

Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot

Jessica Saunders¹  · Priscillia Hunt¹ ·
John S. Hollywood¹

Published online: 12 August 2016

© Springer Science+Business Media Dordrecht 2016

Abstract

Objectives In 2013, the Chicago Police Department conducted a pilot of a predictive policing program designed to reduce gun violence. The program included development of a Strategic Subjects List (SSL) of people estimated to be at highest risk of gun violence who were then referred to local police commanders for a preventive intervention. The purpose of this study is to identify the impact of the pilot on individual- and city-level gun violence, and to test possible drivers of results.

Methods The SSL consisted of 426 people estimated to be at highest risk of gun violence. We used ARIMA models to estimate impacts on city-level homicide trends, and propensity score matching to estimate the effects of being placed on the list on five measures related to gun violence. A mediation analysis and interviews with police leadership and COMPSTAT meeting observations help understand what is driving results.

Results Individuals on the SSL are not more or less likely to become a victim of a homicide or shooting than the comparison group, and this is further supported by city-level analysis. The treated group is more likely to be arrested for a shooting.

Conclusions It is not clear how the predictions should be used in the field. One potential reason why being placed on the list resulted in an increased chance of being arrested for a shooting is that some officers may have used the list as leads to closing shooting cases. The results provide for a discussion about the future of individual-based predictive policing programs.

Keywords Predictive policing · Program evaluation · Propensity score matching · Quasi-experimental design · Risk assessment · Time series analysis

✉ Jessica Saunders
jsaunders@rand.org

¹ RAND Corporation, 1776 Main Street, Santa Monica, CA 90407, USA

Introduction

The term “predictive policing” has garnered significant interest in the law enforcement community as a potential way to increase the likelihood of preventing crime before it occurs. While the term has various operational definitions, predictive policing is typically comprised of two elements: a prediction model that uses an algorithm to identify instances of increased crime risk, and an associated prevention strategy to mitigate and/or reduce those risks (Perry et al. 2013; Ridgeway 2013). With the progress of more advanced analytics (also commonly referred to as predictive analytics, machine learning, or data mining methods), statistical approaches are shown to make more accurate predictions than traditional crime analysis methods in the lab (Berk 2011; Berk and Bleich 2013; Cohen et al. 2007; Gorr and Harries 2003). By leveraging advanced analytics, police departments may be able to more effectively identify future crime targets for preemptive intervention. However, there is little experimental evidence from the field demonstrating whether implementing an advanced analytics predictive model, along with a prevention strategy—“predictive policing”—works to reduce crime, particularly compared to other policing practices in the field.

The impact of both prediction and prevention, that is, a predictive policing program, needs to be tested because there are plenty of reasons to believe improvements in the accuracy of predictions alone may not result in a reduction in crime. First, the predictions may not be actionable because the location and/or time are not precise enough. For example, predictions may identify census blocks at increased risk of crime in the following week, but for most police departments, this period is too long to efficiently and effectively implement a strategy to prevent a crime. Second, since the baseline accuracy of predictions is still relatively low, small improvements can be made to appear as large percentage improvements, when they are rather insufficient to make a difference in the real world. For instance, a method or model may improve the prediction of homicide perpetrators in a city in a year from 1 out of 100,000 people to 6 out of 100,000—a 500 % improvement—but using the average homicide levels in cities, the new approach will still fail to identify nearly 99.5 % of homicide perpetrators. Third, the crime prevention strategies may not work. So in our example, even if a prediction model identifies five more future homicide perpetrators than traditional analysis, the prevention program may not stop them from committing a homicide. Not to mention, they may not even receive the prevention program—the new method identified 6 out of 100,000, meaning 100,000 people would need to receive prevention to avoid 6 homicides. Finally, law enforcement may choose not to use the predictions (noncompliance). And these are only a few of the many reasons that an enhanced prediction model or method may not lead to crime reduction in the field, so there is still a need for research studies to better understand if and how predictive policing programs work in practice.

There is some experimental evidence of the impact of predictive policing strategies on crime, albeit limited. Two peer-reviewed field experiments of explicitly formulated predictive policing programs suggest results are mixed. One study, co-authored by predictive policing software developers, compared the use of predictive policing software to identify micro-places at high risk of crime to crime analysts manually labeling high-risk micro-places; police then conducted additional patrols in the identified places (Mohler et al. 2015). The study found that the predictive policing tool better

recognized future crime risk and reduced crime in comparison to the manual labeling by the crime analysts. But the other study, identifying blocks at high risk within an intelligence-led policing paradigm, did not result in a reduction of property crime compared to business-as-usual hot spots mapping and policing (Hunt et al. 2014). The authors concluded that the failure to identify an effect could be because the program did not in fact work (theory failure), low statistical power, and/or lack of program fidelity in some treatment units (implementation failure). The contradictory findings may be due to differences in study design, experimental control, prediction accuracy, prevention implementation, or a dozen other factors, which are discussed in more detail later.

In order to better understand the effects of an individual-focused predictive policing program in the field, this study analyzes a pilot program implemented in Chicago in 2013 aimed at reducing gun violence. The theory behind the program is not dissimilar to prior efforts to identify, monitor, and deter or incapacitate high-risk or highly active offenders to reduce crime (Abrahamse et al. 1991; Martin and Sherman 1986; Ridgeway et al. 2011; Sherman and Berk 1984). One difference, however, is that the individuals at high risk of being involved in crime in the future were identified using a predictive policing strategy based on a statistical model of co-arrest networks in a policing context. Importantly, the high-risk individuals were not necessarily under official criminal justice supervision nor were they identified through intelligence to be particularly criminally active.

The predictive policing strategy examined in this study refers to the pilot, or first phase, of the Chicago Police Department's (CPD) larger predictive policing program where individuals at highest risk for gun violence were placed on a Strategic Subjects List (SSL). The SSL was disseminated by central command and the prevention strategy was deferred to district commanders who decided the relevant policing intervention strategy for SSL individuals in their district. We test whether the introduction of the SSL affected city-level homicide rates. Furthermore, using individual-level data, we apply propensity score matching methods to estimate the impact of the SSL on the likelihood of high-risk individuals being involved in gun violence. Lastly, we test hypotheses for why the program may or may not have worked.

The rest of this paper is structured as follows: in section two, we provide a review of previous literature on the prediction and prevention of criminal behavior at the individual-level, focusing on linking predictions to practice within the intelligence-led policing paradigm. Section three provides a description of Chicago's SSL program and how it was implemented during the pilot period. In section four, we present the data and quasi-experimental methods used to evaluate program effectiveness and to identify the mechanisms driving outcomes. Section five presents results, and section six concludes with a discussion of what the findings mean for police departments looking to develop and implement similar programs.

Literature review

Predictive policing is a proactive policing model that was popularized, in part, by the development of advanced analytics that were lauded as highly successful in other fields (Beck and McCue 2009; Berry and Linoff 2004; Zikopoulos and Eaton 2011). Now

that analysts have increasing methodological and computational sophistication to predict future crime patterns (Berk and Bleich 2013; Cohen et al. 2007; Perry et al. 2013), many have claimed that police should be using these models to reorient their officers toward future, rather than current, problems (Beck and McCue 2009). The units of crime predictions can fall across a continuum of targets, ranging from large to small geographically sized areas (e.g., predict when and where crime is more likely to occur), all the way down to individuals (e.g., predict who is more likely to be involved in criminal activity). As the models differ in their predicted targets, we might expect these models to be tied to policing strategies for crime control purposes. The policing strategies would likely differ in terms of specificity of the prevention mechanism (e.g., from general prevention to targeted prevention of particular crime types or of specific behaviors) and of the time period (e.g., how far into the future the strategy might operate). However, the best ways to translate predictions into practice are still underdeveloped (Ridgeway 2013).

Predictive policing comes from a long history of proactive policing strategies focused on getting ahead of crime before it occurs (Bordua and Reiss 1966). Since the 1980s, a large body of robust literature has grown on the effectiveness of proactive policing techniques for reducing crime (Braga et al. 2012; Mazerolle et al. 1998, 2000; Sherman and Weisburd 1995; Weisburd and Mazerolle 2000). The evidence-based policing field is replete with studies that demonstrate that the police can proactively prevent and reduce crime, and each share a commonality in the broadest sense with one another and predictive policing—a basis for the selection of a particular target and a mechanism to prevent and reduce crime. Multiple examples of these studies can be found on the Evidenced-Based Policing Matrix (Lum et al. 2011), and one can conceive of integrating predictive analytics into most of these programs.

