

October 2020

ISSUE BRIEF

THE USE OF PREDICTIVE ANALYTICS IN POLICING

Benjamin Carleton | Brittany Cunningham | Zoë Thorkildsen



CNA

BJA
Bureau of Justice Assistance
U.S. Department of Justice



This document contains the best opinion of CNA at the time of issue.

This work was performed under BJA-2018-DP-BX-K015. This project was supported by Cooperative Agreement Number BJA-2018-DP-BX-K015 awarded by the Bureau of Justice Assistance, U.S. Department of Justice. The opinions contained herein are those of the author(s) and do not necessarily represent the official position or policies of the U.S. Department of Justice. References to specific agencies, companies, products, or services should not be considered an endorsement by the author(s) or the U.S. Department of Justice. Rather, the references are illustrations to supplement discussion of the issues.

Recommended citation:

Carleton, B., Cunningham, B., & Thorkildsen, Z. (2020). The Use of Predictive Analytics in Policing. Arlington, VA: CNA.

This issue brief is a product created through a partnership between BJA and CNA. The Using Analytics to Improve Officer Safety work examines granular incident data from 2015–2019 from several local law enforcement agencies to identify incident characteristics (characteristics specific to the incident and related to officer tactical response) associated with officer assaults, injuries, and line-of-duty deaths. Using machine learning techniques, CNA is producing a risk assessment model to link incident characteristics with officer safety outcomes. This work also entails working with participating agencies to identify best practices and recommendations to reduce risks to officer safety in the line of duty.

Table of Contents

Introduction.....	1
Review of existing sources on the use of predictive analytics in policing	2
Predictive Analytics in Policing	3
Approaches to predictive analytics in policing	3
Place-based approaches	4
Examples of place-based approaches	6
Example 1: Indio, California, Strategies for Policing Innovation: Reducing Burglaries Through Predictive Policing and Community Engagement.....	6
Example 2: Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting	6
Person-based approaches.....	7
Example of person-based approach	9
Example 1: The RASOR’S Edge: Focused Deterrence in Cambridge, Everett, and Somerville.....	9
Machine Learning in Policing.....	10
Examples of machine learning used to improve policing practice	12
Example 1: Developing a Practical Forecasting Screener for Domestic Violence Incidents	12
Example 2: Identifying Police Officers at Risk of Adverse Events	13



Table of Contents

Summary of Studies	14
Considerations and Limitations of Predictive Analytics	15
Practical Considerations for Implementation of Predictive Analytics.....	18
Goal of Analytics	18
Data Availability	18
Capabilities and Resources	19
Summary	21
References	22

Introduction

Policing is an evolving field; law enforcement agencies are being asked to do more with limited resources, forcing agencies and their relevant stakeholders (e.g., policy makers, other justice system agencies, community organizations) to continuously look for new ways to reduce crime, keep communities safe, and effectively allocate resources. The use of predictive analytics has evolved in the last

several decades as a promising response to reduce and prevent crime. Predictive analytics in policing “is a data-driven approach to characterizing crime patterns across time and space and leveraging this knowledge for the prevention of crime and disorder” (Fitzpatrick et al. 2019).

Traditionally, law enforcement agencies have operated using primarily reactive measures, such as rapid responses to 911 calls, random patrols, and a greater focus on criminal investigations (Brayne 2017; Fitzpatrick et al. 2019). To operate more proactively, agencies have increasingly employed predictive analytics that informs crime prevention strategies. For example, agencies across the US have implemented a number of strategies (e.g., hot spot detection, targeted offender lists, and risk terrain modeling) and software programs that use a variety of predictive analytics to forecast where and when crimes are most likely to occur and to identify offenders and groups or individuals at risk of becoming victims of crimes.

Predictive analytics builds on traditional crime analysis practices (e.g., identification of crime trends and patterns). In addition to identifying crime trends and patterns based on crimes that have already occurred, predictive analytics goes a step further, forecasting where and when crime is likely to occur or who is likely to be involved in criminal behavior. It equips agencies with knowledge (i.e., data) to help inform where they should target police operations and resources. Agencies can use this knowledge to operate more efficiently and effectively in their crime reduction efforts and resource allocations. It is important to understand that predictive analytics cannot tell the future very well. These predictions rely on past data and assume that future criminal activity will be similar to that reflected in extant data (sometimes factoring in anticipated future changes). This reliance on past data also means that predictive techniques can reinforce systemic bias, racial and otherwise, present in past justice system actions.

Predictive analytics in policing “is a data-driven approach to characterizing crime patterns across time and space and leveraging this knowledge for the prevention of crime and disorder.”

Introduction

We limited the scope of the review to the past 15 years given more recent advancements in predictive technologies used by law enforcement.

The objective of this brief is to provide an accessible resource for law enforcement agencies and their stakeholders (e.g., crime analysts, policy makers, and researchers) interested in learning more about the role of predictive analytics in police operations.

Specifically, this brief offers the following:

- Summarizes the use of predictive analytics to inform policing operations
- Distinguishes between approaches to predictive analytics (person-based and place-based)
- Highlights the emergence of machine learning algorithms as a preferred predictive analytics technique
- Delineates considerations and limitations brought forth in recent literature that law enforcement agencies must consider when using predictive analytics to reduce and prevent crime
- Summarizes several research studies and real-world policing initiatives as examples of how the use of predictive analytics can inform policing practice

REVIEW OF EXISTING SOURCES ON THE USE OF PREDICTIVE ANALYTICS IN POLICING

This brief draws on a review of academic literature, US Department of Justice technical research reports, institutional reports, and news articles published within the last 15 years (2005–2020). We limited the scope of the review to the past 15 years given more recent advancements in predictive technologies used by law enforcement. This is not an exhaustive summary of all available resources on predictive analytics; rather this brief summarizes several commonly cited resources that agencies can use to foster discussions, gain knowledge, and build the capacity to adopt the use of predictive analytics if desired.

