

U.S. Department of Justice

PREDICTIVE ANALYTICS IN LAW ENFORCEMENT: A REPORT BY THE DEPARTMENT OF JUSTICE



November 2014

PREDICTIVE ANALYTICS IN LAW ENFORCEMENT

Federal, State, and Local law enforcement agencies use data analytics in a variety of ways, including to investigate past criminal acts, identify patterns in criminal behavior, allocate scarce enforcement resources, and evaluate program and operational efficacy. Many of these tools are simply faster or automated versions of basic data analysis techniques that have been used for many years. Recently, however, new technologies have enabled law enforcement agencies to harness larger and more intricate data sets. Among these new techniques are sophisticated tools that allow law enforcement agencies to model complex variables and create forward-looking assessments. In this report, we focus on the use of those new, more complex, and forward-looking technologies in domestic law enforcement.¹

Although the term “big data” is often used colloquially to describe a wide range of predictive analytics programs, many of the most common applications do not involve an analysis of the volume and variety of information that are the hallmarks of big data.² While these tools present opportunities and potential challenges that are worthy of consideration, they represent an evolution, rather than a revolution. The vast majority of the tools are less sophisticated than the data models regularly deployed by commercial entities in advertising and marketing, and nothing in use today approaches the depictions in popular culture of technologies that can predict with certainty the future actions of individuals or specific criminal activity.

In the following sections, we provide an overview of the current use of predictive analytics in law enforcement. In Section I, we describe the two major categories of predictive analytics: those that aim to identify places that should receive law enforcement’s attention and those that aim to identify individuals who should receive law enforcement’s attention. In Section II, we provide examples of the various efforts—some of which the Department of Justice (DOJ) has supported through grants and research—of State and Local law enforcement agencies around the country. In Section III, we describe how federal law enforcement agencies within DOJ and the Department of Homeland Security have employed predictive analytics in their own enforcement efforts and to allocate resources. In Section IV, we review the use of predictive methodologies at the Federal and State levels in post-arrest proceedings. Finally, in Sections V and VI we discuss the potential benefits and concerns associated with law enforcement use of big data and predictive analytics, as well as tentative next steps to help ensure that these tools are used effectively and appropriately.

I. PREDICTIVE ANALYTICS IN LAW ENFORCEMENT – METHODS

Current applications of predictive analytics in law enforcement fall largely into two broad categories—(1) predicting places and times that criminal activity is expected to occur (“place-

¹ The report does not consider military or intelligence applications of predictive analytics that are focused abroad.

² See generally EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (May 2014), at 4 (defining “big data” in terms of the “three Vs”—volume, variety, and velocity).

based” methods), and (2) predicting which individuals are likely to perpetrate, or fall victim to, crime (“person-based” methods).³ In this section, we describe both of these categories.

A. Place-Based Methods

Law enforcement agencies have long attempted to identify patterns in criminal activity in order to allocate scarce resources more efficiently. Many agencies have used mapping to focus their efforts on high-crime areas, evoking the familiar image of a city map hanging on the wall of a local precinct, with thumbtacks marking the locations of recent crimes. New technologies are replacing manual techniques, and many police departments now use more sophisticated computer modeling systems to refine their understanding of hot spots, linking offense data to patterns in temperature, time of day, proximity to other structures and facilities, and other variables.⁴

Some of the newest analytical modeling techniques promise to predict with greater precision locations and times at which criminal activity is likely to occur. A seasoned beat officer might tell you that a neighborhood that has recently been victimized by one or more burglaries is likely to be targeted for additional property crimes in the coming days; an analytical method known as “near-repeat modeling” provides evidence-based support for this common understanding by using recent crime data to identify which particular adjacent geographic areas have the greatest risk of additional burglaries in the near future.⁵ Similarly, experienced officers might provide anecdotal evidence that criminal activity often clusters around particular types of facilities such as bars and motels; a technique known as “risk terrain modeling” provides statistical support for the view that crime is often a function of the interaction of specific social and physical factors that attract would-be offenders and create conditions ripe for criminal activity.⁶ Indeed, risk terrain modeling promises to help identify geographic areas at risk for future crime even in the absence of any significant history of criminal activity.

B. Person-Based Methods

Several researchers and law enforcement agencies have embraced the use of data analysis and more sophisticated computing methods to estimate the level of risk associated with individuals. As with the place-based techniques described above, these techniques largely

³ See generally Walter L. Perry, Brian McInnis, Carter C. Price, Susan C. Smith and John S. Hollywood, *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*, Washington, D.C.: RAND Corporation, RR-233-NIJ, 2013 (hereinafter, “RAND Predictive Policing Report”).

⁴ See John E. Eck, Spencer Chainey, James G. Cameron, Michael Leitner, and Ronald E. Wilson, *Mapping Crime: Understanding Hot Spots*, Washington, D.C.: National Institute of Justice, August 2005.

⁵ See Shane D. Johnson, Kate J. Bowers, Dan J. Birks, and Ken Pease, “Predictive Mapping of Crime by ProMap: Accuracy, Units of Analysis, and the Environmental Backcloth,” in David Weisburd, Wim Bernasco, and Gerben J.N. Bruinisma, eds., *Putting Crime in Its Place: Units of Analysis in Geographic Criminology*, New York: Springer, 2009.

⁶ See generally Leslie W. Kennedy, Joel M. Caplan and Eric Piza, *Risk Clusters, Hotspots, and Spatial Intelligence: Risk Terrain Modeling as an Algorithm for Police Resource Allocation Strategies*, 27 J. QUANT. CRIMINOL. 339-362 (2011).

represent an evolution of the assessments that have been part of probation, parole, pretrial detention, and sentencing decisions for many years.

In addition, the use of “social network analysis” is gaining the attention of some State and Local law enforcement agencies.⁷ Based on the theory that crime clusters together in social networks just as it does in space and time, various police departments have begun applying social network analysis to target individuals who are known to have close interpersonal associations with others who have recently perpetrated or been victimized by violent crime. These individuals may receive increased scrutiny, additional support services, or both. Some local law enforcement agencies have paired social network analysis with collaborative community policing methods in efforts to build relationships between the police and at-risk individuals.

II. APPLICATIONS BY STATE AND LOCAL LAW ENFORCEMENT IN INVESTIGATIONS AND RESOURCE ALLOCATION

State and Local law enforcement agencies use an array of predictive tools in allocating resources and conducting investigations. This section discusses those tools, with the discussion organized primarily by whether the tools are place- or person-based, and noting whether or not the projects were funded by DOJ.

A. *Place-Based Analytics: Predicting Crime in Place and Time*

1. DOJ-Supported Research and Development

The Department of Justice has encouraged and funded work related to the development and use of some new predictive and forecasting methodologies. In 2008, former Los Angeles Police Chief William Bratton enlisted the Department’s National Institute of Justice (NIJ) to bring together stakeholders on issues related to predictive policing techniques. NIJ has provided funding to think tanks and universities to study predictive policing and to evaluate pilot projects taking place in states and communities nationwide. NIJ has held two symposia on the use of predictive policing in law enforcement, bringing together researchers, practitioners and law enforcement leaders from around the country. Overall, NIJ has funded more than a dozen law enforcement agencies, researchers, and other entities to develop and implement advanced place-based techniques.

For example, between 2009 and 2012, NIJ awarded approximately \$620,000 to the Shreveport, Louisiana Police Department (SPD) to conduct a pilot program to apply predictive policing strategies to decrease property crime.⁸ During the first phase of the pilot, researchers developed and tested several analytical models using variables traditionally correlated with property crimes, including the presence of residents on probation or parole, data on recent property crime in the area, recent juvenile arrests, and 911 calls to the area. Researchers used these analytical models to identify at-risk areas for intervention and developed intervention

⁷ See generally Andrew V. Papachristos, *The Coming of a Networked Criminology? Using Social Network Analysis in the Study of Crime and Deviance*, *Advances in Criminal Theory*, Vol. 13 (2011), available at http://www.papachristos.org/Publications_2_files/Coming_of_networked_crim.pdf.