Predicting crime using individual-focused modeling

To date, much of the interest in predictive policing has focused on using geospatial modeling to predict future hot spots (Beck and McCue 2009; Groff and La Vigne 2002; Hunt et al. 2014; Mohler et al. 2015; Perry et al. 2013). In fact, researchers started assessing computational methods for what they called “predictive crime mapping” over 10 years ago (Gorr and Harries 2003; Groff and La Vigne 2002). Crime control programs that focus on geographic targets have been met with the most success (Braga 2005; Braga and Weisburd 2010; Sherman 1986; Sherman et al. 1997; Sherman and Weisburd 1995; Weisburd and Mazerolle 2000). Models and methods to predict the criminal behavior of individuals are less prominent in the criminal justice crime prevention literature.

While more recent advanced analytic approaches have focused on improving geographic predictions, predicting future dangerousness of individuals is nothing new. For decades, researchers have been developing clinical predictions of future dangerousness using subjective approaches, such as intelligence or expert opinion assessments of “high risk” (Abrahamse et al. 1991; Bar-Hillel 1980; Martin and Sherman 1986; Pate et al. 1976) and static crime analysis (Braga et al. 2012; Eck et al. 2005; Ridgeway 2013; Sherman and Weisburd 1995). Subjective-based predictions have been replaced by more modern, reliable, and valid actuarial methods (Grove and Meehl 1996; Litwack 2001). Generally, these methods apply mathematical models

to administrative data in order to conduct risk assessments and predict future dangerousness. Examples here include models predicting the risk of persons under community supervision reoffending (Berk 2008, 2011; Berk and Bleich 2013; Berk et al. 2009; Wright et al. 1984) and models assessing the risk that a gang affiliate will be involved in violence as a function of their social relationships (Papachristos 2009; Papachristos et al. 2011, 2012). Thus far, these models have been found to have a moderate level of predictive accuracy (Yang et al. 2010).

These actuarial risk assessments have not been used in a policing context until now. Police traditionally use intelligence and subjective assessments to identify and monitor high-risk individuals, whereas actuarial risk assessments are relatively standard practice in other parts of the criminal system, such as for correctional placement, court sentencing, and probation decisions (Dvoskin and Heilbrun 2001; Quinsey et al. 2000; Wright et al. 1984). These tools and methods are routinely integrated into decisions about supervision and sentencing, but researchers have warned that they were designed to facilitate efficient management of institutional resources, not to target individuals, and caution should be applied to ensure their proper use. The challenges of predicting future offending behavior and people's misunderstanding of predictions have been well documented (Bar-Hillel 1980). For the models assessing a person's future risk of offending or victimization, a key complication is that, while the models can identify increased risk, the overall risk can still be very low. Indeed, a "very high risk" person for homicide might have a risk rate of 1 % per year. According to some scholars, this may still lead to cost-effective policing strategies, as the cost of a "false positive", or someone who is incorrectly identified as a potential offender, is likely less costly than a "false negative," or someone who was incorrectly identified as a non-offender (Berk 2011). To further complicate the issue, researchers note that decisions using actuarial predictions may advance the continued marginalization of economically- and politically-disenfranchised populations (Silver and Miller 2002), which could be more detrimental in a policing context compared to someone in custody who has already been found guilty.

Prevention: intelligence-led policing paradigm

While the prevention strategies connected to any particular prediction may vary in specificity (e.g., general vs. targeted prevention) and the level of proactivity, the process of matching them is most closely described as an Intelligence-Led Policing (ILP) approach. A precise definition of Intelligence-Led Policing is difficult to nail down, but most broadly, it is a strategy that integrates data analysis and intelligence to help police prioritize their targets and activities (Ratcliffe 2002). Intelligence-led policing grew in popularity throughout the 1990s and early 2000s (Cope 2004), and according to Ratcliffe (2002, 2012), the defining characteristic is that intelligence is used as a "decision-making tool" to help the police prioritize their work effectively and efficiently to reduce crime, sometimes using external partnerships as a force multiplier. Intelligence-led policing can be situated at a variety of different organizational levels because "intelligence" can be used to inform programs, strategies, and even larger administrative priorities and policies. There are three components of an intelligence-led policing strategy, described in Ratcliffe's "3i" model (Ratcliffe 2005): *interpreting* the available information (in predictive policing, through the use of a predictive model);

using the information to *influence* an agency's decision-makers to adapt a crime-reduction strategy; and executing the strategy to *impact* criminal behavior and ideally, reduce crime (in predictive policing, through the use of a prevention program).

Evidence on the effectiveness of a well-implemented ILP paradigm in general appears to be lacking in one way or another. ILP is a framework or model and not a singular program, which makes evaluation very challenging because it can be implemented very differently across different settings (treatment heterogeneity). Research tends to focus on organizational issues in adopting a true ILP model at different levels (Ratcliffe 2002, 2005; Ratcliffe and Guidetti 2008). Ratcliffe (2012) reviewed the effectiveness of a few different programs in ILP frameworks with different methods, targets, and settings, and found that the frameworks met with some success. However, it may be that ILP alone is not enough to create success. It is likely the framework must be paired with evidenced-based programs to create positive results (Ratcliffe 2002). This is not unlike other disciplines, such as health (Chinman et al. 2004) and education (Sailor et al. 2008) in which evidence-based frameworks or models are built to include evidence-based programs that can vary over time and across locations.

Connecting individual-focused predictions to practice

ILP is a framework for applying interventions in which specific interventions must be developed separately. There are a number of interventions that can be directed at individual-focused predictions of gun crime because intervening with high-risk individuals is not a new concept. There is research evidence that targeting individuals who are the most criminally active can result in significant reductions in crime (Braga and Weisburd 2012; Gendreau et al. 1996; Lipsey 1999; Loeber and Farrington 1998; Martin and Sherman 1986; Sherman et al. 1997). Additionally, successful programs that work with these high-risk offenders have also demonstrated cost-effectiveness (Caldwell et al. 2006; Foster and Jones 2006). For example, Martin and Sherman (1986) found that Washington, DC's selective apprehension program did arrest repeat offenders more frequently than they would have been arrested otherwise, but the study did not examine whether the increased arrests reduced crime.

Conversely, some research shows that interventions targeting individuals can sometimes backfire (McCord 2003; Sherman 1992; Welsh and Rocque 2014). As an example, some previous proactive interventions, including increased arrest of individuals perceived to be at high risk (selective apprehension) and longer incarceration periods (selective incapacitation), have led to negative social and economic unintended consequences. Auerhahn (1999) found that a selective incapacitation model (Greenwood and Abrahamse 1982) generated a large number of persons falsely predicted to be high-risk offenders, although it did reasonably well at identifying those who were low risk. At the extreme, "Three Strikes and You're Out" laws were intended to incapacitate indefinitely those labeled as lifetime high-rate offenders ("predators"). Gottfredson and Hirschi (1986) found that such lifetime offender groups do not exist (i.e., propensity to engage in crime drops dramatically with age regardless of risk factors present earlier), and thus their incapacitation was unlikely to have prevented crime. Indeed, one study found three strikes was positively associated with homicides and had no statistically significant impact on crime rates (Kovandzic et al. 2004).

Models engaging small groups of criminally active offenders with policing actions have seen more recent interest. A meta-analysis of focused deterrence (also known as “pulling levers” models, e.g., Kennedy 1996 and McGarrell et al. 2006) found such interventions promising at reducing community crime and violence (Braga and Weisburd 2012). However, those selected for focused deterrence interventions were identified through manual police and community efforts, not through predictive analytics. And, notably, the studies did not follow the individuals who were targeted, but instead examined their impact on community crime rates, with the exception of one evaluation that examined reduction of violence within targeted gang sets, but not necessarily program participants (Papachristos and Kirk 2015).

Prevention strategies could affect the likelihood of criminal activity through four potential mechanisms: (1) treatment, which would change the internal motivation to offend (for high-risk people); (2) specific deterrence, which would change the external motivation to offend (for high-risk people); (3) incapacitation, which would limit the ability to offend (for high-risk people); and (4) general deterrence through perceiving more credible deterrence messaging and changing the social environment (for all people); see Fig. 1 for an illustration.