Predictive Analytics in Policing

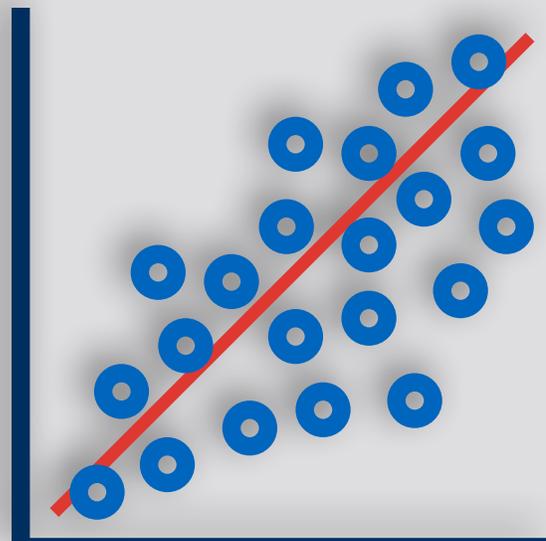
Similar to the evolving nature of policing, predictive analytics has also evolved in its complexity and capabilities, from conventional predictive analytics, such as regression analysis, to more advanced statistical modeling, such as machine-learning algorithms to predict crime and potential offenders and victims. Regression analysis, demonstrated in Figure 1, is a common statistical analysis method to examine the mathematical relationship between observations and determine whether a set of variables does a good job of predicting outcomes for those observations. Machine-learning algorithms find patterns in large amounts of data, do so without being programmed with exact instructions, and then make predictions. As we noted, predictive analytics does not have the power to predict exactly where and when a crime will occur or pinpoint who will commit the next crime; rather **predictive analytics uses past data to inform the risk or probability of such outcomes, accepting certain assumptions inherent in the past data.** Further, because future criminal activity will never be the same as past activity, these techniques cannot produce exact predictions (Perry et al. 2013). The use of predictive analytics creates the potential for unfairly targeting particular communities. We discuss this issue of bias below, along with other noted limitations.

APPROACHES TO PREDICTIVE ANALYTICS IN POLICING

Predictive analytics is often divided into two overarching approaches when referenced in policing: place-based and person-based (Brayne et al. 2015; Ferguson 2019; Fitzpatrick et al. 2019; Moses and Chan 2018; Perry et al. 2013; Selbst 2017).

1. Place-based approaches aim to predict where (geographically) and at what times crime is most likely to occur.
2. Person-based approaches identify offenders, perpetrators, and groups or individuals most at risk of committing crimes or becoming victims of crimes.

FIGURE 1. REGRESSION ANALYSIS



Predictive Analytics in Policing

To further delineate the two approaches, the next section provides descriptions of different predictive analytics techniques for each approach, and examples of how the two approaches have been used in research and in practice.

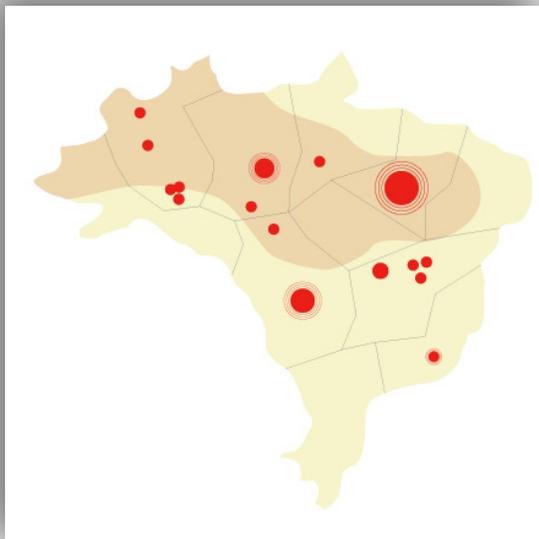
PLACE-BASED APPROACHES

One of the earliest place-based approaches that law enforcement agencies adopted was **hot spot detection** (Fitzpatrick et al. 2019; Perry et al. 2013; Selbst 2017). Hot spot detection, also referred to as “hot spots policing” (Braga and Weisburd 2010) involves mapping crime locations to pinpoint geographically defined clusters of crime, known as “hot spots” (Selbst 2017). Figure 2 shows a sample hot spot detection map, with larger circles representing clusters of hot spots. Crime mapping can be completed using a simple pin map or computerized crime mapping software. Hot spot detection moves agencies away from geographic-based patrols or deployments and toward incident-based patrols or deployments; allocating an agency’s resources to areas that most need them, or most request them. Ideally, the increased focus on hot spots will help prevent future crime in the identified areas and introduce efficiency in police

work because the areas most affected by crime receive the most resources and attention. Critics of hot spot detection argue that place-based enforcement displaces crime from the targeted area to surrounding areas (Reppetto 1976). However, evidence from several studies does not support the argument of crime displacement, suggesting instead that hot spot detection is “more likely to be associated with the diffusion of crime control benefits in surrounding areas rather than crime displacement” (Braga et al. 2012; Braga and Bond 2008; Weisburd et al. 2006).

Since the adoption of hot spot detection, another more advanced predictive technique has been developed to help

FIGURE 2. HOT SPOT DETECTION SAMPLE



Predictive Analytics in Policing

forecast and reduce crime—risk terrain modeling (RTM). RTM starts with the identification of “all factors that are related to a particular outcome for which risk is being assessed” (Caplan et al. 2011). For example, if the outcome were future burglaries, the analysis would include the presence of retail businesses as one of the relevant factors. **After the relevant factors have been identified, “RTM assigns a value signifying the presence, absence or intensity of each factor at every place throughout a given geography” (Caplan et al. 2011). Each factor is represented by a separate map of the same geography (shown in Figure 3), and the maps of the different factors are combined to create the risk terrain map.** Once the maps have been combined, a risk value is assigned to all locations based on the identified factors. A higher risk value indicates a greater likelihood of a crime occurring in that location (Caplan et al. 2011).

Risk terrain modeling (RTM) starts with the identification of “all factors that are related to a particular outcome for which risk is being assessed.”

RTM has helped pave the way for the development of software programs that use machine-learning algorithms to predict where and when crime is likely to occur. While some predictive software programs only analyze historical crime data (i.e., crime type, time, and location) (Brayne 2017; Ferguson 2019; Joh 2017; Kutnowski 2017; Lum and Isaac 2016; Selbst 2017), others pull from the analytical method of RTM, and incorporate variables such as weather patterns, the schedules of major sporting events, and school calendars, in addition to historical crime data (Brayne et al. 2015; Ferguson, 2019; Joh, 2017). These software programs help law enforcement agencies identify areas where future crime is more likely to occur and allocate patrol operations effectively (Brayne 2017).