⁸ See generally Priscillia Hunt, Jessica Saunders, and John S. Hollywood, *Evaluation of the Shreveport Predictive Policing Experiment*, Santa Monica, Cal.: RAND Corporation, RR-531-NIJ, 2014.

models. In the second phase of the pilot, researchers tested the chosen analytical and intervention models on particular districts and compared the results to control districts.

The study is now complete. The RAND Corporation evaluated the program and found that property crime decreased by approximately 35% in the first four months of the seven-month evaluation period as compared with the control districts. After those first four months, however, the SPD reduced its intervention efforts, and property crime reverted back to the same level as the control districts. The RAND evaluation concluded that additional research should be done.

NIJ has also funded research about near repeat methodologies. In 2006 researchers at Temple University used grant funding from NIJ to develop a “near-repeat calculator” to help law enforcement agencies more accurately pinpoint where and when near-repeat effects were most likely to manifest.⁹ In addition, NIJ recently provided approximately \$400,000 to fund a random, controlled study testing near-repeat technologies in Redlands, California and Baltimore, Maryland. Results from this study are expected in late 2015 or early 2016.

NIJ is also funding a series of projects to test the efficacy of Risk Terrain Modeling (RTM). Variables are identified through empirical analysis, literature review, or even professional experience; the factors are subsequently assigned a value and overlain on maps in layers, yielding a map that takes into account all of the various risk factors. NIJ has provided approximately \$500,000 to Rutgers University to study the efficacy of RTM in five cities—Chicago; Colorado Springs; Newark; Kansas City, Missouri; and Glendale, Arizona. This study is expected to conclude in late 2015 or early 2016. In addition, NIJ is providing approximately \$380,000 to the New York City Police Department to study the development of RTM models, including a focus on how risk factors used in such models should be chosen. An evaluation of this study is expected in 2017.

Other DOJ components also have encouraged and supported the use of place-based predictive analytics in policing. The Bureau of Justice Assistance (BJA) provides resources to law enforcement agencies across the United States through its National Training and Technical Assistance Center (NTTAC). Among its highest priority programs is Crime Analysis on Demand, through which BJA offers law enforcement agencies resources to incorporate data analysis into strategic decisions about preventing and responding to crime. Services offered by NTTAC as part of Crime Analysis on Demand include needs assessment; support in addressing analytical gaps; and comprehensive training for crime analysts, including analytical skills, and product development. NTTAC also offers training on predictive policing software.

In 2009, BJA established the Smart Policing Initiative (“SPI”) to support through technical and financial assistance the work of criminal justice agencies in developing data-driven, evidence-based tactics to address their most serious law enforcement challenges. The police department in Columbia, South Carolina, for example, has used SPI funds to develop a predictive project spotlighting repeat and near-repeat burglaries in the North Region of the city, a focus Columbia selected because an initial hot spot analysis revealed high rates of burglaries in the area. Columbia developed a two-tiered strategy in response to the North Region’s burglary

⁹ See National Institute of Justice, *Translating “Near Repeat” Theory into a Geospatial Police Strategy* (June 9, 2014), available at <http://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/near-repeat-theory.aspx>.

problem: (1) visiting residents of homes that had been burgled and conducting near-repeat burglary notifications for homes within a one-block radius, and (2) conducting follow-up security surveys after a subsequent burglary, offering temporary security systems in some cases.

The Department's Office of Community Oriented Policing Services (COPS), also has supported the use of geospatial analysis techniques in community policing initiatives. COPS recently provided approximately \$250,000 to researchers at George Mason University's Center for Evidence-Based Crime Policy to apply hot spot analysis to identify spatial patterns of juvenile crime in Seattle. Through this initiative, the Seattle Police Department implemented community policing intervention in three hot spots identified by a predictive algorithm based on incident reports associated with juvenile arrests, and compared results against three matched hot spots that were identified through different means. This project is now complete, and the researchers expect to release an evaluation in the next three to four months. In addition, COPS funded the development of a "how to" guide for law enforcement agencies to conduct hot spot analysis specific to juvenile crime.

All States, Localities, researchers, and any other entities receiving grant funding from DOJ under the programs described above are required to comply with federal civil rights laws. In particular, Title VI, 42 U.S.C. § 2000d *et seq.*, enacted as part of the Civil Rights Act of 1964, prohibits discrimination on the basis of race, color, and national origin in programs and activities receiving federal financial assistance. In addition, 42 U.S.C. § 3789d, part of the Omnibus Crime Control and Safe Streets Act of 1968, specifically prohibits discrimination on the basis of race, color, religion, national origin, or sex, in programs or activities funded by the Office of Justice Programs, the umbrella entity which includes NIJ and BJA, and in programs funded by COPS' Public Safety Partnership and Community Policing Program.¹⁰

In addition to these legal requirements, BJA provides a suite of resources to State, Local, and Tribal law enforcement and justice agencies to use in their efforts to implement appropriate policies and protections for handling information. COPS also reviews specific state, local, and tribal law enforcement programs to determine whether additional privacy guidelines are recommended. These efforts support and foster information-handling and privacy best practices in the development and use of some new predictive and forecasting methodologies within the law enforcement communities that receive Department grants.

¹⁰ NIJ provided the funding discussed in this paper through grants, with the exceptions of the projects in Shreveport, Louisiana and Chicago, which were funded through cooperative agreements. With grant funding, there is no substantial programmatic involvement between NIJ and the grant recipient; under cooperative agreements, there are greater opportunities for programmatic involvement, including technical assistance and/or consultation. In either event, NIJ typically does not take an active role in designing external projects or studies.

All NIJ grant proposals are reviewed by an independent peer review panel of researchers and practitioners, who base their reviews on criteria set forth in the relevant funding solicitation. NIJ's director reviews all panel assessments and staff reports, and the Assistant Attorney General for the Office of Justice Programs makes all final grant award decisions. Any project that involves human subjects must undergo privacy and human subject reviews.

NIJ and OJP may withdraw funding if a grant recipient or party to a cooperative agreement violates federal law or substantially strays from the parameters of the approved project.

2. Other State and Local Efforts

State and Local law enforcement agencies continue to incorporate place-based predictive analytics into their policing efforts apart from any DOJ funding, often partnering with universities and academics to develop tools and methodologies to address specific law enforcement needs.

In an early experiment in place-based predictive policing, the Richmond, Virginia Police Department (RPD) mined various databases to predict locations and times of heightened criminal activity. The RPD created a model in 2003 that incorporated date, time of day, weather, moon phases, city events, pay days and crime records to identify patterns and relationships that were not necessarily apparent to officers using traditional methods. RPD officers used this information to decide where and when to deploy additional resources.¹¹

More recently, the police department in Santa Cruz, California collaborated with researchers at Santa Clara University to implement a near-repeat algorithm in 2011 aimed at predicting and preventing certain property crimes. Developed using eight years of crime data, the computer model predicted which geographic areas and windows of time were at the highest risk for future crime. Officers were then sent to patrol these areas—approximately the size of one square city block—while they were not handling other service calls. The Santa Cruz Police Department reported a 19% drop in property thefts in the first half of 2012 compared with the same time period in 2011, a result it attributed to its use of predictive mapping.¹²

Researchers from the Rutgers School of Criminal Justice recently conducted a study to determine whether RTM could outperform traditional hot spot mapping in predicting future shootings in Newark, New Jersey.¹³ The Newark RTM, which took into account the area's history of drug arrests, whether the location was a known gang territory, and the presence of at-risk housing in the area, outperformed traditional hot spot models at predicting shootings by approximately 36%.