Three of the four crime control mechanisms focus directly on the high-risk individual, which according to research, is likely to be a much more effective way to target scarce resources because a small minority of offenders commit the majority of crimes (Blumstein 1986). Treatment works through changing the internal motivations for committing a crime and providing high-risk individuals with the skills they need to succeed (Gendreau et al. 1996; Sherman et al. 1997). Deterrence works through three factors associated with punishment: certainty, swiftness, and severity (Becker 1993; Cornish and Clarke 2014). People are assumed to act as if they are trading off the gains from crime with the costs of punishment and to weigh the present more heavily than the future (discounting). So when the expected benefit of committing a crime is greater than the expected costs, people commit crime. The greater the probability the crime prompts swift, certain, and severe punishment, for example, arrest or citation, the less likely someone will commit that crime, thus *detering* the would-be criminal. The program may also reduce crime through an *incapacitation* effect whereby police may immediately increase incapacitation focusing on serving warrants or on violations of probation or parole for predicted individuals, which would remove their opportunity to commit crimes. Removing or reducing some of the criminally active population can even generate an overall “cooling” effect on others who were not targeted, thus leading

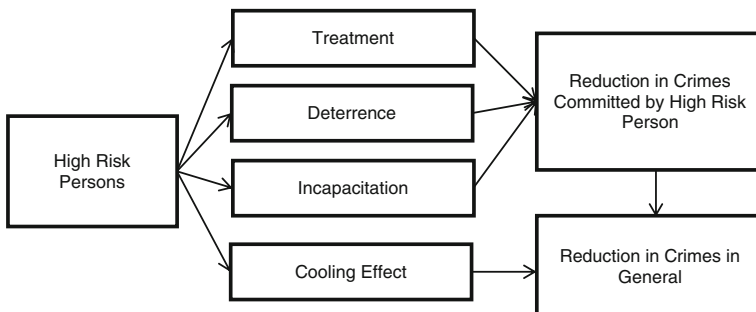


Fig. 1 Crime control mechanisms for individual-focused predictive policing program

to further reduction in crime by those indirectly impacted by the crime control strategy (Braga and Weisburd 2012; Kennedy 1996).

The Chicago SSL predictive policing program

Overview

This study investigates the predictive policing pilot program developed in collaboration between the Chicago Police Department (CPD) and the Illinois Institute of Technology (IIT), funded by the National Institute of Justice.¹ The aim of the pilot was to more efficiently and effectively target limited resources towards subjects in a community at risk for participation in gun violence, either as victims or perpetrators. IIT was responsible for the prediction model and used data contained within the CPD data warehouse to identify prior arrestees at a heightened risk for homicide (e.g., developed automated intelligence). The CPD led in the prevention strategy, using the computer-generated intelligence to action the intelligence.

All five stages of ILP—(1) acquiring information, (2) analyzing intelligence/ information, (3) reviewing and prioritizing, (4) acting on intelligence and tasking responsible parties with the plan, and (5) evaluating the impact of that action—were included in the strategy. CPD worked on step 1 by upgrading their IT infrastructure to allow them to automate the data collection from their data warehouse (Cope 2004). IIT worked on step 2, using the data to estimate an empirical model to generate data-driven intelligence. CPD worked on steps 3 and 4, with the Department of Operations first reviewing the list produced by IIT, and then tasking the district commanders to action the intelligence. While IIT continues to make model improvements and is currently on their fourth version (Lewin and Wernick 2015), this study evaluates implementation of model version 1.0.

CPD prediction model (interpreting)

The prediction model uses social networks (in the form of co-arrests) to previous homicide victims to predict the likelihood of someone becoming a victim of a homicide, and therefore essentially automating the intelligence production in the ILP strategy. The model focuses heavily on the recent body of literature examining correlations between victimization and the social connections to others who were victims of homicide (Papachristos 2009; Papachristos et al. 2011, 2012). The first model specification, version 1.0, estimates the relative risk of a subject being a homicide victim based on two input variables—the number of first-degree co-arrest links and the number of second-degree co-arrest links with previous homicide victims. A “first-degree link” refers to a relationship between a subject and an individual with whom the subject was previously co-arrested who later became a homicide victim. A “second-degree link” refers to a relationship in which a subject was co-arrested with another person who, in turn, was co-arrested with a later homicide victim (see Fig. 2).

¹ This is only one component of the larger long-term collaboration which broadly explores whether and how crime can be predicted.

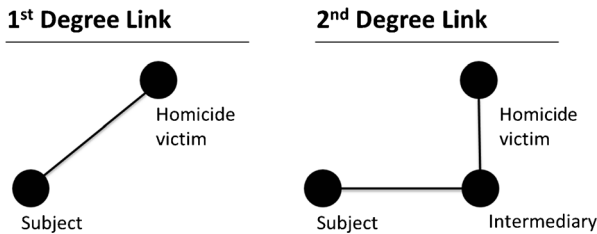


Fig. 2 First- and second-degree co-arrest links

The counts of co-arrest links were summed over the past 5 years and weighted for recency. A quadratic model was fitted to the data, meaning the probability of being a homicide victim increased at an increasing rate with respect to the count of links. Early model tests found that for some subjects (those with many links), the odds of being a homicide victim were thousands of times the risk of the general public over a 2-year follow-up window. To address the numerical instability² reflected by these results, the maximum assigned risk multiplier was “500+”. This study does not evaluate the validity or reliability of the model, but rather focuses on the impact of the predictions being used in practice.

CPD's prevention strategy (influencing and impacting)

The list of subjects and their risk scores were then forwarded to the CPD Deployment Operations Center (DOC). The DOC did a subjective review of the list, assigning subjects from different police districts to the SSL based on their scores and human intelligence collected by CPD. Each district was to be assigned a list of the 20 people in their district with the top risk scores in addition to all subjects with risk scores of “500+”. A total of 426 of some of the highest-risk individuals³ were put onto the SSL version 1.0 on March 26, 2013.

The individuals on the SSL were considered to be “persons of interest” to the CPD. District commanders were responsible for directing their personnel to act on this intelligence and were to be held accountable at regular COMPSTAT meetings. Commanders were not given specific guidance on what treatments to apply to their SSL members; instead, they were expected to tailor interventions appropriately. The main guidance provided by central leadership was to use the programs detailed in Chicago's *Gang Reduction Strategies* (Chicago Police Department 2014) on SSL members when possible,⁴ but commanders were left wide discretion as to what actions their units should take.

² Here, “numerical instability” refers to the likelihood that the estimates that an SSL member's calculated risk of a specified thousands of times more likely to be killed is due more to statistical artifacts from fitting the quadratic curve than an accurate estimate.

³ Initially, CPD said they would put all the highest-risk individuals on the list; however, they decided to vet the list through their Deployment Operations Center (DOC), who made some changes to who would appear on the list, and therefore, the 426 individuals did not represent the highest scoring individuals based on the model.

⁴ Details of GVRS are available at <http://directives.chicagopolice.org/directives/data/a7a57bf0-136d1d31-16513-6d1d-382b311ddf65fd3a.html>

Since the prevention strategy was decentralized to district commanders, we collected qualitative evidence to identify the prevention and intervention strategies used in the field. Specifically, from October 2013 to June 2014, the research team conducted interviews with district commanders (or their executive officers) and observed COMPSTAT meetings. A total of 43 semi-structured interviews were conducted across all 22 police districts. Commanders or executive officers from each district were interviewed at least once and a maximum of three times over the course of 9 months. Interviews lasted between 10 and 30 minutes and consisted of questions regarding the quality of the SSL, uses of the SSL in practice, and suggestions to improve the SSL. The research team also observed 48 COMPSTAT presentations by both area and individual district commanders across 17 COMPSTAT meetings. Structured data collection protocols were completed for both the interviews and the COMPSTAT observations, which were then entered into a spreadsheet. The detailed notes were then coded and content analyzed.

The COMPSTAT observations demonstrate most districts did not focus on intervening with SSL subjects. In over two-thirds (68.8 %) of presentations observed, there was no mention of the SSL. In another 12.5 % presentations, there was an acknowledgment of the SSL, with no further discussion. In less than one in five (18.7 %) presentations, there was both a discussion and executive guidance, which consisted of: (1) allow beat officers to take the lead in contacting SSL subjects, (2) consider using fugitive location and district intelligence teams to locate SSL subjects, and/or (3) change the focus from arresting SSL subjects for minor offenses (for which they would be immediately released) to finding ways to detain SSL subjects over the long term. There was no evidence of executive follow-up on these recommendations at the meetings.

Interview findings indicate two common themes for how commanders recommended addressing SSL subjects to beat officers during the pilot. First, officers (or teams of officers) were assigned to make contact with SSL subjects on varying schedules, usually by going to their home addresses or other locations (7 of 22 districts, or 31.6 %). Second, officers were provided information about the identities of the SSL subjects, and the officers were to make contact with the SSL subjects “if noticed”, especially if the subjects were acting suspiciously (10 districts, 45.4 %). The remaining cluster of districts (5 districts, 22.6 %) reported some combination of both approaches. Interviews indicated that directing officers to increase contact with SSLs was likely the extent of the preventive intervention strategy for the majority of districts. In interviews with CPD staff, it was noted that the central command encouraged the local commanders to take advantage of enhanced legal sanctions authorized through the *Gang Violence Reduction Strategy* (Chicago Police Department 2014), but officers reported using these programs for a very small subset of SSL subjects.

