that Utilize a Quasi-Experimental Design with Counterfactual

Table 1. Studies of Community-Level Violence Reduction Interventions that Utilize a Quasi-Experimental Design with Counterfactual						
Citation	Intervention// Outcome of Interest	Used PSM?	Target Area and Unit of Analysis and Utilization of Comparison Group	Matching Criteria	Did Study Discuss Threat of Competing Inter-ventions?	Stat. Significant Reductions in Outcomes in Treatment Area?
			the program	community experts suggested be used as comparisons		

Table 2. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Police Service Area (PSA) Level						
	Mean					
Unmatched/ Matched	Treated	Comparison	%Bias	% Bias reduction	t-value	p-value
2012 Shooting, Homicide, and Robbery Rate						
Unmatched	107.15	62.98	129.1		2.64	.01*
Matched	107.15	96.751	30.4	76.4	.50	.63
Probation & Parole Supervisees						
Unmatched	180	136.98	54.3		1.07	.29
Matched	180	209.2	-36.8	32.1	-.52	.62
Gang Count						
Unmatched	8.6	4.23	123.3		2.27	.03*
Matched	8.6	9	-11.3	90.8	-.14	.89
Pedestrian & Car Stop Scale						
Unmatched	.50	-.04	66.5		-1.24	.22
Matched	.50	.87	-45.8	31.2	-.49	.64
Percent Hispanic						
Unmatched	1.44	7.45	-78.7		-1.24	.22
Matched	1.44	10.82	-122.8	-56.0	-1.45	.18
Percent Black						
Unmatched	58.50	25.32	206.3		3.54	.00*
Matched	58.50	32.30	163.0	21.0	2.15	.06**
Concentrated Disadvantage						
Unmatched	1.43	-.12	210.6		3.91	.00*
Matched	1.43	.73	95.3	54.8	1.46	.18
Residential Stability						
Unmatched	.18	-.01	23.3		.43	.67
Matched	.18	.34	-19.1	18.0	-.44	.68
Total Population						
Unmatched	16137	23498	-88.2		-1.51	.14
Matched	16137	25562	-112.9	-28.0	-1.53	.16
N=5 PSAs treated matched to 5 untreated						

* $p < 0.05$; ** $p < 0.01$

Table 3. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Census Tract Level						
	Mean					
Unmatched/ Matched	Treated	Comparison	%Bias	% Bias reduction	t-value	p-value
2012 Shooting, Homicide, and Robbery Rate						
Unmatched	35.61	23.21	66.7		3.01	.00*
Matched	35.61	30.96	25.0	62.5	.91	.37
Probation & Parole Supervisees						
Unmatched	53.30	27.67	109.8		4.76	.00*
Matched	53.30	49.96	14.3	86.9	.05	.62
Gang Count						
Unmatched	4.43	2.12	70.8		3.10	.00*
Matched	4.43	4.04	12.0	83.1	.36	.72
Pedestrian & Car Stop Scale						
Unmatched	2.39	1.97	55.7		2.22	.03*
Matched	2.39	2.26	17.3	68.9	.63	.53
Percent Hispanic						
Unmatched	.03	.11	-65.5		-2.26	.02*
Matched	.03	.03	3.1	95.4	.48	.64
Percent Black						
Unmatched	.85	.41	159.0		5.82	.00*
Matched	.85	.81	15.6	90.2	.60	.55
Concentrated Disadvantage						
Unmatched	.98	-.10	145.1		6.23	.00*
Matched	.98	.81	22.5	84.5	.98	.33
Residential Stability						
Unmatched	-.32	.03	-43.7		-1.69	.09**
Matched	-.32	-.28	-3.7	91.6	-.18	.86
Total Population						
Unmatched	3306.5	4002.8	-43.3		-1.82	.07**
Matched	3306.5	3248.9	3.6	91.7	.15	.88
N=23 census tracts treated matched to 23 tracts untreated						