These are a few examples of the many experiments taking place in cities across the country. NIJ is working with DOJ's Bureau of Justice Statistics to survey State and Local law enforcement agencies on their use of predictive techniques in geospatial mapping. The survey will be disseminated to more than 900 State and Local agencies nationwide. Once the survey results are reviewed in late 2015, DOJ should have a more comprehensive picture of the use of various geospatial predictive policing techniques.

All State and Local law enforcement agencies, whether or not they receive federal funding, are required to abide by the federal Violent Crime Control and Law Enforcement Act of 1994, 42 U.S.C. § 14141, which makes it unlawful for any law enforcement agency or agent thereof to engage in a pattern or practice of conduct that deprives individuals of rights guaranteed to them under the Constitution or other federal laws. Where such a pattern or

¹¹ Chandler Harris, *Richmond, Virginia Police Department Helps Lower Crime Rates with Crime Prediction Software*, GOV'T TECH. (Dec. 21, 2008), available at <http://www.govtech.com/public-safety/Richmond-Virginia-Police-Department-Helps-Lower.html>.

¹² Rand Predictive Policing Report, *supra* note 3, at 43.

¹³ See generally Kennedy, Caplan and Piza, *supra* note 6.

practice exists, the Justice Department's Civil Rights Division may bring a civil action to obtain equitable and declaratory relief.

B. *Person-Based Predictive Analytics: Identifying Likely Perpetrators and/or Victims of Criminal Activity*

1. Social Network Analysis

Social Network Analysis (SNA) in the criminal justice context rests on the idea that crime, like infectious disease, follows certain patterns that one can discern through an analysis of social interactions. In the criminal justice context, SNA incorporates the theory that likely perpetrators and victims of homicide are highly overlapping populations, such that an SNA model can help identify individuals most likely to commit future crimes, as well as those who are most likely to be victimized.¹⁴

The Chicago Police Department (CPD), with the help of more than \$3 million from NIJ, is conducting a predictive policing pilot program that, among other things, tests the use of SNA to identify the individuals most at risk of firearms related violence, either as a shooter or a victim.¹⁵ CPD collaborated with academic researchers to develop an SNA model to identify close interpersonal links in social networks in which violence had taken place. Using criminal history records and information about firearms victims, CPD's model generated a Strategic Subjects List (SSL) of the 400 most at-risk individuals within CPD's jurisdiction. According to the model, these individuals had statistical probabilities of being party to violence that were hundreds of times greater than those of an average citizen. Each of the individuals had a prior criminal record and was already known to the CPD. CPD contacts individuals to notify them of their placement on the SSL and of the increased scrutiny that follows from that placement. CPD may, at its discretion, offer social service resources, educational programs or job-placement resources. The CPD's pilot project is expected to end in September 2015, with a final report expected by March 2016.¹⁶

In 2012, COPS funded a university sociologist to work with law enforcement agencies in several Connecticut cities to integrate SNA into their policing strategies. The funding will support local agencies in: (1) identifying patterns of violence among groups; (2) mapping the organizational structure of criminal groups through use of SNA; and (3) identifying which individuals are most at risk of becoming victims or perpetrators of violent crime. From this work, a series of training conferences will be developed to educate criminal justice practitioners on the use of SNA to address urban violent crime.

¹⁴ See, e.g., Andrew V. Papachristos and Christopher Wildeman, *Network Exposure and Homicide Victimization in an African-American Community*, AM. J. PUB. HEALTH: Vol. 104, No. 1 (January 2014), available at <http://ajph.aphapublications.org/doi/abs/10.2105/AJPH.2013.301441>.

¹⁵ The pilot program also includes a geospatial mapping component, which is currently in an early phase of development.

¹⁶ In February 2014, this program was the subject of a *Verge* magazine article that compared the CPD's use of the program to the "pre-crime" technology depicted in the movie *Minority Report*. See Matt Stroud, *The minority report: Chicago's new police computer predicts crimes, but is it racist?*, THE VERGE (Feb. 19, 2014, 9:31 AM), available at <http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist>.

BJA also has funded SNA projects, including one conducted by the Columbia, South Carolina Police Department about gang activity. The project incorporates data from several different sources, including the Columbia Police Department's network of contacts, its existing gang database, and social media. BJA also provides guidance to help law enforcement personnel understand how social media tools and resources can lawfully be used to prevent, mitigate, respond to, and investigate criminal activity while also ensuring appropriate protections.¹⁷

III. APPLICATIONS BY FEDERAL LAW ENFORCEMENT IN INVESTIGATIONS AND RESOURCE ALLOCATION

A. *Department of Justice*

1. **Federal Bureau of Investigation**

As the primary investigative agency of the federal government, the Federal Bureau of Investigation (FBI) is charged with investigating all violations of federal law that are not exclusively assigned to another federal agency and with carrying out investigations within the United States of threats to the national security. This includes investigating international terrorist threats to the United States and conducting activities to meet foreign entities' espionage and intelligence efforts directed against the United States. FBI uses various analytic tools to assist its investigations of prior criminal events and related suspects, as well as to disrupt potential criminal activity.

FBI uses geospatial mapping techniques. For instance, FBI field offices collect information about locations of interest and community demographics within their area of responsibility and use computer mapping technology to depict this information visually and show relationships among the data. Sometimes called domain awareness, this analysis helps each field office review relevant information in its area of responsibility in an integrated format and can help track crime trends.

Most FBI data analysis is targeted to particular subjects (*i.e.*, looking for information about particular individuals or suspects); however, FBI also conducts data analytics and threat-based analyses to assist in the prioritization of resources; to focus investigative and analytic efforts; to identify intelligence collection gaps; and, on a limited basis, to identify sub-sets of persons who may bear further investigative scrutiny. To conduct these analyses, FBI compares a variety of data sources, including information from FBI case files, open source or commercial databases, and information lawfully collected by FBI or other government agencies. This information includes, among other things, biographical information, biometric information, financial information, location information, associates and affiliations, employment and business information, visa and immigration information, travel information, and criminal and investigative history information.

¹⁷ See Global Justice Information Sharing Initiative, *Developing a Policy on the Use of Social Media in Intelligence and Investigative Activities: Guidance and Recommendations* (February 2013), available at <https://it.ojp.gov/gist/132/Developing-a-Policy-on-the-Use-of-Social-Media-in-Intelligence-and-Investigative-Activities--Guidance-and-Recommendations->.

FBI also uses advanced techniques to analyze Suspicious Activity Reports (SARs). The Bank Secrecy Act (BSA) requires U.S. financial institutions (including banks, money service businesses, securities firms, and casinos) to maintain records of cash purchases of negotiable instruments; file reports of cash transactions exceeding \$10,000; and report suspicious activity that might indicate money laundering, tax evasion, or other criminal activity. These records are filed with the Financial Crimes Enforcement Network (FinCEN).

FBI conducts and analyzes [REDACTED]

Like other FBI activity, data analytics are governed by FBI's policies and procedures that protect privacy and civil liberties. Under the Attorney General's Guidelines for Domestic FBI Operations (AGG-Dom) and FBI's Domestic Investigations and Operations Guide (DIOG), no investigative activity may be taken solely on the basis of activities that are protected by the First Amendment or based solely on the race, ethnicity, national origin or religion of the subject, or a combination of such factors. In addition, all FBI activities must comply with the Constitution and all applicable statutes, Executive Orders, Department of Justice regulations and policies, and Attorney General guidelines.

2. Bureau of Alcohol, Tobacco and Firearms and Explosives

The Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF) targets and prevents the illegal use and trafficking of firearms, the illegal use and storage of explosives, acts of terrorism, acts of arson and bombings, and the illegal diversion of alcohol and tobacco products. In carrying out its work, ATF uses various forms of data analysis for strategic planning, resource allocation, and investigations of prior criminal events.