In sum, findings from the observations of COMPSTAT meetings and interviews with district-level administration suggest the topic of SSL subjects received relatively little attention. Overall, the observations and interview respondents indicate there was no practical direction about what to do with individuals on the SSL, little executive or administrative attention paid to the pilot, and little to no follow-up with district commanders. These findings led the research team to question whether this should be considered a prevention strategy.

Methods

This study uses an intent-to-treat analysis to estimate the impact of the CPD SSL predictive policing program version 1.0 on city-wide homicide levels and on incidence rates of individual-level involvement with gun crime. Outcomes of the pilot are analyzed using two quasi-experimental methods. First, we perform an interrupted time-series analysis of city-level data to determine whether the introduction of the SSL changed the aggregate homicide crime trend. Second, we conduct propensity score analysis using individual-level data in order to determine whether being on the SSL affected the likelihood of being involved with gun violence. To do so, we exploit the fact that some of the subjects on the SSL identified as being at highest risk were not treated (i.e., their score met the criteria for inclusion on the SSL but they were excluded for reasons we describe later). Last, we test hypotheses regarding the mechanisms driving the effects.

Data

For the city-level analysis, this study used publically available data on CPD's website (gis.chicagopolice.org). The outcome was monthly homicides from January 2004 through April 2014. Examining a 10.75-year period, there was an average of 38 monthly homicides ($SD = 10.65$). Homicide counts followed a distinct seasonal pattern and a general linear trend downward over the entire period.

For the individual-level analysis, we used two datasets provided by CPD that cannot be made public due to the sensitive nature of the data (for aggregated descriptive statistics, see Table 1). The first dataset was a person-level file of 873,281 individuals with arrest histories prior to March 2013, including data on (1) demographics (gender, age at most proximate arrest, race), (2) arrest history (number and type), (3) social network variables (number of first- and second-degree co-arrestees who were victims of homicide), and (4) the risk score generated by IIT. This file also contained four outcomes of interest for the 1-year follow-up period: (1) murder victim, (2) shooting victim, (3) arrest for murder, and (4) arrest for a shooting. The second dataset contained all recorded police contact with the 17,754 arrestees with at least one first- or second-degree association with a homicide victim and law enforcement from 1980 through the end of the observation window. This file contained 542,636 records, with 464,006 dating before the intervention period.

Analysis

City-level outcomes: interrupted time-series design (ARIMA)

This paper uses a time-series analysis to better understand whether the introduction of the SSL version 1.0 affected the homicide trend in Chicago. Specifically, we tested whether it led to a reduction of gun violence at an aggregate level. The ARIMA models were based on two components: (1) a model of homicides as a time series, which makes inferences about the underlying process in the dependent outcome using different time-series variational components, and (2) an impact assessment of the introduction of the SSL version 1.0 (McCleary et al. 1980). The first step was to identify the model for the time series, which characterizes autoregressive, non-seasonal differences, and lagged forecast

Table 1 Sample descriptive statistics and balance table

Variables	SSLs <i>n</i> = 426		Unweighted comp group <i>n</i> = 17,754		Weighted comparison group ESS = 273		Std. effect size	K Smirnov statistic	<i>p</i>
	Mean or %	SD	Mean or %	SD	Mean or %	SD			
Demographics									
Male	95.8 %		90.4 %		96.9 %		-0.06	.01	.29
African-American	77 %		77.5 %		76.9 %		0.00	0.00	1
Age at last arrest	22.57	4.76	26.23	9.02	22.91	5.05	-0.70	.05	.74
Prior arrests									
Murder	.05	.25	.03	.19	.06	.24	-0.15	.01	1
Sexual assault	.02	.15	.02	.16	.02	.13	0.02	0	1
Robbery	.49	.91	.37	.77	.53	.88	-0.04	.03	1
Aggravated battery	.25	.57	.17	.45	.32	.65	-.12	.04	1
Burglary	.35	.81	.31	.81	.48	.97	-.17	.06	.51
Theft	.73	2.08	.77	2.81	.87	2.19	-.07	.02	1
MVT	.77	1.24	.61	1.16	.75	1.18	.02	.02	1
Arrest	3.76	4.62	3.60	4.51	3.98	4.82	-.04	.03	1
Outstanding warrant	1.21	1.73	.88	1.47	1.18	1.75	.02	.04	.88
Prior contact with CPD									
All arrests	18.87	12.19	11.28	9.51	18.13	12.74	.07	.06	1.00
Contact cards	40.83	84.12	12.18	20.18	34.46	58.55	.08	.07	.40
Victimizations	.85	1.18	0.82	2.13	.95	1.56	-.08	.03	.31
Risk									
# 1st Degree	1.16	1.08	.24	.46	1.04	.92	.12	.07	.42
# 2nd Degree	7.28	5.70	1.46	1.40	6.25	5.31	.18	.12	.01
Risk score	268.58	195.15	30.48	34.61	208.62	177.14	.31	.16	<.01

errors, and the second step was to model in intervention effects (Bruinsma and Weisburd 2014). Once the appropriate model was selected, we also conducted a series of sensitivity analyses to test if there were other breaks in the time series not related to SSL.⁵

This study uses the ARIMAX R procedure (Ohri 2013) to test the impact of the SSL on overall city-level homicides in Chicago. ARIMAX identified the best fitting model of the entire time trend as ARIMA (0,0,0)(2,0,1), meaning the homicide series does not display non-seasonal variation, yet it is characterized by the following seasonal

⁵ It is important to note that Chicago has gone through transformative changes over this time period, including a new Superintendent in 2011 and the integration of COMPSTAT to provide oversight. In addition to changes in leadership and management style, CPD has implemented a large number of homicide reduction strategies during this time period, including the multiple changes to the Gang Violence Reduction Strategy, and gang call-ins across different districts starting in 2010.

processes: two seasonal autoregressive terms, no differences across seasons, and one seasonal moving average term.

Individual level outcomes: propensity score matching design

Chicago PD implemented the program citywide and was not willing to randomize high-risk subjects to the SSL, so an experimental setting for evaluating program effectiveness on the individual-level was not achieved. CPD originally stated that they would treat the top 20 individuals in each district, and everyone who scored above a certain threshold (500+), which would lend itself to statistical approaches for quasi-experimental settings such as a regression discontinuity design or instrumental variable approach. However, there was, in fact, quite a substantial overlap in risk scores between the SSL and a group of individuals that did not end up on the list, who became our pool of non-treated potential comparison cases (see Fig. 3). This overlap of treated (i.e., on the SSL) and untreated (i.e., not on the SSL) happened for two reasons: (1) some districts did not have a large number of the highest-risk people in their area of operation, so individuals with slightly lower scores appeared on their list; and (2) the DOC had some discretion on who to put on the SSL, particularly when there were a lot of very high-risk individuals, so not all of the highest scoring individuals were ultimately placed on the SSL. This allowed us to match risk scores, along with other observed features that research indicates are associated with predicting future criminal offending, in order to generate a weighted comparison group that was almost statistically indistinguishable from the SSL on all observable measures.⁶

Individuals were selected into treatment by their risk scores and, therefore, they differed systematically from eligible comparison cases. In order to control for these differences, case weights, w_i , were estimated to alter the covariate distribution (including the risk scores) for the comparison group. This results in equivalent comparison and treatment groups (to create treatment effects on the treated estimates, as opposed to average treatment effects). These weights were created using boosted regression to provide the conditional odds of receiving treatment where x_i is the vector of control variables and $p(x_i)$ is the estimated conditional probability of receiving treatment for an individual with control variables equal to x_i , also known as the propensity score (McCaffrey et al. 2004; Rosenbaum and Rubin 1983):

$$w_i = \frac{p(x_i)}{1-p(x_i)}.$$

Using the weighting scores generated in the TWANG R package (Ridgeway et al. 2014), the following equation reduces the bias of estimates of the effective average treatment effect on the treated (EATT):

$$EATT = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i} - \frac{\sum_{i=1}^n y_i w_i (1-t_i)}{\sum_{i=1}^n w_i (1-t_i)},$$

⁶ While there is always the possibility that the groups are different on unobservable variables, we have captured many of the important research-validated criminogenic factors. That is why we specify the approach reduces, rather than eliminates, bias.