* $p < 0.05$; ** $p < 0.01$

Table 4. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Block Group Level

Unmatched/ Matched	Mean Treated	Comparison	%Bias	% Bias reduction	t-value	p-value
2012 Shooting, Homicide, and Robbery Rate						
Unmatched	12.72	6.59	67.4		6.74	.00*
Matched	12.72	11.64	11.9	82.3	.50	.62
Probation & Parole Supervisees						
Unmatched	16.01	8.05	77.9		7.50	.00*
Matched	16.01	15.40	6.0	92.3	.34	.73
Gang Count						
Unmatched	1.35	.61	60.4		5.87	.00*
Matched	1.35	1.33	1.1	98.1	.06	.96
Pedestrian & Car Stop Scale						
Unmatched	2.01	1.41	79.2		5.99	.00*
Matched	2.01	2.03	-1.8	97.7	-.10	.92
Percent Hispanic						
Unmatched	2.48	11.92	-66.9		-4.11	.00*
Matched	2.48	1.45	7.3	89.1	1.65	.10
Percent Black						
Unmatched	87.58	45.82	142.8		9.32	.00*
Matched	87.58	90.56	-10.2	92.9	-1.14	.25
Concentrated Disadvantage						
Unmatched	.78	-.05	119.5		9.53	.00*
Matched	.78	.73	6.4	94.6	.39	.70
Residential Stability						
Unmatched	-.24	.02	-30.1		-2.30	.00*
Matched	-.24	-.21	-3.1	89.9	-.21	.84
Total Population						
Unmatched	1011.50	1144	-24.5		-2.00	.05**
Matched	1011.50	1011.6	-0.0	99.9	.00	1.00

N=72 block groups treated matched to 72 block groups untreated

* p<.05 **p<.10

Figure 2a. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Census Tract Level

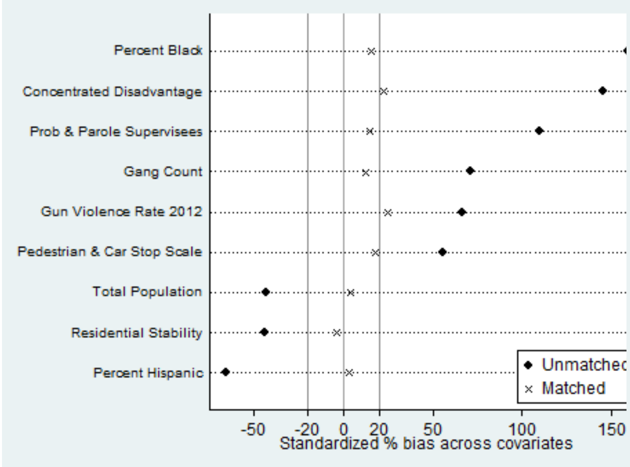

Figure 2b. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Block Group Level

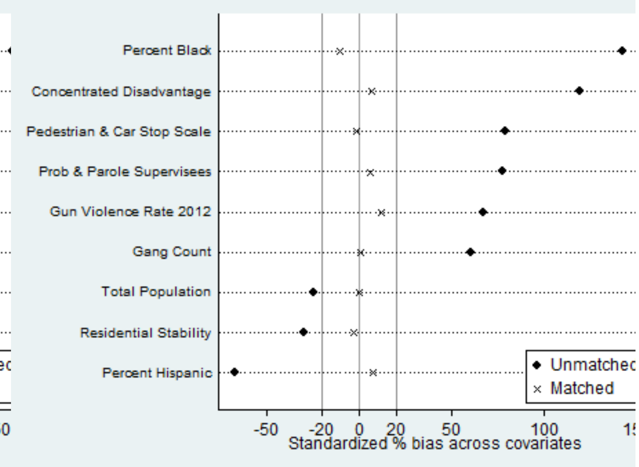

Figure 3. Post-Match Geographic Output for Propensity Score Matching of Communities at the Census Tract Level