For example, ATF's Frontline model, an operational planning tool, helps ensure that the agency is maximizing its resources. Every field division must: (1) assess the violent crime environment within its jurisdiction; (2) create a strategic plan to address the violent crime issues identified in the assessment; (3) direct available resources in support of the strategic plan; and (4) measure the impact of its efforts, retooling as necessary to meet identified goals and objectives. The central component of Frontline is the domain assessment process. Each field division gathers and analyzes intelligence to identify its unique violent crime environment, develop strategies to address the drivers of violent crime in the jurisdiction, and prioritize resources for inspection, investigation, and enforcement activities.

In addition to this overall approach to intelligence and data-driven management, ATF uses data analysis to develop specific leads in criminal cases, as well as to establish patterns of

criminal activity. For example, the Bomb Arson Tracking System (BATS) permits State, Local, and other Federal law enforcement agencies to share information related to bomb and arson investigations and incidents. BATS helps track the material components of bombs, intended targets, and other methods across investigations. Analyzing this data can identify patterns and help connect seemingly disparate criminal events. Another example is GangNet, a program that tracks data about gangs and gang incidents and facilitates information-sharing across departments, agencies, states, and regions. The data can be used to discern trends, relationships, and patterns in gang behavior and demographics.

ATF also created e-Trace, a web-based crime-gun tracing system, to streamline tracing efforts for national and international law enforcement agencies. The e-Trace database allows users to query firearms recovery data such as possessor name and associates, recovery location and date, type of crime, serial number and other identifiers. The system links separate recoveries that have matching data. ATF analyzes e-Trace and multiple sales data against FBI Uniform Crime Reporting data and census data to predict crime hot spots, and to target those hot spots through enhanced enforcement initiatives.

3. U.S. Marshals Service

The United States Marshals Service (USMS) apprehends federal fugitives, protects the federal judiciary, operates the Witness Security Program, houses and transports federal prisoners, and manages and sells seized assets as part of the Asset Forfeiture Program.

USMS uses various forms of data analytics in carrying out its work. In particular, the Firearm Incident Risk Examination program uses arrest, conviction, and current charge data to analyze shooting incidents between law enforcement and criminal offenders. The program identifies criminal and demographic attributes associated with firearm violence and uses the resulting analysis to assess the relative risk of violence in various apprehension or arrest scenarios. The analysis also informs risk-mitigation training and post-incident assessments. This program does not purport to predict future incidents.

USMS also uses data analysis in order to allocate personnel and other resources. Funding for local task forces is determined, in part, based on trend analyses of warrant and closed-case databases, to determine geographical areas likely to have the greatest number of fugitives.

4. Drug Enforcement Administration

The Drug Enforcement Administration (DEA) enforces federal controlled substances laws and regulations and supports programs that seek to reduce the availability of illicit drugs in domestic and international markets. DEA investigates the principal members of organizations involved in the growing, manufacture, or distribution of controlled substances. DEA also coordinates with Federal, State, Local and international law enforcement on mutual drug enforcement efforts.

Although DEA does not use forward-looking predictive analytics with big data sets, it uses other forms of data analysis in carrying out its work. In particular, the Agency maintains two important databases, which it uses in part to identify potential criminal activity or regulatory violations.

One of these databases, called ARCOS, houses data collected from manufacturers and distributors of Schedule I, II, or III narcotic controlled substances, who must report to DEA the sale, purchase, loss, or inventory adjustment of these controlled substances. DEA can use the database to monitor trends in the flow of controlled substances from their point of manufacture through commercial distribution channels to point of sale or distribution at the dispensing and retail level. DEA reviews this data to ensure that purchase, sale, and other transaction reports match. It also reviews the data for suspicious activity, such as massive or recurrent losses. Such suspicious activity could lead DEA to investigate a previously unidentified target.

Similar to ARCOS reporting, DEA registrants at all levels, including practitioners and pharmacies, must report all thefts and significant losses of controlled substances. This information is maintained in the Drug Theft Loss Database. As with ARCOS, DEA reviews this database for suspicious activities, which may lead DEA to investigate a previously unknown target.

B. *Department of Homeland Security*

In the conduct of its law enforcement, immigration, border, and transportation security missions, DHS collects an enormous amount of data related to the facilitation of travelers and trade. This data includes, but is not limited to, information on international imports and exports, international and domestic air passenger manifests, visa application data, border crossing data, and investigative data. Use of these broad datasets helps DHS identify and stop security threats, but also sweeps up detailed personal information about people who do not pose a risk to homeland security. DHS conducts privacy assessments of its programs that use personal data, and maintains policies to ensure that all data is used in accordance with the purposes for which it was gathered, consistent with the laws and policies governing the systems in which it was retained and analyzed. Analytical rules and processes relied upon to support DHS activities are subjected to human review in the inception of the analysis and many front line personnel still employ discretion in screening and enforcement decision making. Thus the underlying facts and analytical rules are generally visible to and understood by analysts and oversight personnel, and often to front line employees.

DHS is harnessing big data with information technology advances and the use of predictive analytics to enhance its security missions. DHS Science and Technology (S&T) is conducting research and development that is paving the way for next generation mission capabilities that will be relevant to law enforcement. In addition, S&T has established a big data laboratory that is being used to examine technical opportunities to improve the law enforcement mission in collaboration with DHS law enforcement components and organizations.

At the agency level, DHS is developing the DHS Data Framework, a wholly owned central data repository in both classified and unclassified domains. The DHS Data Framework will enable scalable and controlled aggregation of DHS datasets, for use by DHS components while protecting privacy. By allowing searches across a wide range of datasets, the Data Framework will enable agencies to uncover illicit activity that might otherwise be difficult to detect. This allows DHS to take immediate action when appropriate and also supports the development and enhancement of enforcement strategies.

In addition to the use of the DHS Data Framework, individual DHS law enforcement components are using big data and predictive analytics in furtherance of their missions. We describe below the component programs that DHS identified.¹⁸

1. U.S. Customs and Border Protection

U.S. Customs and Border Protection (CBP) works to keep the border secure while facilitating lawful travel and trade. CBP oversees customs, immigration, border security, and agricultural protection at the border and border checkpoints. CBP uses predictive analytics in its Automated Targeting System (ATS), which analyzes large volumes of trade and travel data to conduct risk assessments and identify cargo and persons requiring additional screening or inspection. ATS applies the following predictive analytics techniques:

- **User Defined Rules.** CBP personnel can develop rules to filter incoming data [REDACTED] to identify targets for further investigation. These rules typically are fairly specific; for example [REDACTED]. CBP management reviews all user defined rules before they are implemented.
- **Threat Modeling Predictive Algorithms.** CBP's threat algorithms are generated through statistical modeling based on historical data. These predictive models are typically broader than user defined rules and are developed in response to more generalized threats. Implementation generally follows formal briefings to CBP leadership. [REDACTED] and to respond to various threats.
- **Automated Targeting System-Land Predictive Analytics Targeting Rules.** CBP has deployed [REDACTED]—to determine which vehicles should be subject to secondary screening.

2. U.S. Immigration and Customs Enforcement

Immigration and Customs Enforcement (ICE) is a law enforcement entity that enforces federal laws governing border control, customs, trade, and immigration. ICE relies on predictive analytics in a number of programs, including the following:

¹⁸ The Transportation Security Administration (TSA) is not discussed here, because DHS does not consider TSA's screening activity to be a law enforcement function. Similarly, we do not discuss DHS programs that may be in development but not yet operational.

- FALCON Data Analysis and Research for Trade Transparency System (FALCON-DARTTS) is a data mining investigative tool containing [REDACTED] FALCON-DARTTS analyzes trade and financial data to identify statistically anomalous transactions that may warrant investigation for trade-based money-laundering and other import/export crimes. ICE investigators can also [REDACTED]
[REDACTED]
[REDACTED]
- FALCON-Roadrunner applies big data analytics to large-scale trade data systems associated with the import/export border enforcement functions of CBP and ICE. FALCON-Roadrunner analyzes trade- and law enforcement-related data from multiple sources across government, private sector, and foreign customs organizations to identify statistically anomalous trade transactions that may warrant investigation of export violations.