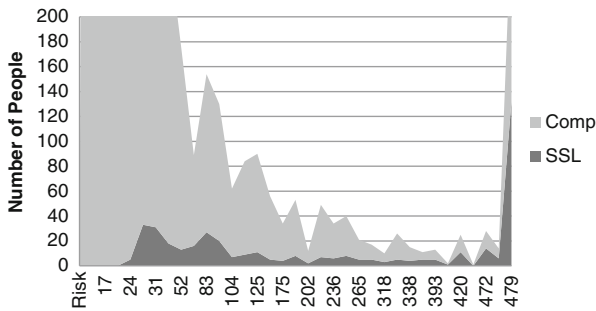


Fig. 3 Overlap of risk scores between SSLs and comparison group

where y is the outcome variable and t is the treatment indicator, for all individuals i . After applying the propensity score weights, the comparison group was reduced from 17,754 unweighted cases to an effective sample size of 273, and the only significantly different pre-treatment covariates were the number of second-degree associates and the risk scores. Since these two variables were still unbalanced (e.g., the treatment group still had higher scores than the weighted comparison), we used these variables as covariates in all weighted regressions and estimated doubly-robust models (Funk et al. 2011; Huber 1973). Doubly-robust models add remaining unbalanced variables as predictors to control for any residual between-group differences to create the closest statistical match between treatment and control groups as possible. According to scholars, this is a valid way of controlling for unbalanced and missing variables for causal modeling with propensity score matching (Kang and Schafer 2007).

Estimates of the impact of the predictive policing strategy could be biased if there were other programs targeting our control and/or treatment groups. While there were many violence reduction initiatives taking place in Chicago, including call-ins, almost none of the SSL subjects were included. An exception was one district (out of 22), which was a pilot area of the program that issued focused deterrence notification letters to SSL subjects. The commander (or her designee), along with a representative from a non-profit that coordinates services to ex-offenders, visited the SSL subjects or their families to let the subject know that he or she was identified as being at heightened risk for homicide victimization. This was the only place where a formal “treatment” was offered as a way to prevent gun crime, but we had no way to track if any SSL subject received services. As only one district participated, less than 5 % the SSLs were subject to this intervention, so we would not expect this to be driving our results.

Mediation analysis

In an effort to better understand what may be driving an association between inclusion on the SSL and our individual-level outcomes (involvement in gun violence), we conducted a mediation analysis. The mediation analysis allows us to explain the mechanisms that underlie an observed relationship between the SSL and involvement in the commission of a gun crime via the contacts with police. That is, we hypothesize the program was designed to enhance the *deterrence message* (through the probability of getting caught) not only directly but also by delivering prevention strategies that could affect decision-making that would lead to a weapons offense (as either an

offender or a victim). It may also have an *incapacitation effect* by removing the SSLs from the community at a higher rate. In this analysis, we try to tease out whether the individual-level effects of being on the SSL are based on deterrence or incapacitation.

Since the prevention strategy most consistently described for every district of Chicago was contact with SSL members, we considered contact with police as a mediator variable by first testing whether those on the SSL experienced greater contact with police than the comparison group. We then conducted a mediation analysis in which the likelihood of involvement with a gun crime as either a victim or offender depends on being on the SSL (direct effect) and on the extent of contact with police, which is a function of being on the SSL (mediation effect).

Results

Impact

City-level homicide rates

Examining the raw data, it is clear that homicide trends display a high degree of seasonality, with more homicides occurring in the warm months.⁷ There is a negative linear trend across the series ($b = -0.0267$), shown as the dotted line in Fig. 4, with overall monthly homicides decreasing from January 2004 through September 2014. Visually, it appears the homicide trend was falling prior to the introduction of the SSL in April 2013; the SSL version 1.0 was released to the district commanders at the end of March. The ARIMA analysis was conducted to statistically test whether the introduction of the SSL affected the monthly homicide levels in Chicago.

When we entered the month that the SSL was introduced to the ARIMA (0,0,0)(2,0,1) model in April, we found that there was a decrease in monthly homicides of 3.90 (95 % CI: 0.22, 7.58). However, sensitivity analyses examining decreases in homicide that pre- and post-date the SSL program show that this decrease is likely to be part of an overall trend downward, and not specifically due to the SSL intervention. Specifically, we conducted a set of “placebo tests”, where we modeled in a fake intervention date and tested whether there was still an effect. The analysis indicated a statistically significant reduction in monthly homicides when using 7 months pre- and post-SSL introduction. This demonstrates there was a statistically significant reduction in monthly homicides each month above and beyond the longer-term linear trend prior to the intervention (Table 2). With this sensitivity analysis in mind, we conclude that the statistically significant reduction in monthly homicides predated the introduction of the SSL, and that the SSL did not cause further reduction in the average number of monthly homicides above and beyond the pre-existing trend.

Impact on individual risk of homicide

Between March 2013 and March 2014, there were 405 homicides in Chicago. Seventy-nine percent of the homicide victims had a criminal history in the years prior to their deaths, 16 % had at least one association with a homicide victim, and 1 % of them were

⁷ With the exception of the winter of 2012 which did not experience the same degree of a “cooling off” period.

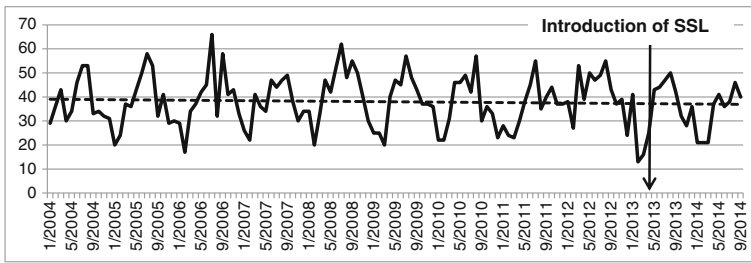


Fig. 4 Monthly homicides in Chicago from January 2004 to September 2014

listed on the SSL (see Fig. 5). Looking at this from another perspective, 0.7 % of the 426 SSL subjects were homicide victims, 0.4 % of the 17,754 associates were homicide victims, 0.029 % of the 855,527 former arrestees with no associates were homicide victims, and 0.003 % of the rest of the almost 2 million Chicago residents without any criminal record were victims of homicide.

A greater proportion of individuals on the SSL were involved in a shooting as either a victim or an arrestee, 6.8 % ($n = 29$), than comparable former arrestees who were not placed on the list but had at least one first- or second-degree association with a homicide victim (3.2 %, $n = 529$), and 0.2 % of former arrestees with no linkages to prior homicide victims ($n = 1,939$). However, once other demographics, criminal history variables, and social network risk have been controlled for using propensity score weighting and doubly-robust regression modeling, being on the SSL did not significantly reduce the likelihood of being a murder or shooting victim, or being arrested for murder. Results indicate those placed on the SSL were 2.88 times more likely to be arrested for a shooting (Table 3).

Mediation analysis

Seventy-seven percent of the SSL subjects had at least one contact card over the year following the intervention, with a mean of 8.6 contact cards, and 60 % were arrested at some point, with a mean of 1.53 arrests. In fact, almost 90 % had some sort of interaction

Table 2 Placebo sensitivity analyses: all homicides

Intervention month	Intervention coefficient	SE
Oct 2012	-4.23*	2.15
Nov 2012	-4.13*	2.18
Dec 2012	-4.52**	1.39
Jan 2013	-4.46**	1.68
Feb 2013	-6.33**	1.89
March 2013	-5.54**	1.93
April 2013	-3.90**	1.88
May 2013	-3.00	2.36

* $p < 0.05$, ** $p < 0.01$

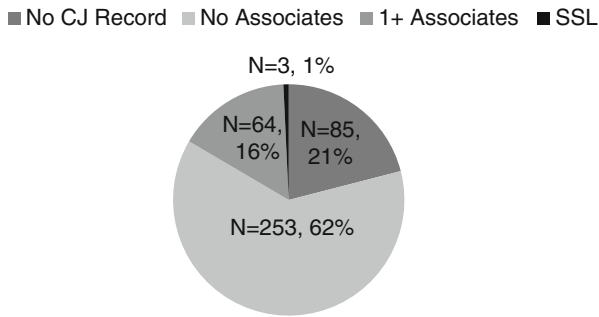


Fig. 5 Homicides in Chicago from March 2013 to March 2014 by risk group

with the Chicago PD (mean = 10.72 interactions) during the year-long observation window. This increased surveillance does appear to be caused by being placed on the SSL. Individuals on SSL were 50 % more likely to have at least one contact card and 39 % more likely to have any interaction (including arrests, contact cards, victimizations, court appearances, etc.) with the Chicago PD than their matched comparisons in the year following the intervention. There was no statistically significant difference in their probability of being arrested or incapacitated⁸ (see Table 4). One possibility for this result, however, is that, given the emphasis by commanders to make contact with this group, these differences are due to increased reporting of contact cards for SSL subjects.