FIGURE 3. SAMPLE MAP LAYERS USED IN RTM



Predictive Analytics in Policing

EXAMPLES OF PLACE-BASED APPROACHES

EXAMPLE 1: [Indio, California, Strategies for Policing Innovation: Reducing Burglaries Through Predictive Policing and Community Engagement](#)

The Indio, California, Police Department (Indio PD) implemented a Bureau of Justice Assistance (BJA)-funded Strategies for Policing Innovation (SPI) program addressing truancy that used predictive analytics to address burglaries in the community (Parker and Martinez 2014). In collaboration with Dr. Robert Nash Parker, a criminologist from the University of California, Riverside, Indio PD developed a computer model that predicts, by census block group, where burglaries are likely to occur within the City of Indio. The model incorporated an RTM approach that included data on socioeconomic characteristics of the block group, probation data, data on school absences, truancy arrests, and historical (lagged) truancy rates. The model showed a strong relationship between historical truancy rates and current burglary rates. In response, Indio PD implemented an aggressive public information campaign and educational outreach program to counter predicted truancy in the community. Interrupted time series analysis of the impact of the holistic truancy reduction efforts found reductions in burglaries after the program was implemented.

EXAMPLE 2: [Risk Terrain Modeling: Brokering Criminological Theory and GIS¹ Methods for Crime Forecasting](#)

In a study of shootings (2011), Caplan and colleagues used RTM to forecast the risk of future shootings in Irvington, New Jersey. Data were divided into three six-month time periods: January to June 2007 (Period 1), July to December 2007 (Period 2), and January to June 2008 (Period 3). The researchers compared risk terrain maps using Period 1 and Period 2 data with actual shooting incidents from Period 2 and Period 3, respectively. To create the two risk terrain maps, Caplan and colleagues first identified three relevant factors previously found to predict shootings: dwellings of known gang members, locations of retail business infrastructure, and locations of drug arrests. The researchers created separate density maps of each factor and assigned risk values to locations based on the density of the particular factor. The maps display density value using a color scale, with lighter shades indicating lower risk and darker shades indicating greater risk. The separate factor maps combine to create two risk terrain maps, which use a similar color scale. The first risk terrain map showed Period 1 risk terrain with actual shooting incidents from Period 2 mapped. The second risk terrain map showed Period 2 risk terrain with

¹ Geographic Information System

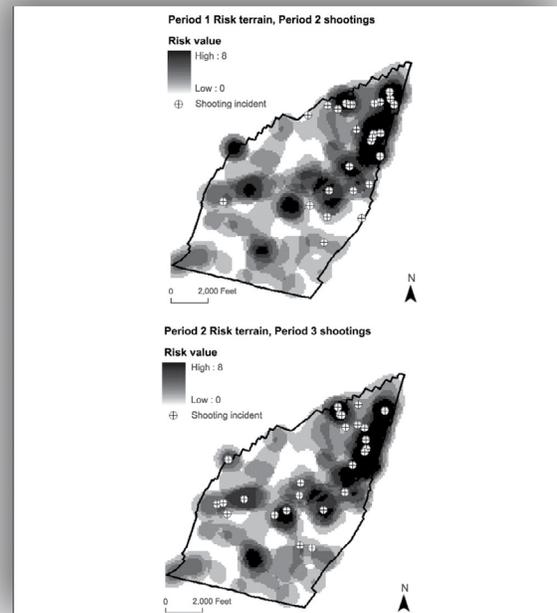
Predictive Analytics in Policing

actual shooting incidents from Period 3 mapped. As depicted in Figure 4, the future shooting incidents appeared to be located in higher risk locations. Results from logistic regression analysis for the Period 1 risk terrain suggested that for every increased unit of risk (on a scale of 0 to 8), the likelihood of a shooting significantly increased by at least 69 percent. For the Period 2 risk terrain, this figure was 56 percent.

PERSON-BASED APPROACHES

Person-based approaches look specifically at data and information related to offenders and victims of crimes and aim to predict who is most likely to be a future offender or victim. Because victimization is closely tied to one's proximity to at-risk groups and individuals and at-risk locations, many of the place-based approaches (e.g., hot spot detection) and offender-focused approaches (e.g., targeted offender lists) can also be used to predict likely victims of crime (Perry et al. 2013). One type of person-based approach is the creation of targeted offender lists (Garrett 2018; Joh 2017; Selbst 2017). **Targeted offender lists** focus on individuals who are known to be chronic offenders or are most likely to offend in the future. The Chicago Police Department's (CPD) targeted offender list, also known as the "heat list" or "Strategic Subject List," used an algorithm to identify individuals most at risk of being offenders or victims of crime in the future (Joh 2017; Selbst 2017). After much criticism claiming that the list unnecessarily and unfairly targeted people for police attention, CPD adapted how it used the list (Hollywood 2016; Stroud 2016). CPD transitioned to also using it for "custom notification," in which CPD actively conducted outreach

FIGURE 4. RISK TERRAIN WITH ACTUAL MAPPED SHOOTINGS



Source: Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). Risk terrain modeling: Brokerage criminological theory and GIS methods for crime forecasting. *Justice Quarterly*, 28(2), 360-381.

Predictive Analytics in Policing

to individuals on the list to notify them that they were on the department's radar as persons at great risk for victimization, and to warn them that future criminal activity would be prosecuted (Hollywood 2016; Joh 2017). A 2016 study of CPD's list found that individuals on the list were not more or less likely to become victims of a homicide or a shooting than the comparison group (Saunders et al. 2016). However, those on the list were more likely to be arrested for shootings. Both of these results raised questions about the list's effectiveness and purpose. Subsequently, the list was again revised and then renamed to the Crime and Victimization Risk Model (CVRM) (Hollywood et al. 2019). Despite revisions, a 2019 evaluation found that the new CVRM tool was "not operationally suitable" (Hollywood et al. 2019).

The Los Angeles Police Department (LAPD) used a different approach to targeting potential offenders. Through a points-based system, LAPD kept track of "the worst of the worst" offenders (Brayne 2017). Individuals received one point for certain outcomes, such as police contact; a greater number of points demonstrated a greater threat to the community (Brayne 2017; Selbst 2017). However, the uses of both CPD's heat list and LAPD's point-based system came with considerable controversy. Questions about whether these techniques are fair or biased were raised, and arguments that these techniques unfairly target minority communities were made (Ferguson 2017; Puente 2019). These questions and arguments, along with public outcry, led the LAPD to end its point-based system in August of 2018 (Puente 2019) and CPD to end its "Strategic Subject List" in November of 2019 (Gorner and Sweeney 2020).