3. National Protection and Programs Directorate

The National Protection and Programs Directorate (NPPD) is charged with protecting and enhancing the nation's critical physical and cyber infrastructure. The Federal Protective Service (FPS) is a subcomponent of the NPPD and provides security and law enforcement services to federally owned and leased facilities. FPS worked with the DHS Science and Technology Directorate to develop the Threat Analysis Systems Requirements (TASR). TASR is designed to undertake a systematic evaluation and analysis of FPS threat-related systems, processes, and tools by reviewing, among other information, data sources, evidence-based countermeasures, and Information Technology systems and security requirements.

4. United States Coast Guard

The United States Coast Guard defends the maritime borders of the United States and works to protect the maritime economy, the maritime environment, and assist people in peril at sea. The Coast Guard has developed a program called Geographic Information Systems to perform statistical analysis to identify those times and locations where nefarious activity is most likely to occur. [REDACTED]
[REDACTED]
[REDACTED]

IV. APPLICATIONS IN POST-ARREST PROCEEDINGS

A number of Federal and State entities also have begun to use, or are considering using, predictive analytics in post-arrest proceedings, including at hearings on pretrial detention, at sentencing, and in determining appropriate supervision levels for individuals on probation or parole. As with other uses of predictive analytics in law enforcement, much of this development is at an early stage. Some examples are discussed below.

A. *Predictive Analytics to Aid Decision-Making relating to Pre-trial Release*

Some jurisdictions have begun to use big data-based predictive analytics to make pretrial detention decisions. These efforts are geared to identifying the individuals most at risk to violate conditions of release, and to addressing concerns that judges may detain low-risk offenders too often.

1. **Federal Pretrial Services Risk Assessment**

Federal agencies have implemented risk assessment analysis only recently. After one failed attempt to adopt such a tool in the early 1990s, the Administrative Office of the U.S. Courts and the Office of the Federal Detention Trustee revisited the issue in January 2009, publishing a substantial research report that set out to promote the use of alternatives to pretrial detention. The primary recommendation that emerged from that effort was for the development and implementation of an evidence-based pretrial risk assessment tool. The Office of Probation and Parole led this project, working with Dr. Christopher Lowenkamp, a nationally recognized expert in risk assessment in the community corrections context. They developed a tool—the Pretrial Services Risk Assessment (PTRA)—that was piloted in several districts before being introduced nationwide in September 2011.¹⁹

When a defendant makes an initial court appearance, the magistrate judge typically requires a pretrial services report based on an investigation by a pretrial services officer. The officer interviews the offender and prepares a report that documents the defendant's case information and criminal history, an assessment of whether the defendant is likely to appear for future court proceedings or presents a danger to the community, and the officer's recommendation as to whether the defendant should be released.

In developing a tool for use at the federal level, the Administrative Office considered adopting existing models from Virginia and Washington, DC, but concluded that the federal offender population was sufficiently distinct to warrant the development of a new tool using only federal data. Accordingly, PTRA was developed using the same data used in the 2009 research report, *i.e.*, case information and criminal history supplied by the Administrative Office on all persons charged with federal offenses between 2001 and 2007 who were processed by the federal pretrial services system. Using bivariate analyses and multivariate logistic regression models, eleven factors were identified as most predictive: (1) felony convictions; (2) other pending charges; (3) prior failures to appear; (4) current charge; (5) seriousness of current charge; (6) employment status; (7) substance abuse; (8) age at interview; (9) education level; (10) citizenship; and (11) home ownership. The resulting instrument produces a score that allows the offender to be classified into risk categories that are associated with rates of failure to appear, new criminal arrest, and technical violations leading to revocation. The instrument also includes nine unscored items that do not affect the risk score but were tracked for potential use in the future.

¹⁹ See generally Timothy P. Cadigan, James L. Johnson, & Christopher Lowenkamp, *The Re-Validation of the Federal Pretrial Services Risk Assessment (PTRA)*, FED. PROBATION: Vol. 76, No. 2 (Sept. 2012), available at http://www.uscourts.gov/uscourts/FederalCourts/PPS/Fedprob/2012-09/01_federal_ptr.html.

Early results from pilot programs in the District of Nebraska and Western District of North Carolina indicated that use of PTRAs led to more recommendations for release by the pretrial services officer, and increased actual release rates. A re-validation of the instrument, conducted using data on cases opened between 2010 and 2011, suggested that PTRAs have strong predictive validity with respect to accurately classifying defendants' risk level. The re-validation study also indicated that the unscored factors were not sufficiently predictive for inclusion in the tool, leading to their removal.

Despite these early indications of its utility, the actual use of PTRAs thus far has not been as widespread as Administrative Office had hoped. Although PTRAs are routinely generated by the pretrial services officer, they are very rarely included in the report itself, and often are completed after the report has been finalized. One possible explanation for this is lack of judicial buy-in. PTRAs were developed and implemented without judicial involvement, and anecdotal evidence suggests that judges do not find PTRAs helpful to their decision-making and prefer not to see them in the pretrial services report. Administrative Office efforts to expand the use of PTRAs are continuing.

2. Other Pretrial Risk Assessment Tools

The State of Kentucky and seven cities and counties throughout the country are using a tool called the Public Safety Assessment-Court (PSA-Court) to help guide judicial decision-making relating to pretrial release. This tool is the product of the Laura and John Arnold Foundation, which set out in 2011 to develop a universal risk assessment tool that could improve judges' release and detention decisions.²⁰ The Foundation's research team built a dataset drawn from 1.5 million cases from across the country and studied the cases in which the defendant had been released pretrial. The researchers identified hundreds of risk factors, but ultimately focused on nine factors – including age at current arrest, prior criminal convictions, and any prior failure to appear, among others. Demographic factors like socioeconomic status, education level, and neighborhood were not among the factors selected. Significantly, the research also disclosed that offender interviews, which are time-consuming for pretrial officers, are not necessary. Information derived from those interviews did not improve the predictive value of the tool.

The tool's outputs—rankings for each individual on three risk scales (new criminal activity, new violent criminal activity, and failure to appear)—do not bind judges. But the Foundation notes that in the six months after Kentucky deployed the PSA-Court tool in July 2013, the arrest rate for released defendants declined from 10% to 8.5%, even though slightly more defendants were released pending trial. At the same time, the rate of those who failed to appear for court did not increase. In evaluating the tool, the Foundation concluded that PSA-Court had no discriminatory impact on minorities or women.

²⁰ This discussion is drawn from Laura and John Arnold Foundation, *Results from the First Six Months of the Public Safety Assessment – Court in Kentucky* (July 2014), available at <http://www.arnoldfoundation.org/sites/default/files/pdf/PSA-Court%20Kentucky%206-Month%20Report.pdf>, and from Telephone Interview Anne Milgram, Vice President of Criminal Justice, Laura and John Arnold Foundation (Oct. 29, 2014).

B. *Use of Predictive Analytics in Decisions Regarding Supervision Level for Individuals on Probation/Parole*

Probation and parole agencies historically have attempted to assess individual risk in order to determine an appropriate level of supervision for parolees and to prioritize resources to monitor those individuals at greatest risk to reoffend. Such assessments have relied upon individuals' criminal history and the nature of the offenses for which they were sentenced. Recently, probation and parole departments have incorporated more advanced data analysis techniques and computer modeling into their risk assessments. Some of these methods are discussed below.