Results of the impact analysis show the only statistically significant difference between the comparison and SSL group was an increase in arrest for shooting. Therefore, we analyzed whether the impact of SSL membership on being arrested for a shooting was mediated by additional contact. Results indicate that the total relationship between treatment and the outcome is 1.18, and most of that comes from the direct effect—not mediation (Fig. 6). In other words, the additional contact with police did not result in an increased likelihood for arrests for shooting, that is, the list was not a catalyst for arresting people for shootings. Rather, individuals on the list were people more likely to be arrested for a shooting regardless of the increased contact.

Discussion

Several important findings have emerged from the current pilot. First, while using arrestee social networks improved the identification of future homicide victims, the number was still too low in the pilot to make a meaningful impact on crime. The pilot version 1.0 of the model identified less than 1 % of homicide victims (3 out of 405) so there is certainly room for improvement. Second, the prevention strategy associated with the predictive strategy was not well developed and only led to increased contact with a group of people already in relatively frequent contact with police. As such, the main result of this study is that at-risk individuals were not more or less likely to become victims of a homicide or shooting as a result of the SSL, and this is further supported by city-level analysis finding no effect on the city homicide trend.

⁸ Most arrestees were not incapacitated for any significant period of time, but rather were booked into the Cook County jail and released within a few hours to a few days.

Table 3 Doubly robust treatment estimates

	Estimate (SD)	<i>t</i> test	Exp(b)
Shooting victim	-.221 (.397)	-.558, <i>p</i> = .58	.802
Shooting arrest	1.06 (.430)	2.46, <i>p</i> = .01*	2.88
Murder victim	.037 (.669)	.055, <i>p</i> = .96	1.04
Murder arrest	.453 (.462)	.981, <i>p</i> = .33	1.57
Any weapon outcome	.168 (.300)	.559, <i>p</i> = .58	1.18

We do find, however, that SSL subjects were more likely to be arrested for a shooting. The effect size was rather large, such that those placed on the SSL were 2.88 times more likely than their matched counterparts to be arrested for a shooting, although this is based on a small absolute number of shootings—only 9 individuals from the SSL were arrested for a shooting in the year after being placed on the list, against 5 from the matched control group (or 84 from the unmatched comparison group). This raises some questions as to whether this is a “positive” or “negative” finding—did the program lead to an increase in shootings perpetrated by SSL subjects due to some sort of backfire or harmful effect, similar to one that has been identified in prior research (McCord 2003; Sherman 1992; Welsh and Rocque 2014), or were they more likely to be arrested for a shooting they perpetrated because they were under more intense surveillance and intelligence-gathering activities?

We explore this question by examining the date of the shooting associated with the arrests, hypothesizing that a backfire effect could not occur *before* the introduction of the SSL. Unfortunately, the data do not lead to much clarification on this point because of missingness—there was a shooting date associated with the arrest in only 56 % ($n = 5$) of the SSLs (total $n = 9$) and 81 % ($n = 68$) of the unweighted comparison group (total $n = 84$). In the SSL group, 80 % of the shootings occurred after the intervention date compared to the 88 % of the eligible unweighted comparison cases. Therefore, based on these available data, there does not appear to be a difference in the timing of the shootings that resulted in arrests during our observation window between SSL members and to the unweighted comparison group (see Table 5).

While missing data make it impossible to confirm if the shootings associated with arrests occurred before or after treatment for 44 % of the SSL subjects, there are some reasons to believe the program did not increase shootings. If the existence of the SSL caused a backfire of shootings that would not have happened otherwise, committed by a group of people that the Chicago PD were more closely monitoring (and had significantly more contact with), we would have expected more mentions of the SSL

Table 4 Doubly-robust difference in contact with CPD between SSL subjects and their matched comparisons

	Estimate (SD)	<i>t</i> test	Exp(b)
Contact cards	.393 (.147)	2.68, <i>p</i> = .007	1.48
Arrest	.195 (.145)	1.34, <i>p</i> = .18	1.22
Any interaction	.332 (.134)	2.48, <i>p</i> = .01	1.39

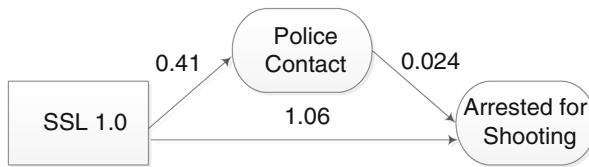


Fig. 6 Mediation analysis results

subjects during interviews and COMPSTAT meetings. We found no evidence such a surge occurred, either in the media or in our interviews, and interviewees were typically forthright about problems with the pilot.

Second, there is a large overlap between victims and offenders, and we find a program impact only on arrest for shootings and on none of the other indicators. Indeed, the risk of an SSL subject being a victim was lower than that of the control group, just not to a statistically significant level. Further evidence disputing the backfire theory is that we find no indirect relationship between increased police contact on being arrested for a shooting. Rather, it appears the list had a direct effect of increasing the probability of arrest, which did not work through police contact. We offer one explanation of this provided to us by the Chicago PD. They emphasize that the list was used as an intelligence-gathering source. Meaning, when there was a shooting, the police looked at the members of the SSL as possible suspects. This suggests that the impact of the SSL was on clearing shootings, but not on gun violence in general, during the observation window. However, a key question is why this did not lead to a reduction in the perpetration of gun violence.

A further discussion of drivers of the null, city-level result is also warranted. The effect on the city-wide homicide trend is difficult to detect because homicide was trending downward before the introduction of the SSL. In the year after the introduction of the SSL, a greater proportion of the 426 SSL individuals were victims of homicides (0.7 %) than the 17,754 individuals who were previously known to law enforcement but were not placed on the list (0.036 %) or the over 2 million Chicago residents without criminal records (0.003 %). However, controlling for demographics, criminal history, and co-arrest social network characteristics, those on the SSL were no less likely to be the victims of a homicide or a shooting than those who were not placed on the list. Again, it is not clear if this is due to the absence of a defined prevention strategy or a lack of impact because this sort of approach cannot work (e.g., theory failure vs. implementation failure).

The prediction model in this study was a first version model based on a simple calculation of the number of first- and second-degree co-arrestees who were homicide victims, inspired by prior work by Papachristos and colleagues (Papachristos 2009;

Table 5 Timing of shooting associated with post-treatment arrest

Group	Total arrests for shooting	Arrests with valid shooting date	# Shootings after intervention	% Shootings after intervention
Treatment ($n = 426$)	9	5	4	80 %
Unweighted comparison group ($n = 17,754$)	84	68	60	88 %

Papachristos et al. 2011, 2012). Although outside the scope of this study, there is clearly a question about how well the model performed in predicting homicide victimization. Those placed on the SSL were twice as likely to be victims of a homicide as others with arrest records, and 233 times more likely to be victims of homicide than the average Chicago resident. However, even with those increased odds, the individuals on SSL still only experienced a 0.7 % homicide rate over 12 months, illustrating how difficult it is to predict low-incidence events. Since the first list development, the algorithm has been improved, and now, according to ITT, 29 % of the top 400 subjects were accurately predicted to be involved in gun violence over an 18-month window (Lewin and Wernick 2015).

In terms of how to operationalize the predictions, the algorithm did not differentiate between “high-risk” versus “high-threat” individuals, which was found to be problematic in the field. There were no formal differentiations in the SSL predictive model between persons who were “high threat” in terms of being violent threats to their local community and persons who were “high risk” of becoming a victim based on their associations or lifestyle attributes, for example, substance use disorders and gambling, but who were not violent themselves. Assuming these two groups require different interventions, group identification would be necessary information for police to devise targeted prevention strategies. According to IIT and CPD, in the newer iteration of the SSL, specific risk factors are presented along with the list of individuals to facilitate the appropriate selection of an intervention.

Conclusions

Although the findings of this study are more relevant for cities, particularly those experiencing gang-related homicides, the conclusions do offer important, general insights into the potential implications of predictive policing programs. The pilot effort does not appear to have been successful in reducing gun violence, although it may have improved justice by identifying more perpetrators. It does not appear that there were any unintended crime consequences, such as a violent backfire effect. The SSL program continues to evolve through improvements in the statistical algorithms of the prediction model.