Like LAPD's point-based system, a software program was developed by a private company to help improve officer safety in the field by producing a threat score for a person or address when a 911 call comes in. This threat score is based on a number of factors including criminal history, address, and social media use (Joh 2017; Selbst 2017). The software program is used when responding to 911 calls, rather than to create a running targeted offender list.

Predictive Analytics in Policing

EXAMPLE OF PERSON-BASED APPROACH

EXAMPLE 1: [The RASOR'S Edge: Focused Deterrence in Cambridge, Everett, and Somerville](#)

Through funding from the BJA SPI program, Cambridge, Everett, and Somerville, Massachusetts, collaborated on the Regional Analytics for the Safety of Our Residents (RASOR) effort (Uchida et al. 2016). RASOR implemented person-based predictive policing methods as an innovative method of identifying individuals for intervention using focused deterrence. The three agencies shared information about impact players, habitual offenders, crime data, known associates, and other material, contributing to the development of a predictive model for identifying high-risk individuals operating across the three communities. Research partners from Justice & Security Strategies, Inc., created a predictive model using data on about 280,000 individuals known to the three agencies; this model was used to calculate a social harm score for 150 individuals for a rigorous evaluation of focused deterrence. The social harm score was effective in identifying high-risk individuals, though the evaluation of the focused deterrence effort returned mixed results.

Machine Learning in Policing

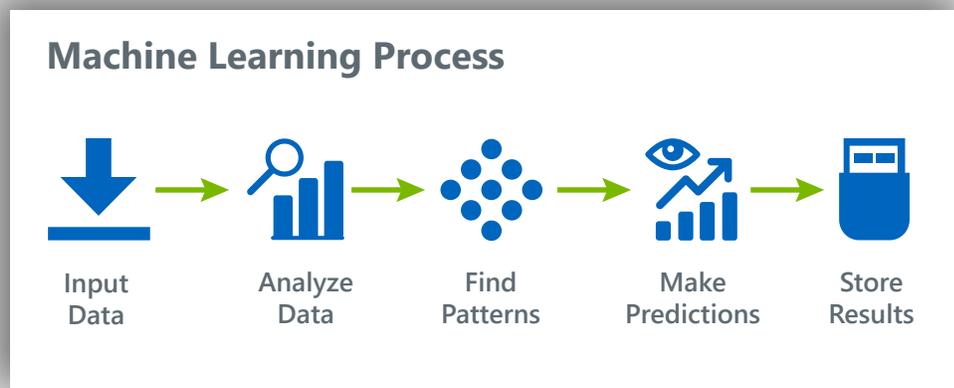
The goal of machine learning is “to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data...”

Predictive analytics in law enforcement relies on the limits of the data and the analytic tools (e.g., regression, RTM), the statistical capabilities of crime analysts, and the intended goal (e.g., detect hot spots, target future offenders or victims) of the chosen analytic technique. We continually see agencies transitioning from conventional predictive analytics (e.g., regression) to more advanced

techniques such as machine learning. Machine learning’s prospects at improving prediction accuracy and its better suitability for the complex nature of criminological data have helped to spur this transition.

The goal of machine learning is “to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest” (Murphy 2012). Figure 5 shows the basic steps of the machine learning process. In the context of policing, past crime data is analyzed using a machine-learning algorithm to detect patterns and relationships that conventional predictive analytics is unable to uncover. Then, the data pertaining to the uncovered patterns are further analyzed to improve predictions.

FIGURE 5. MACHINE LEARNING PROCESS



Machine Learning in Policing

Machine learning can overcome several limitations of conventional predictive analytics. One such limitation is that the conventional regression model cannot effectively incorporate costs into the model (Berk and He 2005). *Cost*, in this sense, refers to the consequences of making incorrect predictions—that is, the false positives or false negatives that can result from the predictive analysis. Berk and Bleich (2013) stress that the consequences or weight of false negatives and false positives are not the same and that costs need to be incorporated into predictive models (Berk and Bleich 2013; Berk and He 2005). For example, as described in Berk and Bleich (2013), when forecasting parole success for individuals, the cost of paroling an individual who will fail and may commit a serious crime is not equal to the cost of denying parole for an individual that will succeed.

Machine learning offers a number of helpful capabilities. Below, we highlight several reasons that machine learning is emerging as a preferred statistical method over other predictive analytics techniques. The reasons mentioned do not represent all of the potential benefits of machine learning, but do summarize the major themes from existing literature.

- **PATTERN IDENTIFICATION:** Machine-learning algorithms are able to identify patterns and find relationships in large datasets that a human would not traditionally identify (Babuta et al. 2018; Selbst 2017). Based on identified patterns and relationships, machine-learning algorithms continuously adapt and are able to discover additional patterns (Berk and Bleich 2013; Lum and Isaac 2016). In policing, the identification of patterns might inform an agency about the times of day when a particular neighborhood is more likely to experience a specific type of crime.
- **DATA COMPLEXITY:** Much of the literature on predictive analytics in law enforcement stresses the complex nature of criminological data (Berk and Bleich 2013; Brennan and Oliver 2013; Carton et al. 2016). Because human behavior and neighborhoods, among other characteristics that influence crime, are dynamic, linear and more conventional techniques are not appropriate for the complexity of criminological data (Walker 2007).
- **ACCURACY:** One of the strongest arguments for machine learning is that it is more accurate than conventional predictive models (Babuta et al. 2018; Berk and Bleich 2013; Carton et al. 2016). Accuracy is critical for the field of policing, which often deals with life-and-death situations. Researchers argue that with more complex data, such as criminological data, machine-learning algorithms have greater forecasting accuracy than more conventional models (Berk and Bleich 2013).



Machine Learning in Policing

EXAMPLES OF MACHINE LEARNING USED TO IMPROVE POLICING PRACTICE

In addition to its integration in practice, machine learning has also been used in research studies to address questions in law enforcement. These studies help to inform operations and strategies. Below, we highlight two studies that use machine learning to improve existing law enforcement operations.