1. Federal Post-Conviction Risk Assessment (PCRA)

The Administrative Office of the United States Courts' Office of Probation and Pretrial Services has long used risk assessment tools to help determine an appropriate level of supervision for individuals on post-conviction supervision.²¹ The federal probation system has used some form of risk assessment since the 1970s. In the early 1990s, the Federal Judicial Center, the education and research arm of the federal court system, began developing a new risk model for the federal probation system based on factors identified by a multivariate regression analysis of federal cases. The tool, known as the Risk Prediction Index (RPI), was approved for all new probation cases in 1997 and included information about the age of the offender at the start of supervision, number of prior arrests, whether a weapon was used to commit the instant offense, employment status, history of drug and alcohol abuse, history of circumventing supervision, education level, and whether the individual was living with a spouse and/or children at the time supervision began. Probation officers used the score to help determine an appropriate level of supervision.

In 2009, the Administrative Office wished to create a new federal probation risk and needs assessment tool that, unlike the RPI, could take into account "dynamic" factors associated with recidivism, such as antisocial attitudes and associates, and could allow regular reassessment to permit officers to determine whether supervision strategies were reducing that risk. Working again with Dr. Lowenkamp, the Administrative Office set about developing an instrument that could (1) identify the offender's risk level; (2) suggest ways to bring down the offender's risk level; and (3) identify the possible obstacles to intervention.

Informed by existing research, developers used a variety of regression analyses to determine the most predictive elements to be used in the instrument, including criminal history, education level, employment history, substance abuse, social networks and self-reported attitudes and cognitions. The resulting tool, called the Post-Conviction Risk Assessment (PCRA), includes "scored" items that have been demonstrated by the Administrative Office's own empirical research to be statistically significant predictors of recidivism. These items contribute

²¹ Administrative Office of the United States Courts, Office of Probation and Pretrial Services, *An Overview of the Federal Post Conviction Risk Assessment* (September 2011), available at http://www.uscourts.gov/uscourts/FederalCourts/PPS/PCRA_Sep_2011.pdf (hereinafter, "PCRA Report"); Interview with John Fitzgerald, Chief, Criminal Law Policy Staff, Administrative Office of U.S. Courts, Probation and Pretrial Services Office, in Washington, D.C. (October 27, 2014) (various members of Criminal Law Policy Staff involved in discussion).

to a determination of the offender's predicted risk level, and include criminal history factors, education level, employment history, drug and/or alcohol abuse, and family dynamics (whether the offender lives with a spouse or children, family stability, relations with peers and other social support). The tool also includes various unscored items that do not contribute to the individual's risk level but are being tested to determine if they can improve prediction. Taking into account the factors identified above, the PCRA classifies each offender within one of four risk categories. The vast majority of offenders—roughly 85%—are classified as low or low/moderate risk, with a rearrest rate of around 5%. Individuals classified as low-risk are subject to minimal supervision requirements and districts are given the discretion to have these offenders check in with their probation officers through mail and electronic means. For other offenders, the PCRA helps probation officers determine an appropriate level of supervision and support services on an individualized basis.

The PCRA was implemented between 2010 and 2012. The AO's internal, preliminary evaluation indicated that the PCRA is fairly accurate with respect to predicting the likelihood that offenders in each risk category will be rearrested, performing similarly to the RPI or the U.S. Sentencing Commission's Criminal History Score.

2. State and Local Efforts

Philadelphia

The Philadelphia Adult Probation and Parole Department (APPD), a county-level probation and parole agency, recently developed a tool to assess the relative risk of recidivism for each individual under the APPD's supervision (the "APPD Risk Tool" or "Risk Tool").²² The APPD hoped to target its supervision resources more efficiently, to those most at risk of reoffending.

Prior to the APPD's development of the Risk Tool, the majority of offenders released on probation or parole in Philadelphia were assigned to general supervision, regardless of their offense of conviction, prior criminal record, or the risk they might pose to the community. Typically, individuals placed in general supervision were required to report in person once a month and to meet with a supervision officer for twenty to thirty minutes. In 2005, APPD personnel approached the University of Pennsylvania's Jerry Lee Center of Criminology, expressing a desire to create a system that would allow APPD to focus its resources on individuals who were the most likely to commit serious crimes in the future, while substantially scaling down its supervision of individuals who were thought to present little risk of reoffending.

²² For a complete discussion of the development and application of the APPD Risk Tool, see Geoffrey C. Barnes and Jordan M. Hyatt, *Classifying Adult Probationers by Forecasting Future Offending: Final Technical Report* (Mar. 2012) (evaluation supported with NIJ grant funding), available at <https://www.ncjrs.gov/pdffiles1/nij/grants/238082.pdf>. The discussion herein is informed largely by that report, as well as through telephone interviews with Dr. Richard Berk, Professor of Criminology and Statistics, University of Pennsylvania (Oct. 6, 2014) and Dr. Ellen Kurtz, Director of Research, Philadelphia Adult Department of Probation and Parole (Oct. 9, 2014).

Building from data from more than 100,000 past APPD parole cases, a Penn researcher and the APPD constructed the Risk Tool using a statistical method known as “random forest” modeling. Random forest modeling is a method by which a large number of individual regression trees—typically hundreds or thousands—are run separately, with each tree selecting predictive factors at random. The individual regression trees then each “vote” on an outcome, and the outcome with the most “votes” becomes the overall outcome for the model—in this case, an estimated risk level.

The APPD Risk Tool was first implemented in March 2009, and the most recent model has been in effect since November 2011. The current version incorporates twelve predictive factors, including the offender’s age at the beginning of the probation case, residential zip code, and various other factors relating to the individual’s criminal history. According to the Risk Tool, the variable with the most predictive value in Philadelphia is the number of prior stays in the county’s prison system, followed closely by the offender’s zip code, the amount of time between the instant offense and the most recent prior serious offense, the offender’s current age, and the age at which the offender was first charged as an adult.

The APPD Risk Tool evaluates each new parole case at the intake level, classifying it as high-, moderate-, or low-risk, and the assigned risk level determines the degree of supervision the offender will receive.²³ Individuals considered high-risk are predicted to commit one or more serious offenses within the first two years after the case start. These individuals take drug tests every week, and parole officers conduct field visits twice per month. High-risk individuals who violate the terms and conditions of their parole are typically subject to re-arrest immediately. Moderate-risk individuals are predicted to commit only non-serious offenses during this time period. These individuals must meet with their parole officers once a month, and parole officers do not conduct field visits. Low-risk offenders are not predicted to commit any new offense of any kind and are subject to very little supervision; they must meet with their parole officers only four times per year.

Since implementing the Risk Tool in 2009, the APPD has seen a substantial decrease in reoffending among its adult probation and parole population. For the high-risk population, the rate of reoffending with a serious crime in the first two years decreased from 21% in 2009 to 10% in 2011; among the moderate-risk population it decreased from 11% to 5%; and among individuals classified as low-risk, from 4% to 2%.²⁴ While overall crime rates are down, it is notable that APPD used fewer supervision officers than it had in 2009.

²³ The number of individuals that could be classified in each risk level was largely determined by the APPD’s estimates of how many individuals it *could* effectively supervise in each risk group. In other words, the APPD had to estimate how many offenders could be classified as high-risk, taking into consideration the limited resources it possessed to provide the kind of intense supervision that these individuals would receive. The APPD ultimately decided that had the resources to classify approximately 15% of its probationers as high-risk, 25%-30% as moderate-risk, and 55%-60% as low-risk. Barnes and Hyatt, *supra* note 22 at 7.

²⁴ The last statistic suggests that low-risk individuals actually suffer detrimental effects from over-supervision, a theory that finds some support in academic literature. See Christopher T. Lowenkamp & Edward J. Latessa, *Understanding the Risk Principle: How and Why Correctional Interventions Can Harm Low-Risk Offenders*, TOPICS IN COMMUNITY CORRECTIONS (U.S. Dept. of Justice, National Institute of Corrections, D.C.), 2004, at 3, 6, available at http://www.uc.edu/content/dam/uc/ccjr/docs/articles/ticc04_final_complete.pdf.