In terms of how the SSL program could be improved, it may be that both better predictions for the likelihood of being involved in gun violence and prevention strategies are necessary in order for an individual-based predictive policing program to work in the field. The discrepancy between observed outcomes and predicted risk is operationally significant, but statistically reasonably small given the difficulty of predicting a low-probability event. Also, it may be that the absolute quantification of the risk factor is less important in practice than the relative ranking of the subjects, which we did not attempt to measure. Since this study, based on version 1.0 of the model, the researchers who developed the SSL algorithm have progressively improved the predictive performance of the model, which is currently in version 4 (Lewin and Wernick 2015).

Regardless of how “well” a model performs, there will always be a concern with misidentifying people as not going to commit or be a victim of gun violence (false negatives) and misidentification of people as going to commit or be a victim of gun

violence (false positives). This problem is well researched and the problem with false negatives and positives must be considered in terms of what it is being used for (Aitchison and Dunsmore 1980; Benjamini and Hochberg 1995). Models can be specified to penalize false negatives or positives more strongly, and it could be that a criminal justice application would have a different threshold than, for instance, a medical screening tool (Berk 2011). The problem is actually a fairly standard issue of modern statistical practice of large-scale multiple testing in the analysis of high-dimensional data. In their study to identify potentially problematic officers, for example, Ridgeway and MacDonald (2009) used a false discovery rate approach and selected a threshold of 0.50, which implies that the cost of failing to identify a problem officer equals the cost of flagging a good officer. Analysts and program developers should consider this trade-off for their purposes. Additionally, the consideration of this trade-off and thresholds should explicitly take into account the prevention strategy—if risk scores are used for a benign (and inexpensive) treatment, false positives are less of a problem; however, once the treatment becomes invasive, detrimental, and expensive, these false positives can become a huge problem. Conversely, false negatives in this application can be very costly, as they represent a missed opportunity to prevent shootings.

Second, and perhaps more importantly, law enforcement needs better information about what to do with the predictions—the “prevention” part of predictive policing. Indeed, the district commanders may have been cautious about intervening without the explicit direction of more senior administration, since these individuals were initially identified as potential victims. As this study shows, by providing almost no guidance to district commanders and police officers in the field, either up front or on an ongoing basis, they all generally opted for recommending to their officers to increase “contact” with individuals on the list in varying forms and levels of effort. And our analysis shows the officers did just that—we find a statistically significant increase in police contacts. However, it is not at all evident that contacting people at greater risk of being involved in violence—especially without further guidance on what to say to them or otherwise how to follow up—is the relevant strategy to reduce violence. Alternatively, we did not find convincing evidence that increased contact resulted in a backfiring effect, but the possibility cannot be ruled out.

The finding that the list had a direct effect on arrest, rather than victimization, raises privacy and civil rights considerations that must be carefully considered, especially for predictions that are targeted at vulnerable groups at high risk of victimization. Both local and national media openly ask whether the CPD SSL pilot constitutes racial profiling (Erbenraut 2014; Llenas 2014; Stroud 2014). A review of the legal and constitutional issues involved in using predictions for criminal justice purposes notes that, while it is not legal to use protected classes as predictors (Starr 2014), classifications that have differential impact on different protected classes, such as racial groups, that are not designed to have this impact, are legal (Tonry 1987). Tonry (1987) also argues that the ethical issues with using prediction in a policing context are less controversial than in other criminal justice settings because they are necessary for the cost-effective distribution of scarce resources, and their decisions will ultimately be reviewed by impartial judges before punishment is delivered. However, using predictions to identify individuals in the community for increased police scrutiny has not been subject to judicial review.

Acknowledgments We would like to thank the Chicago Police Department and Dr. Miles Wernick from the Illinois Institute of Technology for their participation and support of this evaluation. We would also like to acknowledge research assistance provided by Sam Cooper and Alessandra Sienra-Canas. This publication was made possible by Award Number 2009-IJ-CX-K114 - Predictive Policing Analytic & Evaluation Research Support awarded by the National Institute of Justice, Office of Justice Programs. The opinions, findings, conclusions and recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.

References

- Abrahamse, A. F., Ebener, P. A., Greenwood, P. W., Fitzgerald, N., & Kosin, T. E. (1991). An experimental evaluation of the Phoenix repeat offender program. *Justice Quarterly*, 8(2), 141–168.
- Aitchison, J., & Dunsmore, I. R. (1980). *Statistical prediction analysis*. CUP Archive.
- Auerhahn, K. (1999). Selective incapacitation and the problem of prediction. *Criminology*, 37(4), 703–734.
- Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3), 211–233.
- Beck, C., & McCue, C. (2009). Predictive policing: what can we learn from Wal-Mart and Amazon about fighting crime in a recession? *Police Chief*, 76(11), 18.
- Becker, G. S. (1993). Nobel lecture: the economic way of looking at behavior. *Journal of Political Economy*, 101, 385–409.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, 57, 289–300.
- Berk, R. (2008). Forecasting methods in crime and justice. *Annual Review of Law and Social Science*, 4, 219–238.
- Berk, R. (2011). Asymmetric loss functions for forecasting in criminal justice settings. *Journal of Quantitative Criminology*, 27(1), 107–123.
- Berk, R., & Bleich, J. (2013). Statistical procedures for forecasting criminal behavior. *Criminology and Public Policy*, 12(3), 513–544.
- 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 the Royal Statistical Society: Series A (Statistics in Society)*, 172(1), 191–211.
- Berry, M. J., & Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. Indianapolis, IN: John Wiley & Sons.
- Blumstein, A. (1986). *Criminal Careers and “Career Criminals”* (Vol. 2). Washington, DC: National Academies.
- Bordua, D. J., & Reiss, A. J., Jr. (1966). Command, control, and charisma: reflections on police bureaucracy. *American Journal of Sociology*, 72, 68–76.
- Braga, A. (2005). Hot spots policing and crime prevention: a systematic review of randomized controlled trials. *Journal of Experimental Criminology*, 1(3), 317–342.
- Braga, A., & Weisburd, D. L. (2010). *Policing problem places: Crime hot spots and effective prevention*. New York, NY: Oxford University Press on Demand.
- Braga, A., & Weisburd, D. L. (2012). The effects of focused deterrence strategies on crime: a systematic review and meta-analysis of the empirical evidence. *Journal of Research in Crime and Delinquency*, 49(3), 323–358.
- Braga, A., Papachristos, A. V., & Hureau, D. M. (2012). The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly*, 1–31.
- Bruinsma, G., & Weisburd, D. (2014). *Encyclopedia of Criminology and Criminal Justice*. New York: Springer.
- Caldwell, M. F., Vitacco, M., & Van Rybroek, G. J. (2006). Are violent delinquents worth treating? A cost-benefit analysis. *Journal of Research in Crime and Delinquency*, 43(2), 148–168.
- Chicago Police Department. (2014). *Gang Violence Reaction Strategy: General Order G10-01*. Chicago, IL.
- Chinman, M., Imm, P., & Wandersman, A. (2004). *Getting To Outcomes™ 2004*. Santa Monica: Rand Corporation.
- Cohen, J., Gorr, W. L., & Olligschlaeger, A. M. (2007). Leading indicators and spatial interactions: a crime-forecasting model for proactive police deployment. *Geographical Analysis*, 39(1), 105–127.