EXAMPLE 1: [Developing a Practical Forecasting Screener for Domestic Violence Incidents](#)

Berk and colleagues (2005) used data mining techniques, a form of machine learning, to develop a screening tool that officers in the Los Angeles County Sheriff's Department could use to forecast future domestic violence incidents at particular households in Los Angeles County. Deputies dispatched to households for incidents that likely involved domestic violence conducted an initial 30-question screener. Deputies asked these questions to victims and others present at the scene of the incident. Over a three-month follow-up period, deputies recorded all new dispatches to the initially selected households. The researchers hypothesized that certain questions from the original screener would be better predictors of the households that had additional domestic violence calls for service; deputies could use these relevant questions as a shorter screener. The department collected complete data for 516 households; of those, 109 households (21 percent) had another call for service within the three-month follow-up period. The study found that the short screener instrument correctly forecasted future calls for service about 60 percent of the time and future calls involving domestic violence misdemeanors and felonies about 50 percent of the time. Ultimately, the authors of the study concluded "that for households prone to domestic violence it is possible to develop quick-response threat assessment instruments that can be used successfully in the field by law enforcement personnel." However, they cautioned that the predictors used to create the screener will likely vary based on the site, and that local police departments should develop their own screeners rather than use one from another jurisdiction.

Machine Learning in Policing

EXAMPLE 2: Identifying Police Officers at Risk of Adverse Events

Carton and colleagues (2016) used a machine learning technique known as a random forest model² to improve the identification of police officers at risk of adverse events (e.g., citizen complaint, use of force). The study team worked with the Charlotte-Mecklenburg Police Department to develop a machine-learning algorithm to identify officers that were at risk for an adverse event. Carton and colleagues found that their random forest model significantly outperformed the existing Early Intervention System. The machine learning method flagged 12 percent more high-risk officers (true positives) and flagged 32 percent fewer low-risk officers (false positives) (Carton et al. 2016). These improved results are significant because they will allow the department to more accurately target training and other interventions to the officers at highest risk of an adverse event. In addition, the improved accuracy will help reduce wasted time and unnecessary administrative work on low-risk officers (Carton et al. 2016).

As a supplement to this study, Carton and colleagues looked at dispatch-level data to identify factors that were most predictive of an adverse event between an officer and a citizen. They found that factors such as travel time to the event and officer-initiated dispatches were more predictive of adverse outcomes for officers. This finding will allow the department to implement more effective early interventions that address these factors, and reduce the chances of future adverse events (Carton et al. 2016). Carton and colleagues suggest that dispatch-level models provide the opportunity for “predictive risk-based dispatch decisions,” in which officers who are identified as being at higher risk of an adverse event can be held back from certain calls. This has important implications for both citizen and officer safety, in that the ability to hold back a high-risk officer can decrease the chances of a potentially hostile or violent encounter. Notably, holding back a high-risk officer should not be the only intervention. Law enforcement agencies should provide the necessary training and other support (e.g., dispatch back-up) to high-risk officers to ensure that the underlying issues increasing risk of an adverse event are properly addressed.

² Random forest is a widely used machine learning technique that aggregates predictions from a large number of decision trees to gain better performance. A decision tree works like a flow chart, where each observation is classified according to whether it meets the criteria at each branch. The random forest algorithm aggregates a large number of small decision trees and averages the results of the small decision trees to make predictions.

Summary of Studies

Below is a summary of the examples of studies discussed in this brief. The studies used either place-based and person-based predictive analytics approaches or machine-learning techniques.

Predictive analytics approach	Study title	Locale	Results
Risk terrain model	Reducing Burglaries Through Predictive Policing & Community Engagement	Indio, CA	Model showed strong relationship between historical truancy rates and current burglary rates.
Risk terrain model & geographic information system (GIS) model	Risk Terrain Modeling: Brokering Criminological Theory and GIS Methods for Crime Forecasting	Irvington, NJ	Period 1 RTM suggested that for every unit of increased risk, likelihood of shooting significantly increased by at least 69%. For Period 2, figure was 56%.
Targeted offender list	Predictions Put into Practice: A Quasi-Experimental Evaluation of Chicago's Predictive Policing Pilot	Chicago, IL	Individuals on the list were not more or less likely to become a victim of a homicide or a shooting than the comparison group. Those on the list were more likely to be arrested for a shooting.
Social harm score (threat score) & focused deterrence	The RASOR'S Edge: Focused Deterrence in Cambridge, Everett, and Somerville	Cambridge, Everett, and Somerville, MA	The social harm score was effective in identifying high-risk individuals, though the evaluation of the focused deterrence effort returned mixed results.
Data mining techniques (machine learning)	Developing a Practical Forecasting Screener for Domestic Violence Incidents	Los Angeles County, CA	The short screener instrument correctly forecasted future calls for service about 60% of the time and future calls involving domestic violence misdemeanors and felonies about 50% of the time.
Random forest (machine learning)	Identifying Police Officers at Risk of Adverse Events	Charlotte, NC	The machine learning method flagged 12% more high-risk officers, representing true positives, and flagged 32% fewer low-risk officers, representing false positives.

Considerations and Limitations of Predictive Analytics

Overall, findings about the effectiveness of predictive analytics in reducing crime are mixed (Ferguson 2019). The results often depend on several factors, such as the particular analytics used, the context in which they are used, and how well they are implemented. Several law enforcement agencies have found the use of predictive analytics and associated software to be effective at reducing crime (Bond et al. 2014; Turner et al. 2014). For example, in 2013, the Atlanta Police Department conducted an initial 90-day deployment of a machine-learning software program in two urban policing zones. Results from the trial period showed that the two policing zones that used the program saw crime reductions of 8 and 9 percent, while policing zones that did not experienced crime increases between 1 and 8 percent (Turner et al. 2014). Conversely, some studies have found the use of predictive analytics to be ineffective at reducing crime (Hunt et al. 2014). A 2014 RAND study found no differences in property crime for districts using predictive analytics versus those that did not (Hunt et al. 2014).

There are several limitations to be aware of before deciding to use predictive analytics as a tool to inform policing operations, including the potential biased nature of predictive analytics and the data it relies on.

There are several limitations to be aware of before deciding to use predictive analytics as a tool to inform policing operations, including the potential biased nature of predictive analytics and the data it relies on.