Maryland

The Maryland Department of Public Safety and Correctional Services (DPSCS) employs a risk assessment tool similar to Philadelphia's for supervising its probation and parole population.²⁵ The Maryland program got its start in 2007, when DPSCS noted that over 30% of all people arrested for murder in Baltimore were under DPSCS supervision at the time of their arrest. At the time, Maryland had a stratified supervision system, with 11 different risk levels. DPSCS wished to streamline this system, focusing its efforts on the individuals most likely to commit, or become victims of, violent crime.

Working with the same University of Pennsylvania researchers who constructed the Philadelphia model, DPSCS constructed a risk assessment tool based on a review of murder and shooting data in Baltimore City. Like the Philadelphia program, the Maryland risk assessment tool uses random forest modeling to assign a risk level to each new individual under supervision. The tool considers a number of factors, including the offender's age, sex, education level, employment and marital status, zip code, criminal history, and any history of drug or psychiatric treatment.

Using the tool, DPSCS identified 2,300 offenders as having the highest propensities for violence out of the approximately 70,000 individuals under DPSCS supervision. These individuals then were enrolled in a heightened supervision program known as the Violence Prevention Initiative (VPI). The program requires offenders to check in with their community supervision officers five days a week and meet with an officer in person on a weekly basis. By contrast, the lowest-risk individuals may check in through a kiosk system. In addition to the automated risk assessment tool, DPSCS considers other factors, such as whether the offender is a known high-ranking gang member, whether he or she has been a shooting victim in the recent past, and the offender's history of supervision by the VPI unit. If those factors are present, the offender may be placed in VPI even if such placement is not recommended by the risk assessment tool.

DPSCS policy requires probation officers to seek a warrant immediately when any VPI offender is re-arrested. If the individual does not commit a new offense within 90 days, he is allowed to transition out of VPI to a lower supervision level. Maryland currently has between 3,000 and 4,000 individuals enrolled in VPI statewide. The state also established a version of the VPI program for juvenile offenders.

C. *Risk Analysis Calculations in the Sentencing Context*

As States face budget shortfalls and strive to reduce overcrowding at correctional facilities, the past few years have seen a resurgence in states' interest in using risk assessment in sentencing.²⁶ Several states have used new technologies and computer modeling techniques that

²⁵ This discussion is drawn from telephone interviews with Tammy Brown, Executive Director, Governor's Office of Crime Control and Prevention (Oct. 14, 2014), and Kristen Mahoney, Deputy Director for Policy at the Bureau of Justice Assistance (formerly policy advisor for Maryland Governor Martin O'Malley) (Oct. 7, 2014), and from Governor's Office of Crime Control & Prevention, *Fact Sheet: Violence Prevention Initiative* (May 2014), available at <http://www.goccp.maryland.gov/msac/documents/FactSheets/VPI.pdf>.

²⁶ John Monahan and Jennifer L. Skeem, *Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning*, 26 FED. SENT. R. 158 (2014).

have the power to synthesize multiple data sets to estimate risk for committing future crimes on an individual-offender basis. Some examples are discussed below.

Virginia

In 1994, the Virginia General Assembly passed a comprehensive sentencing reform package which, among other things: (1) abolished the use of parole in the state; and (2) created the Virginia Criminal Sentencing Commission (VCSC) to develop, implement, and administer felony sentencing guidelines in the Commonwealth. To offset the increase in prison population that was expected to result from the elimination of parole, the legislation required the VCSC to develop an empirically based system that could be used to divert 25% of the lowest-risk individuals who would otherwise be facing incarceration under Virginia's sentencing guidelines away from corrections institutions to alternative punishments.

The VCSC developed the Nonviolent Offender Risk Assessment Instrument (RAI) with the goal of identifying, among individuals convicted of non-violent fraud, larceny, and drug offenses and who had no previous history of violent crime, those who had the lowest likelihood of committing another felony crime within three years.²⁷ The VCSC employed logistical regression analysis to identify the factors most predictive of recidivism. The VCSC ultimately identified eleven specific factors: gender, age, marital status, employment status, whether the offender acted alone when committing the crime, whether the offender was convicted of additional offenses at the same time, whether the offender had been arrested or incarcerated in the past year, prior criminal record, prior drug felony convictions, prior incarcerations as an adult, and prior incarcerations as a juvenile.²⁸ Each of these factors was weighted based on its relative degree of predictive power and incorporated into a risk assessment worksheet to be completed for each eligible offender.

The tool was tested in six of Virginia's 31 judicial circuits, and an independent evaluation by the National Center for State Courts (NCSC) found that the RAI "provides an objective, reliable, transparent, and more accurate alternative to assessing an offender's potential for recidivism than the traditional reliance on judicial intuition or perceptual shorthand."²⁹ The tool was implemented for statewide use in 2003. In July of 2013, the VCSC refined the RAI, removing from the model certain demographic information on employment and marital history. The VCSC also implemented a slightly different tool for drug offenders than that used for those convicted of fraud or larceny.

²⁷ In addition to the cited sources, this discussion is informed in large part by a Telephone Interview with Meredith Farrar-Owens, Director, Virginia Criminal Sentencing Commission (Oct. 16, 2014).

²⁸ Brian J. Ostrom, Matthew Kleiman and Fred Cheesman, *Offender Risk Assessment in Virginia: A Three-Stage Evaluation*, National Center for State Courts (2002), at 27. This evaluation was funded through a grant from the National Institute of Justice. The VCSC affirmatively decided not to use race as a predictor variable, despite the fact that it showed statistical significance during model development. This is because the Commission believed that the race variable was actually serving as a proxy for other variables that were difficult to measure, such as economic deprivation, educational disadvantage, family instability, and limited employment opportunities, which factors may disproportionately affect the African-American community. *Id.* at 27.

²⁹ Matthew Kleinman, Brian J. Ostrom and Fred L. Cheeseman, *Using Risk Assessment to Inform Sentencing Decisions for Nonviolent Offenders in Virginia*, 53 CRIME & DELINQUENCY 106, 126 (January 2007), available at <http://www.sagepub.com/spohnstudy/articles/5/Kleiman.pdf>.

Like the Virginia sentencing guidelines themselves, the application of the RAI recommendation is discretionary; the judge may implement or disregard the recommendation as he or she sees fit. The NCSC study found that of all cases recommended by the RAI for diversion, judges diverted approximately 40% of the time for larceny offenses, 59.6% of the time for drug offenses, and 65% of the time for fraud offenses.

In 1999, the Virginia Assembly again directed the VCSC to develop a risk assessment tool; this time, however, the tool was to be used to increase, rather than decrease, the punishment recommended by the sentencing guidelines.³⁰ The new tool, in use since July 2002, identifies the sex offenders with the highest probability of committing another sex crime or other crime against a person, such as assault, kidnapping, murder, robbery, or stalking, within the next five years. Like the RAI, the tool was developed using logistic regression analysis to identify the factors most indicative of recidivism risk. For the sex offender tool, these factors included the offender's criminal history, the offender's relationship to the victim, and details of the crime of sentence, in addition to certain demographic information such as age, employment status, and education level (whether the offender completed the 9th grade). Individuals identified as high-risk by the tool (approximately 5% of all sex offenders) are subject to a maximum sentence as much as three times as high as that recommended under the Virginia sentencing guidelines. Again, the recommendation produced by the sex offender risk assessment tool, like the Virginia sentencing guidelines themselves, is entirely discretionary. The VSCS noted that most judges continue to sentence sex offenders within the traditional guidelines range, without making any upward adjustment.

Utah

The Utah Sentencing Commission has been using a tool known as the Level of Service Inventory-Revised (LSI-R) for approximately a decade as part of the presentence investigation report.³¹ The LSI-R consists of 54 variables, including criminal history, education, employment, financial status, residential stability, family/marital status, substance abuse, and social attitudes. The LSI-R classifies individual offenders into one of five risk categories, though jurisdictions using the tool can customize the categories according to their own needs. In Utah, scores from the LSI-R are provided to sentencing courts, which consider them alongside the state's sentencing guidelines. Utah also uses LSI-R scores in probation and parole determinations.