- Cope, N. (2004). 'Intelligence led policing or policing led intelligence?' Integrating volume crime analysis into policing. *British Journal of Criminology*, 44(2), 188–203.
- Cornish, D. B., & Clarke, R. V. (2014). *The reasoning criminal: Rational choice perspectives on offending*. New Brunswick, NJ: Transaction Publishers.
- Dvoskin, J. A., & Heilbrun, K. (2001). Risk assessment and release decision-making: Toward resolving the great debate. *American Academy of Psychiatry and the Law*, 29, 6–10.
- Eck, J., Chainey, S., Cameron, J., & Wilson, R. (2005). *Mapping crime: Understanding hotspots* (Vol. NCJ 209393). Washington, DC: National Institute of Justice.
- Erbentraut, J. (2014). Chicago's controversial new police program prompts fear of racial profiling. *The Huffington Post*.
- Foster, E. M., & Jones, D. (2006). Can a costly intervention be cost-effective?: an analysis of violence prevention. *Archives of General Psychiatry*, 63(11), 1284–1291.
- Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American Journal of Epidemiology*, 173(7), 761–767.
- Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: what works! *Criminology*, 34(4), 575–608.
- Gorr, W., & Harries, R. (2003). Introduction to crime forecasting. *International Journal of Forecasting*, 19(4), 551–555.
- Gottfredson, M., & Hirschi, T. (1986). The true value of Lambda would appear to be zero: an essay on career criminals, criminal careers, selective incapacitation, cohort studies, and related topics*. *Criminology*, 24(2), 213–234.
- Greenwood, P. W., & Abrahamse, A. F. (1982). *Selective incapacitation*. Santa Monica: Rand Corporation.
- Groff, E. R., & La Vigne, N. G. (2002). Forecasting the future of predictive crime mapping. *Crime Prevention Studies*, 13, 29–58.
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293.
- Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 1, 799–821.
- Hunt, P., Saunders, J., & Hollywood, J. S. (2014). *Evaluation of the Shreveport Predictive Policing Experiment*. Santa Monica: RAND Corporation.
- Kang, J. D., & Schafer, J. L. (2007). Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 25, 523–539.
- Kennedy, D. M. (1996). Pulling levers: chronic offenders, high-crime settings, and a theory of prevention. *Valparaiso University Law Review*, 31, 449.
- Kovandzic, T. V., Sloan, J. J., III, & Vieraitis, L. M. (2004). "Striking out" as crime reduction policy: the impact of "three strikes" laws on crime rates in US cities. *Justice Quarterly*, 21(2), 207–239.
- Lewin, J., & Wernick, M. (2015). *Chicago Police Department Data Analytics and Predictive Policing*. Paper presented at the International Association of Chief's of Police, Chicago, IL.
- Lipsey, M. W. (1999). Can intervention rehabilitate serious delinquents? *The Annals of the American Academy of Political and Social Science*, 564(1), 142–166.
- Litwack, T. R., & 2. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law*, 7, 409.
- Llenas, B. (2014). Brave New World of "Predictive Policing" Raises Specter of High-Tech Racial Profiling, *Fox News Latino*.
- Loeber, R., & Farrington, D. P. (1998). *Serious and violent juvenile offenders: Risk factors and successful interventions*. Thousand Oaks, CA: Sage Publications.
- Lum, C., Koper, C. S., & Telep, C. W. (2011). The evidence-based policing matrix. *Journal of Experimental Criminology*, 7(1), 3–26.
- Martin, S. E., & Sherman, L. W. (1986). Selective apprehension: a police strategy for repeat offenders. *Criminology*, 24, 155.
- Mazerolle, L. G., Kadleck, C., & Roehl, J. (1998). Controlling drug and disorder problems: the role of place managers. *Criminology*, 36(2), 371–404.
- Mazerolle, L. G., Ready, J., Terrill, W., & Waring, E. (2000). Problem-oriented policing in public housing: the Jersey City evaluation. *Justice Quarterly*, 17(1), 129–158.
- McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods*, 9(4), 403.
- McCleary, R., Hay, R. A., Meidinger, E. E., & McDowall, D. (1980). *Applied time series analysis for the social sciences*. Beverly Hills: Sage Publications.

- McCord, J. (2003). Cures that harm: unanticipated outcomes of crime prevention programs. *The Annals of the American Academy of Political and Social Science*, 587(1), 16–30.
- McGarrell, E. F., Chermak, S., Wilson, J. M., & Corsaro, N. (2006). Reducing homicide through a “lever-pulling” strategy. *Justice Quarterly*, 23(02), 214–231.
- Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., & Brantingham, P. J. (2015). Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, 110(512), 1399–1411.
- Ohri, A. (2013). *Forecasting and Time Series Models R for Business Analytics* (pp. 241–258). Springer.
- Papachristos, A. (2009). Murder by structure: dominance relations and the social structure of gang homicide. *American Journal of Sociology*, 115(1), 74–128.
- Papachristos, A. V., & Kirk, D. S. (2015). Changing the street dynamic. *Criminology and Public Policy*, 14(3), 525–558.
- Papachristos, A., Braga, A., & Hureau, D. (2011). Six-degrees of violent victimization: Social networks and the risk of gunshot injury.
- Papachristos, A., Braga, A., & Hureau, D. (2012). Social networks and the risk of gunshot injury. *Journal of Urban Health*, 89(6), 992–1003.
- Pate, T., Bowers, R. A., & Parks, R. (1976). *Three approaches to criminal apprehension in Kansas City: An evaluation report*. Washington, DC: Police Foundation.
- Perry, W. L., McInnis, B., Price, C. C., Smith, S. C., & Hollywood, J. S. (2013). *Predictive policing: The role of crime forecasting in law enforcement operations*. Santa Monica, CA: Rand Corporation.
- Quinsey, V. L., Harris, G. T., & Rice, M. E. (2000). Violent Offenders: Appraising and Managing Risk. *Psychiatric Services*, 51(3), 395
- Ratcliffe, J. (2002). Intelligence-led policing and the problems of turning rhetoric into practice. *Policing and Society*, 12(1), 53–66.
- Ratcliffe, J. (2005). The effectiveness of police intelligence management: a New Zealand case study. *Police Practice and Research*, 6(5), 435–451.
- Ratcliffe, J. H. (2012). *Intelligence-led policing*. New York, NY: Routledge.
- Ratcliffe, J. H., & Guidetti, R. (2008). State police investigative structure and the adoption of intelligence-led policing. *Policing: An International Journal of Police Strategies and Management*, 31(1), 109–128.
- Ridgeway, G. (2013). Linking prediction and prevention. *Criminology and Public Policy*, 12(3), 545–550.
- Ridgeway, G., & MacDonald, J. M. (2009). Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association*, 104(486), 661–668.
- Ridgeway, G., Braga, A. A., Tita, G., & Pierce, G. L. (2011). Intervening in gun markets: an experiment to assess the impact of targeted gun-law messaging. *Journal of Experimental Criminology*, 7(1), 103–109.
- Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., & Griffin, B. A. (2014). Toolkit for Weighting and Analysis of Nonequivalent Groups: A tutorial for the twang package. *R vignette. RAND*.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.
- Sailor, W., Dunlap, G., Sugai, G., & Homer, R. (2008). *Handbook of positive behavior support*. New York, NY: Springer.
- Sherman, L. W. (1986). Policing communities: what works? *Crime and justice*, 343–386.
- Sherman, L. W. (1992). The influence of criminology on criminal law: evaluating arrests for misdemeanor domestic violence. *Journal of Criminal Law and Criminology*, 83, 1–45.
- Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review*, 49, 261–272.
- Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime “hot spots”: a randomized, controlled trial. *Justice Quarterly*, 12(4), 625–648.
- Sherman, L. W., Gottfredson, D., MacKenzie, D., Eck, J., Reuter, P., & Bushway, S. (1997). *Preventing crime: What works, what doesn't, what's promising: A report to the United States Congress*. Washington, DC: US Department of Justice, Office of Justice Programs.
- Silver, E., & Miller, L. L. (2002). A cautionary note on the use of actuarial risk assessment tools for social control. *Crime and Delinquency*, 48(1), 138–161.
- Starr, S. B. (2014). Evidence-based sentencing and the scientific rationalization of discrimination. *Stanford Law Review*, 66, 803–953.
- Stroud, M. (2014). The Minority Report: Chicago's new police computer predicts crimes, but is it racist? *The Verge*.
- Tonry, M. (1987). Prediction and classification: legal and ethical issues. *Crime and Justice*, 9, 367–413.
- Weisburd, D., & Mazerolle, L. G. (2000). Crime and disorder in drug hot spots: implications for theory and practice in policing. *Police Quarterly*, 3(3), 331–349.

- Welsh, B. C., & Rocque, M. (2014). When crime prevention harms: a review of systematic reviews. *Journal of Experimental Criminology*, 10(3), 245–266.
- Wright, K. N., Clear, T. R., & Dickson, P. (1984). Universal applicability of probation risk-assessment instruments. *Criminology*, 22(1), 113–134.
- Yang, M., Wong, S. C., & Coid, J. (2010). The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*, 136(5), 740.
- Zikopoulos, P., & Eaton, C. (2011). *Understanding big data: Analytics for enterprise class hadoop and streaming data*. McGraw-Hill Osborne Media.

Jessica Saunders is a Senior Criminologist at RAND and Professor with Pardee RAND Graduate School. Her research focuses on policing and policing reform, school safety, and criminal justice policy evaluation.

Priscillia Hunt is an Economist at RAND, Professor with Pardee RAND Graduate School, and Research Fellow of the Institute for the Study of Labor (IZA). Her research interests focus on the economics of crime, including studies of criminal behavior, assessing operations of the criminal justice system, and evaluating impacts of criminal justice policy.

John S. Hollywood is a senior operations researcher at the RAND Corporation. He conducts research on criminal justice technologies, to include leading a general technical assessment of predictive policing for the National Institute of Justice. He has also led multiple studies to identify and prioritize top science and technology-related needs for criminal justice for NIJ.