- **BIAS:** One of the primary criticisms of predictive analytics is that it perpetuates historical biases and is built using biased data (Babuta et al. 2018; Brayne 2017; Cino 2017; Ensign et al. 2017; Fitzpatrick et al. 2019; Joh 2017; Perry et al. 2013; Selbst 2017; Shapiro 2019). Within the field of policing, the vast majority of crime data are collected by the police themselves, based only on the crimes that are reported (Joh 2017; Lum and Isaac 2016; Selbst 2017). If police allocate more resources to certain neighborhoods, then crime data from those neighborhoods will be overrepresented in subsequent predictive models (Cino 2017; Ensign et al. 2017; Joh 2017; Selbst 2017). This can be referred to as “algorithmic discrimination” (Selbst 2017). Ultimately, any existing biases in resource allocation and police enforcement will inherently be reflected in any analysis based on those data. If police have allocated more resources in one neighborhood and thus have made more arrests in that neighborhood, then analyses may unfairly predict that crime

Considerations and Limitations of Predictive Analytics

is more likely to occur in that neighborhood. The same can be said for person-based approaches, such as targeted offender lists, which may rely on past criminal history data. If that past criminal history data contains bias (e.g., African-Americans are more likely to be stopped and detained), then those biases will persist in subsequent predictions.

- **BLACK BOX:** Machine-learning algorithms present another significant problem in policing. Many researchers and critics consider machine-learning algorithms to be “black boxes” because they cannot be fully understood by those who use them to make decisions and those who the decisions impact (Babuta et al. 2018; Brennan and Oliver 2013; Cino 2017; Joh 2017; Perry et al. 2013). More specifically, it can be nearly impossible to identify the particular factors that led a machine-learning model to a final decision. This lack of transparency obscures how decisions are made and makes it difficult to challenge a decision (Babuta et al. 2018; Brennan and Oliver, 2013; Joh 2017). It also hinders targeted interventions, since the impacts of specific individual factors used in the model are unknown. Consequently, agencies are often unable to justify their choices of variables. The black box nature of machine-learning algorithms has led to concerns about due process and the civil rights of defendants and other vulnerable populations (Babuta et al. 2018; Kutnowski 2017; Perry et al. 2013; Selbst 2017). Addressing the lack of transparency, Babuta and colleagues (2018) recommend that machine-learning models should not be the sole source of decision-making; instead, agencies should use them in conjunction with officer discretion. This also helps ensure that law enforcement officers and agencies are held accountable for the decisions they make. Machine-learning models and predictive analytics are tools in the law enforcement toolbox. They do not drive operations on their own. Rather, predictive analytics and other relevant factors inform operations and decision-making.
- **PRIVATIZED SOFTWARE:** Because law enforcement agencies are simply the clients, they are not fully aware of how the predictive analytics is used or how certain algorithms make decisions (Cino 2017; Joh 2017). When working with private companies, agencies have to sign contracts that may include nondisclosure agreements (Garrett 2018). As a result, agencies are not able to share crucial information with the public, which perpetuates the idea of the “black box.”

Considerations and Limitations of Predictive Analytics

- **FEASIBILITY:** Feasibility in this context refers to law enforcement agencies' willingness, ability, and capacity to use predictive analytics properly or to adopt predictive analytics software. Being understaffed, lacking crime analysts, or having limited time for in-service training can further complicate the adoption of new analytical methods or software. Additional obstacles include the actual utilization of the new methods, questions about the legitimacy of those methods, and the fact that few agencies are designed to support advanced crime analysis (Santos and Taylor 2014).

Practical Considerations for Implementation of Predictive Analytics

Agencies that want to utilize predictive analytics techniques must consider certain factors about the analytic tools, resources, and local context to assess their feasibility. Here, we describe key factors for consideration.

GOAL OF ANALYTICS

Agencies should consider the crime problem or intelligence gap they will use predictive analytics to address. Certain predictive analytics techniques are more appropriate for particular problems. Table 1 outlines the primary uses of the predictive analytics techniques described in this brief.

TABLE 1. PRIMARY USES OF PREDICTIVE ANALYTICS TECHNIQUES

Predictive analytics technique	Primary uses
Regression analysis	Identifying areas at increased risk of crime and identifying variables that have a positive association with increased risk of crime
Hot spot detection	Identifying areas with increased concentration of crime
Risk terrain modeling (RTM)	Identifying areas at increased risk of crime based on environmental factors that increase vulnerability for specific locations
Targeted offender lists	Identifying individuals at increased risk of becoming offenders or victims
Machine learning algorithms	Identifying areas at increased risk of crime

Sources: (Joh 2017; Kennedy et al. 2011; Perry et al. 2013)

DATA AVAILABILITY

Different predictive analytics techniques require different data sources to support their implementation. An agency will need to assess whether it collects the necessary data in proper formats to use predictive analytics techniques. If the data are not available, the agency may consider developing more robust data collection and records management systems to use predictive techniques in the future. Table 2 presents commonly required types and sources of data associated with each technique described in this brief.

Practical Considerations for Implementation of Predictive Analytics

TABLE 2. GENERAL DATA REQUIREMENTS FOR PREDICTIVE ANALYTICS TECHNIQUES

Predictive analytics technique	General data requirements
Regression analysis	Historical crime data (e.g., type, location, time), data for independent variables identified as being relevant (e.g., demographic data)
Hot spot detection	Historical crime data (e.g., type, location, time)
Risk terrain modeling (RTM)	Historical crime data (e.g., type, location, time), environmental risk factors data (e.g., liquor stores, gas stations), calls for service data (as needed)
Targeted offender lists	Arrest data, gang affiliations data, criminal activity trends, victimization data
Machine learning algorithms	Historical crime data (e.g., type, location, time), incident report data; CAD data; other data as desired (e.g., weather pattern data, school calendar data)

Sources: (Joh 2017; Kennedy et al. 2016; Perry et al. 2013; <https://data.cityofchicago.org/Public-Safety/Strategic-Subject-List/4aki-r3np>)

CAPABILITIES AND RESOURCES

The predictive analytics techniques described in this brief all require technical skills, such as statistical analysis skills, statistical programming language skills, and data management skills. These capabilities may be fulfilled by agency personnel (e.g., crime analyst) or by outside providers such as research partners or vendors. Different techniques also require increasingly sophisticated skills and advanced training. Table 3 illustrates some of the technical skills needed to implement the predictive analytics techniques described in this brief.