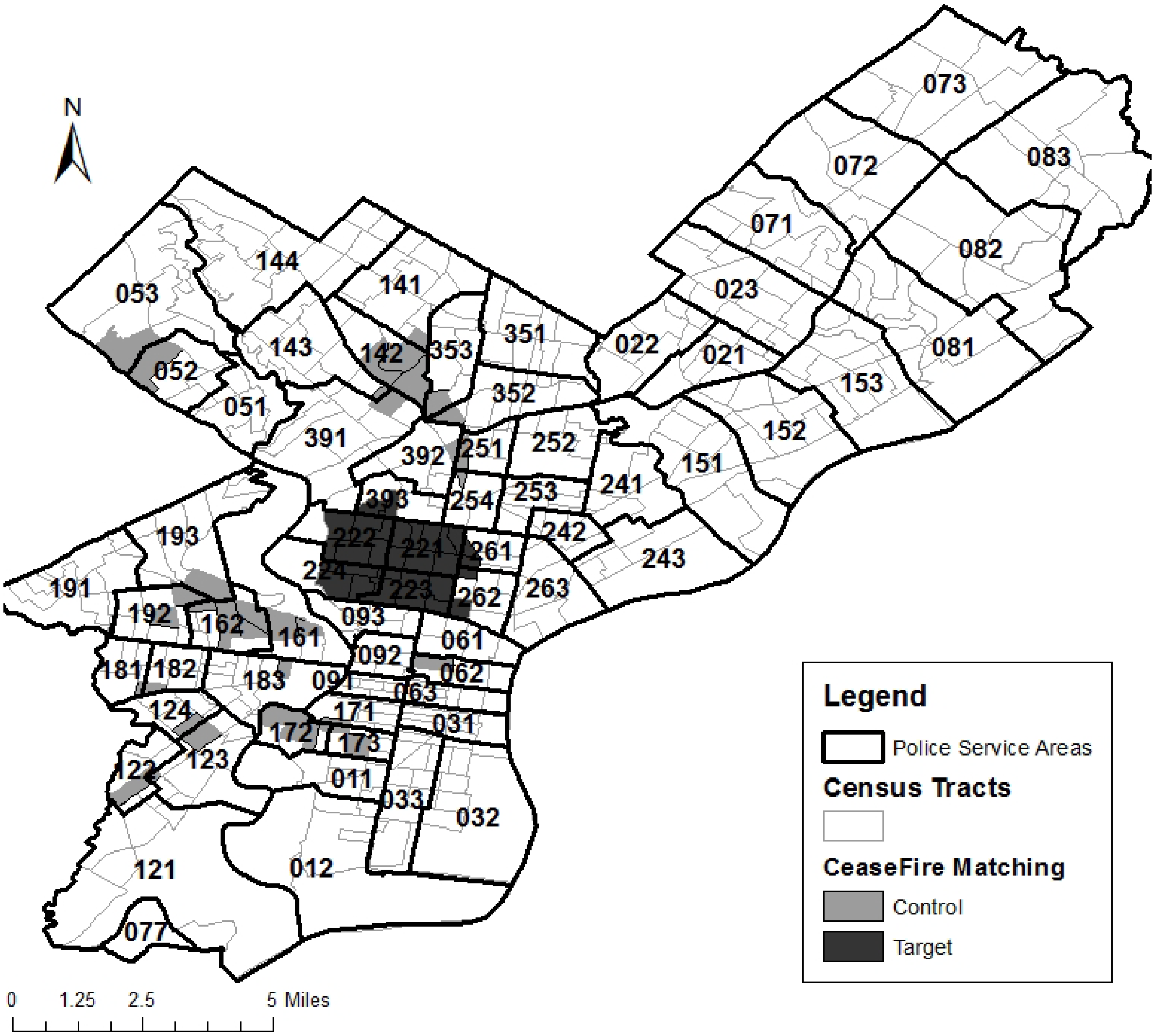

Figure 4. Post-Match Geographic Output for Propensity Score Matching of Communities at the Block Group Level