Pennsylvania

In 2010, the Pennsylvania General Assembly passed comprehensive sentencing reform legislation, which, among other things, instructed the Pennsylvania Commission on Sentencing (PCS) to develop a risk assessment tool for courts to use in sentencing criminal offenders. Unlike Virginia's RAI, however, which is narrowly geared toward allowing alternative sentences

³⁰ See Virginia Criminal Sentencing Commission, *Assessing Risk Among Sex Offenders in Virginia* (January 2001), at 7, available at http://www.vcsc.virginia.gov/sex_off_report.pdf.

³¹ See *Risk Redux*, *supra* note 26, at 159-60.

for certain individuals convicted of specific non-violent offenses, Pennsylvania's tool is intended to apply to substantially all offenders in the system.³²

The Pennsylvania risk assessment tool is still in the development phase, but initial discussions suggest that predictive factors will likely include gender, age, details of the offense of sentencing, and criminal history. The PCS hopes to test the tool in 2015 but there is no definitive timeline for implementation. Pennsylvania's sentencing guidelines, like Virginia's, are discretionary, and judges would not be required to take the predicted risk level into account when sentencing a particular offender.

Maryland

The Maryland State Commission on Criminal Sentencing Policy (MSCCSP) is also developing a risk assessment tool for use in sentencing.³³ Maryland's efforts to use risk assessment in sentencing came about as a result of the state judiciary's interest in evidence-based sentencing. Maryland is still in the very early stages of developing its sentencing tool and has yet to determine what predictive factors will be included or how the tool will be incorporated into a sentencing recommendation. Maryland may be considering a framework like Virginia's Nonviolent Offender RAI, which uses the assessment as a ground for recommending an alternative to incarceration for low-risk offenders, rather than as a basis for enhancing the sentences of individuals deemed to present a higher risk. Maryland's sentencing guidelines are voluntary; accordingly, like the other jurisdictions using these tools, their precise application would be entirely within the judges' discretion.

Several states in addition to those discussed above are actively considering the use of similar statistical instruments to assess risk. From the growing number of jurisdictions using or developing these tools, the interest in using risk assessment techniques to inform sentencing decisions appears to have been gathering momentum over the past decade. The use of risk assessment tools in sentencing—particular those that take into account demographic characteristics such as education level and socioeconomic background—raise a number of concerns, which are discussed in more detail below.

V. POTENTIAL BENEFITS AND POLICY CHALLENGES

When developed correctly, data-based methodologies can lead to law enforcement decisions based on factors and variables that empirically correlate with risk, rather than on human instincts and prejudices about places and people that are not necessarily grounded in fact. There is inherent value in reducing the kind of subjective decision-making that can result in faulty or inequitable application of the law.

³² Telephone Interview with Mark Bergstrom, Executive Director, Pennsylvania Commission on Sentencing (Oct. 9, 2014).

³³ Telephone Interview with David A. Soulé, Executive Director, Maryland State Commission on Criminal Sentencing Policy (Oct. 15, 2014).

Predictive analytics also might promote more efficient use of law enforcement resources, which, in this era of tightening budgets, is particularly significant. If police resources are deployed with greater efficiency, there should be a corresponding gain in public safety for any number of reasons: offenders might be apprehended more often, an enhanced police presence might deter crime, and police might acquire additional opportunities to build more productive relationships with the communities they serve. The degree of marginal benefit afforded by big-data predictive analytics—as opposed to the kind of analytics that police forces were using until recently—is still being explored, and ought to continue to be a focal point of research.

But just as big-data predictive analytics might yield significant benefits if deployed correctly, these same techniques raise policy questions and potential concerns, depending on the variables at play and the purpose for which the analysis is brought to bear. To some extent, these issues are the same ones that would apply to “little-data” analytics: any analysis that takes characteristics associated with groups—and then brings those characteristics to bear in decision-making about individuals—raises issues about which variables should be considered and the appropriate contexts to which the analysis should be applied.

Legal authorities provide some guidance about the degree to which race, national origin, and other protected or immutable characteristics may be considered. Critics have noted that proxies for protected characteristics, or for socioeconomic characteristics, can make their way into analyses as well. Even when the variables seem neutral, any model is susceptible to importing any biases reflected in the underlying data.

There is also a fundamental question about what decisions should be based on historical, broad-based data rather than on individualized conduct. It would make little sense to deploy resources without an understanding of where they are most needed, and fewer concerns have been raised about the potential for misuse of data for these purposes (although, at a basic level, additional police deployment can mean additional law enforcement scrutiny for individuals who live in those areas). On the other hand, making decisions about sentencing—where individual liberty is at stake in the most fundamental way—based on historical data about other people, raises additional questions. As the Attorney General has cautioned, criminal sentences should not be based on static factors and immutable characteristics unrelated to the criminal conduct at issue, such as a defendant’s education level, socioeconomic status, or neighborhood of residence. Reliance on these factors may unintentionally exacerbate unjust disparities in our criminal justice system by, for example, resulting in different sentences for two otherwise similar defendants who reside in different zip codes. Likewise, the length of a defendant’s prison term should not be adjusted simply because a statistical analysis has suggested that *other* offenders with similar demographic profiles will likely commit a future crime. Instead, equal justice demands that sentencing determinations be based primarily on the defendant’s own conduct and criminal history.

VI. CONCLUSION

Law enforcement agencies at the Federal, State, and Local level have long shared the goals that make up the mission statement of the Department of Justice: ensuring public safety, preventing and controlling crime, seeking just punishment for those who break the law, and ensuring the fair and impartial administration of justice. To fulfill this mission, law enforcement agencies seek out new and better tools to focus resources on preventing crime and protecting the

public. These principles also require law enforcement to be mindful of how new tools impact civil liberties or might lead to unfairness.

Predictive policing methods should be considered against the backdrop of these core law enforcement missions and in comparison to the traditional methods and practices they might replace. It is critical that we examine and understand both the potential value of these new methods as well as any unintended consequences they may produce. Some supporters of these methods have urged that they are more accurate and, at least arguably, more objective, than human decision-makers who currently make assessments like probation or parole determinations based on their individual judgment. These assertions should be tested. To that end, the Department recently asked the United States Sentencing Commission to study the use of data analytics in both the sentencing and reentry contexts, and to issue policy recommendations based on its analysis; the Commission has agreed to undertake this very important initiative.

More broadly, we should continue the conversation about predictive analytics methods that has already begun between key stakeholders—ranging from law enforcement agencies to academics, community leaders, and civil society groups. As the conversation continues, we should consider a range of questions, such as:

- Does use of a given predictive analytics tool lead to an improvement in public safety outcomes when compared to existing law enforcement methods?
- Does the use of a given technique result in a greater or lesser disparate impact on marginalized communities than the use of existing law enforcement methods?
- Can any given technique be modified to further minimize any disparate impact without compromising its predictive value?
- Are there certain factors that should not be relied upon by predictive analytics models? Does the answer depend on how the results of the model are used?
- How should techniques be evaluated on each of these questions over time?
- Are there training guidelines, best practices, or other protections that law enforcement agencies can adopt to ensure that predictive analytics are being used in a manner that ensures civil rights protections?

The Department of Justice is also eager to work collaboratively with the full range of stakeholders to develop best practices in this emerging area. To that end, as a follow-up to this report, the Department will work with stakeholders and develop guidance for the use of predictive analytics by State and Local law enforcement agencies—those on the cutting edge of predictive policing.

The development of predictive techniques should continue to be driven by the core law enforcement goals of protecting the public and ensuring fairness in our justice system. These goals are complementary, not contradictory, and it is the role of all who care deeply about law enforcement's mission to ensure that predictive methods live up to their promise.