Practical Considerations for Implementation of Predictive Analytics

TABLE 3. TECHNICAL SKILLSETS NEEDED FOR PREDICTIVE ANALYTICS TECHNIQUES

Predictive analytics techniques	Technical skillsets
Hot spot detection	<ul style="list-style-type: none"> • Use of common statistical analytical tools (e.g., Excel, GIS) • Programming skills • Ability to synthesize large amounts of data from various sources ranging from police reports to mapping and GIS data to recognize trends and patterns
Regression analysis	<ul style="list-style-type: none"> • Use of statistical software packages and languages • Understanding of statistical concepts (e.g., probability, variance) and modeling • Ability to link multiple datasets
Targeted offender lists	<ul style="list-style-type: none"> • Ability to link multiple datasets • Standardized approach to obtain intelligence from street officers, detectives, crime analysts, and other agencies • Understanding of crime patterns and institutionalization of gangs • Ability to synthesize large amounts of data from various sources (e.g., incident reports, CAD data, case files) to recognize trends and patterns
Risk terrain modeling (RTM)	<ul style="list-style-type: none"> • Understanding of statistical concepts (e.g., probability) • Understanding of geospatial modeling, mapping and coding • Use of GIS software
Machine learning algorithms	<ul style="list-style-type: none"> • Understanding of statistical concepts (e.g., probability, analysis of variance, hypothesis testing) and statistical modeling • Ability to link multiple datasets • Programming skills

Sources: (Perry et al. 2013; Bureau of Justice Assistance, 2020)

Summary



In the last decade, law enforcement agencies across the country have increasingly adopted predictive analytics to reduce and prevent crime. This includes conventional and more advanced predictive analytics for place-based (e.g., hot spot detection) and person-based predictions (e.g., targeted offender lists). The use of predictive analytics has empowered agencies to take more proactive approaches to crime reduction, rather than conventional reactive approaches, such as rapid responses to 911 calls. However, not all agencies have found the use of predictive analytics to be more effective than their current operations.

As the field of policing has evolved, so have the capabilities and complexities of predictive analytics. Machine learning is emerging within the field as a superior option for forecasting crime. Researchers argue that machine learning is better suited for the complex nature of criminological data and that machine learning's ability to incorporate costs (i.e., consequences) improves the accuracy of prediction. However, the decision to allocate resources (e.g., staff, money) toward predictive analytics strategies and software should be well informed. The proper use of predictive analytics to reduce and prevent crime requires careful planning and training.

References

- Babuta, A., Oswald, M., & Rinik, C. (2018). Machine Learning Algorithms and Police Decision-Making: Legal, Ethical and Regulatory Challenges. Royal United Services Institute for Defence and Security Studies.
- Berk, R. A., & Bleich, J. (2013). "Statistical procedures for forecasting criminal behavior: A comparative assessment." *Criminology & Pub. Pol'y*, 12, 513.
- Berk, R. A., He, Y., & Sorenson, S. B. (2005). "Developing a practical forecasting screener for domestic violence incidents." *Evaluation Review*, 29(4), 358–383.
- Bond, B. J., Hajjar, L., Ryan, A., & White, M. D. (2014). Lowell, Massachusetts, Smart Policing Initiative: Reducing property crime in targeted hot spots. Washington, DC: Bureau of Justice Assistance.
- Braga, A. A., & Bond, B. J. (2008). "Policing crime and disorder hot spots: A randomized controlled trial." *Criminology*, 46(3), 577–607.
- Braga, A. A., Papachristos, A., & Hureau, D. (2012). Hot spots policing effects on crime. *Campbell Systematic Reviews*, 8(1), 1–96.
- Braga, A. A., & Weisburd, D. L. (2006). "Police innovation and crime prevention: Lessons learned from police research over the past 20 years." *This review draws upon material available in David L. Weisburd and Anthony A. Braga.(Eds.)*. Washington, DC: National Institute of Justice.
- Braga, A. A., & Weisburd, D. (2010). *Policing problem places: Crime hot spots and effective prevention*. Oxford University Press on Demand.
- Brayne, S. (2017). "Big data surveillance: The case of policing." *American sociological review*, 82(5), 977–1008.
- Brayne, S., Rosenblat, A., & Boyd, D. (2015). Predictive policing. *Data & Civil Rights: A New Era Of Policing And Justice*.
- Brennan, T., & Oliver, W. L. (2013). "Emergence of machine learning techniques in criminology: implications of complexity in our data and in research questions." *Criminology & Pub. Pol'y*, 12, 551.
- Bureau of Justice Assistance. 2020. "Project Safe Neighborhoods Webinar on Targeting Offenders and Creating Targeted Offender Lists." Accessed September 4, 2020. <https://psntta.org/tta-resources/webinars/psn-webinar-on-targeting-offenders-and-creating-targeted-offender-lists/>.
- Caplan, J. M., Kennedy, L. W., & Miller, J. (2011). "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting." *Justice Quarterly*, 28(2), 360–381.
- Carton, S., Helsby, J., Joseph, K., Mahmud, A., Park, Y., Walsh, J., & Ghani, R. (2016). "Identifying police officers at risk of adverse events." In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 67–76).
- Cino, J. G. (2017). "Deploying the Secret Police: The Use of Algorithms in the Criminal Justice System." *Ga. St. UL Rev.*, 34, 1073.
- Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2017). "Runaway feedback loops in predictive policing." *arXiv preprint arXiv:1706.09847*.