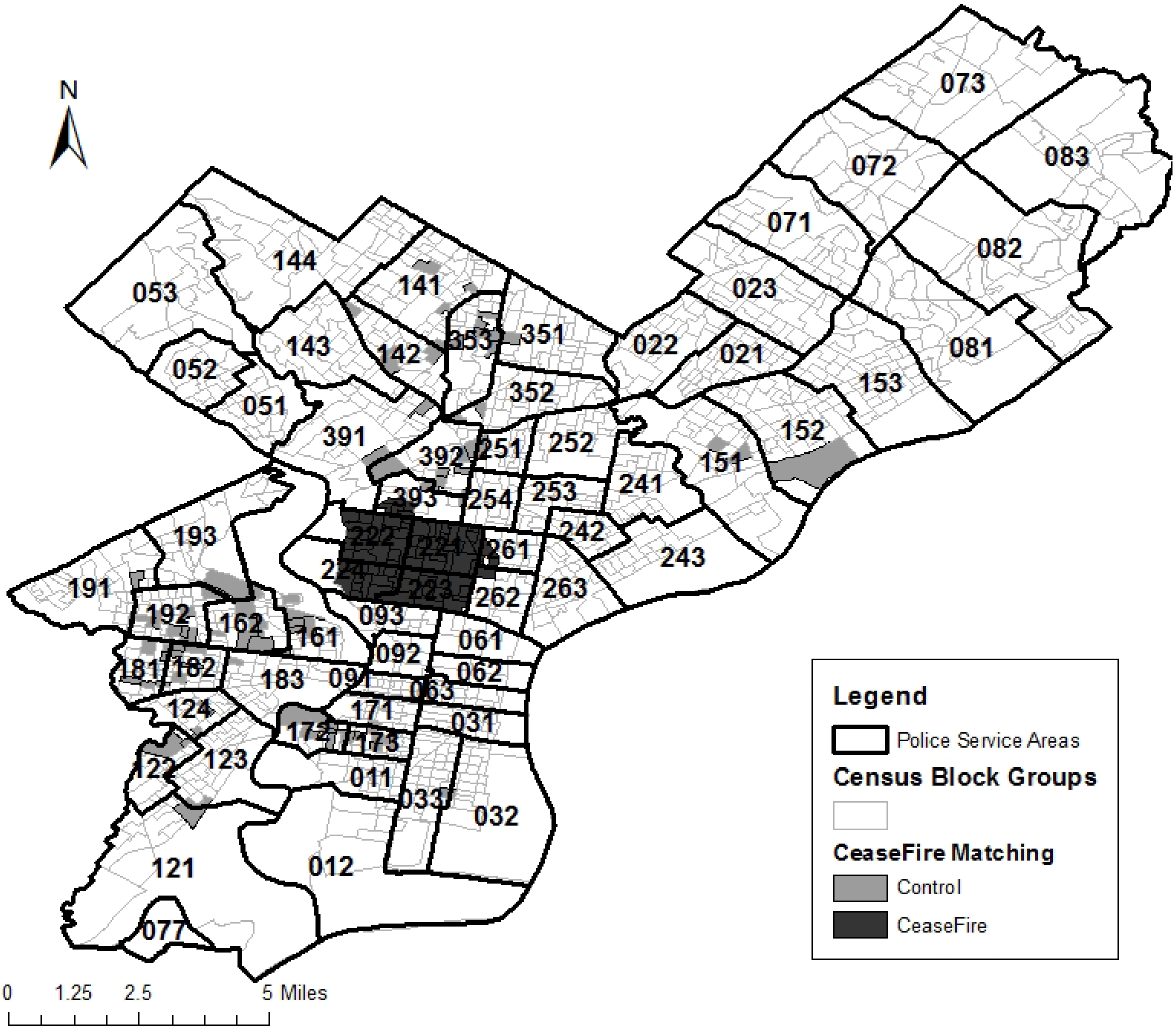

Table 5. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Block Group Level, Northern PSAs (PSA 221, 222 and southern portion of 393)

	Mean					
Unmatched/ Matched	Treated	Comparison	%Bias	% Bias reduction	t-value	p-value
2012 Shooting, Homicide, and Robbery Rate						
Unmatched	12.44	6.59	64.0		5.24	.00
Matched	12.44	14.30	-20.3	68.2	-.60	.55
Probation & Parole Supervisees						
Unmatched	16.59	8.05	84.0		6.55	.00
Matched	16.59	15.52	10.5	87.5	.45	.65
Gang Count						
Unmatched	1.17	.61	49.4		3.68	.00
Matched	1.17	1.41	-20.8	57.9	-.90	.37
Pedestrian & Car Stop Scale						
Unmatched	2.28	1.41	118.2		6.97	.00
Matched	2.28	2.43	-20.5	82.6	-.87	.38
Percent Hispanic						
Unmatched	2.59	11.92	-65.6		-3.25	.00
Matched	2.59	2.29	2.1	96.7	.25	.80
Percent Black						
Unmatched	90.07	45.82	151.9		7.91	.00
Matched	90.07	89.04	3.5	97.7	.32	.75
Concentrated Disadvantage						
Unmatched	.91	-.05	149.4		8.98	.00
Matched	.91	.86	8.5	94.3	.42	.67
Residential Stability						
Unmatched	-.13	.02	-17.6		-1.08	.28
Matched	-.13	-.16	3.6	79.6	.19	.85
Total Population						
Unmatched	1087.2	1144	-10.2		-.69	.49
Matched	1087.2	953.17	21.1	-136.0	1.11	.27