References

- Ferguson, A. G. (2019). Predictive Policing Theory. Chapter 24: *The Cambridge Handbook of Policing in the United States* (ed. Tamara Rice Lave & Eric J. Miller), Cambridge Univ. Press.
- Ferguson, A. G. (2017). "The Police Are Using Computer Algorithms to Tell If You're a Threat." *Time*. <https://time.com/4966125/police-departments-algorithms-chicago/>
- Fitzpatrick, D. J., Gorr, W. L., & Neill, D. B. (2019). "Keeping Score: Predictive Analytics in Policing." *Annual Review of Criminology*.
- Garrett, B. L. (2018). "Evidence-Informed Criminal Justice." *Geo. Wash. L. Rev.*, 86, 1490.
- Garrison, A. H. (2RAND009). "The influence of research on criminal justice policy making." *Professional Issues in Criminal Justice*, 4(1), 9-21.
- Gorner, J., & Sweeney, A. (2020). "For years Chicago police rated the risk of tens of thousands being caught up in violence. That controversial effort has quietly been ended." *Chicago Tribune*. <http://www.chicagotribune.com/news/criminal-justice/ct-chicago-police-strategic-subject-list-ended-20200125-spn4kjmrxrh4tmktdjckhtox4i-story.html>
- Hollywood, J. S. (2016). "CPD's 'heat list' and the dilemma of predictive policing." *Crain's Chicago Business*. <https://www.chicagobusiness.com/article/20160919/OPINION/160919856/chicago-police-s-heat-list-and-what-to-do-with-predictive-policing>
- Hollywood, J. S., McKay, K. N., Woods, D., & Agniel, D. (2019). *Real-Time Crime Centers in Chicago: Evaluation of the Chicago Police Department's Strategic Decision Support Centers*. RAND Corporation.
- Hunt, P., Saunders, J., & Hollywood, J. S. (2014). *Evaluation of the Shreveport predictive policing experiment*. RAND Corporation.
- Joh, E. E. (2017). "Feeding the machine: Policing, crime data, & algorithms." *Wm. & Mary Bill Rts. J.*, 26, 287.
- Kennedy, L. W., Caplan, J. M., & Piza, E. (2011). Risk clusters, hotspots, and spatial intelligence: risk terrain modeling as an algorithm for police resource allocation strategies. *Journal of quantitative criminology*, 27(3), 339–362.
- Kennedy, L. W., Caplan, J. M., Piza, E. L., & Buccine-Schraeder, H. (2016). Vulnerability and exposure to crime: Applying risk terrain modeling to the study of assault in Chicago. *Applied Spatial Analysis and Policy*, 9(4), 529–548.
- Kutnowski, M. (2017). "The ethical dangers and merits of predictive policing." *Journal of community safety and well-being*, 2(1), 13–17.
- Lum, K., & Isaac, W. (2016). "To predict and serve?" *Significance*, 13(5), 14–19.
- Moses, L., & Chan, J. (2018). "Algorithmic prediction in policing: assumptions, evaluation, and accountability." *Policing and society*, 28(7), 806–822.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- Parker, R. N., & Martinez E. (2014). *Indio, California Smart Policing Initiative: Reducing Burglaries through Predictive Policing and Community Engagement*. Washington, DC: Bureau of Justice Assistance.

References

- 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*. RAND Corporation.
- Puente, M. (2019). "LAPD ends another data-driven crime program touted to target violent offenders." *Los Angeles Times*. <https://www.latimes.com/local/lanow/la-me-laser-lapd-crime-data-program-20190412-story.html>
- Reppetto, T. A. (1976). Crime prevention and the displacement phenomenon. *Crime & Delinquency*, 22(2), 166–177.
- Santos, R. B., & Taylor, B. (2014). "The integration of crime analysis into police patrol work." *Policing: an international journal of police strategies & management*.
- Saunders, J., Hunt, P., & Hollywood, J. S. (2016). "Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot." *Journal of Experimental Criminology*, 12(3), 347–371.
- Selbst, A. D. (2017). "Disparate impact in big data policing." *Ga. L. Rev.*, 52, 109.
- Shapiro, A. (2019). "Predictive Policing for Reform? Indeterminacy and Intervention in Big Data Policing." *Surveillance & Society*, 17(3/4), 456–472.
- Stroud, M. (2016). "Chicago's predictive policing tool just failed a major test." *The Verge*. <https://www.theverge.com/2016/8/19/12552384/chicago-heat-list-tool-failed-rand-test>.
- Turner, G., Brantingham, D. J., & Mohler, D. G. (2014). "Predictive Policing in Action in Atlanta, Georgia." *The Police Chief*, 81, 72–74.
- Uchida, C. D., Swatt, M., Davis, J. S., Connor, C., Shutinya, M., Phillips, W., & Wagner, D. (2016). *The RASOR'S Edge: Focused Deterrence in Cambridge, Everett, and Somerville*. Justice & Security Strategies, Inc.
- Walker, J. T. (2007). "Advancing science and research in criminal justice/criminology: Complex systems theory and non-linear analyses." *Justice Quarterly*, 24(4), 555–581.
- Weisburd, D., Wyckoff, L. A., Ready, J., Eck, J. E., Hinkle, J. C., & Gajewski, F. (2006). Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits. *Criminology*, 44(3), 549–592.

About CNA

CNA is a nonprofit research and analysis organization dedicated to the safety and security of the nation. It operates the Institute for Public Research—which serves civilian government agencies—and the Center for Naval Analyses, the Department of the Navy’s federally funded research and development center (FFRDC). CNA is dedicated to developing actionable solutions to complex problems of national importance. With nearly 700 scientists, analysts and professional staff, CNA takes a real-world approach to gathering data, working side-by-side with operators and decision makers around the world. CNA’s research portfolio includes global security and great power competition, homeland security, emergency management, criminal justice, public health, data management, systems analysis, naval operations and fleet and operational readiness.



About BJA

BJA helps to make American communities safer by strengthening the nation's criminal justice system: Its grants, training and technical assistance, and policy development services provide state, local, and tribal governments with the cutting edge tools and best practices they need to reduce violent and drug-related crime, support law enforcement, and combat victimization.

BJA is a component of the *Office of Justice Programs, U.S. Department of Justice*, which also includes the *Bureau of Justice Statistics, National Institute of Justice, Office of Juvenile Justice and Delinquency Prevention, Office for Victims of Crime*, and *Office of Sex Offender Sentencing, Monitoring, Apprehending, Registering, and Tracking*.

BJA Mission

BJA provides leadership and services in grant administration and criminal justice policy development to support local, state, and tribal law enforcement in achieving safer communities. BJA supports programs and initiatives in the areas of law enforcement, justice information sharing, countering terrorism, managing offenders, combating drug crime and abuse, adjudication, advancing tribal justice, crime prevention, protecting vulnerable populations, and capacity building. Driving BJA's work in the field are the following principles:

- Emphasize local control.
- Build relationships in the field.
- Provide training and technical assistance in support of efforts to prevent crime, drug abuse, and violence at the national, state, and local levels.
- Develop collaborations and partnerships.
- Promote capacity building through planning.
- Streamline the administration of grants.
- Increase training and technical assistance.
- Create accountability of projects.
- Encourage innovation.
- Communicate the value of justice efforts to decision makers at every level.

To learn more about BJA, follow us on [Facebook](#) and [Twitter](#).



CNA

BJA

Bureau of Justice Assistance
U.S. Department of Justice