N=46 block groups treated matched to 46 block groups untreated

* p<.05 **p<.10

Table 6. Post-Match Bias Reduction Statistics for Propensity Score Matching of Communities at the Block Group Level, Hot Spot Areas

Unmatched/ Matched	Mean Treated	Comparison	%Bias	% Bias reduction	t-value	p-value
2012 Shooting, Homicide, and Robbery Rate						
Unmatched	11.37	6.68	54.5		3.24	.00
Matched	11.37	10.89	5.6	89.7	.21	.84
Probation & Parole Supervisees						
Unmatched	18.82	8.03	103.9		6.46	.00
Matched	18.82	18.04	7.6	92.7	.25	.80
Gang Count						
Unmatched	2.18	.55	122.7		8.66	.00
Matched	2.18	2.11	5.4	95.6	.18	.86
Pedestrian & Car Stop Scale						
Unmatched	2.11	1.35	99.5		4.80	.00
Matched	2.11	2.14	-4.7	95.3	-.16	.88
Percent Hispanic						
Unmatched	1.16	12.76	-80.0		-3.01	.00
Matched	1.16	.84	2.2	97.3	.55	.59
Percent Black						
Unmatched	95.67	47.85	175.9		6.68	.00
Matched	95.67	96.64	-3.6	98.0	-.63	.53
Concentrated Disadvantage						
Unmatched	1.01	-.02	141.5		7.46	.00
Matched	1.01	.78	31.5	77.7	1.34	.19
Residential Stability						
Unmatched	-.10	.02	-15.7		-.67	.50
Matched	-.10	-.08	-1.8	88.3	-.08	.94
Total Population						
Unmatched	1065	1149.4	-15.3		-.80	.42
Matched	1065	1015.9	8.9	41.7	.37	.71

N=28 block groups treated matched to 28 block groups untreated

Table 7. ARIMA Time Series Models Estimating the Treatment Effect of Philadelphia CeaseFire

Treatment/Comparison	Interval	ARIMA	Impact	SE	Z value	p-value
Larger Treatment Area						
CeaseFire (5 PSAs)	month	(1,0,0)	-.788*	.387	-2.03	.042
Comparisons	month	(2,0,0)	-.478**	.171	-2.77	.003
Northern Treatment Area						
CeaseFire (3 PSAs)	month	(1,0,0)	-.838*	.359	-2.33	.010
Comparisons	month	(0,0,0)	-.563*	0.282	-1.99	.023
Gun Crime Hot Spots						
CeaseFire	month	(1,0,0)	-1.080*	0.457	-2.36	.018
Comparisons	month	(2,0,0)	-.763	0.438	-1.74	.081

* $p < .05$, ** $p < .01$

[i] Philadelphia CeaseFire originally emerged as a program in 2011 as small pilot test to begin to build neighborhood support for full implementation of the Cure Violence Public Health Model. The pilot was comprised of three outreach workers and a supervisor, and their work was focused on one small hot spot within one Police Service Area in North Philadelphia.

[ii] PSAs were established in 2009, after the new Police Commissioner (Charles Ramsey) was hired. Each PSA is headed by a police lieutenant, and includes an average of three sergeants and thirty-nine officers who are responsible for patrolling the area. The idea is to increase police-community contact and officer involvement in the communities. The PSA model is considered a foundation of Philadelphia's neighborhood policing strategy (Joyce 2016).

[iii] Maps representing additional output with autocorrelated concentrated disadvantage included as a balancing variable are available from the authors upon request.