ANNUAL REVIEWS

Annual Review of Criminology

Policing in the Era of Big Data

Greg Ridgeway

Department of Criminology and Department of Statistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA; email: gridge@upenn.edu

Annu. Rev. Criminol. 2018. 1:401–19

First published as a Review in Advance on September 27, 2017

The *Annual Review of Criminology* is online at criminol.annualreviews.org

https://doi.org/10.1146/annurev-criminol-062217-114209

Keywords

policing, Big Data, predictive policing, open data

Abstract

Fifty years ago, the 1967 President's Commission on Law Enforcement and Administration of Justice urged the rapid adoption of information technology to improve the effectiveness, efficiency, and fairness of the criminal justice system, including policing. They predicted that we could make great progress on the challenge of crime if only we could deliver the right information to the right police officer at the right time. In this twenty-first century era of Big Data, all the technologies described in the 1967 Commission report are widely available and accessible to police departments. This review characterizes what Big Data means for policing, discusses the technologies making Big Data possible, describes how police departments are putting Big Data to use, and assesses how close we are coming to realizing the vision offered in 1967. Although police may be rich in data, we still need to improve the extraction of information and knowledge from that data and put them to use to decrease crime and improve clearance rates.

> Modern technology can make many specific contributions to criminal administration. The most significant will come from the use of computers to collect and analyze the masses of data the system needs to understand the crime control process.
>
> —*The Challenge of Crime in a Free Society*, 1967

INTRODUCTION

Five decades ago, the 1967 President's Commission on Law Enforcement and Administration of Justice foresaw all the potential of information technology to improve the effectiveness, efficiency, and fairness of the criminal justice system. For police, the report proposed the development of "portable recording devices" to facilitate data collection, computers that could automate the dispatch of patrol cars closest to calls for service, networked alarms that could notify nearby officers without the intervention of a dispatcher, alteration of police deployments in real-time as data reveal emerging problems, and new wireless networks to reduce communication congestion.

As early as 1964, the St. Louis Police Department was using data and modeling strategies for patrol allocation (Crowther 1964). Over the next decade, researchers were active in developing the mathematical foundation, data requirements, and computational resources to align police resources with the times and places that the public needed them (Chaiken & Dormont 1978). Other technologies, such as automatic vehicle location (AVL), also emerged during this period, signaling that police visionaries imagined harnessing data and computing in new ways to improve police performance.

Nonetheless, these early studies discuss the challenges of police data science in that era; challenges included programs created using languages with limited support (Michigan Algorithm Decoder), lack of technical expertise or rough user interfaces [the Los Angeles Police Department (LAPD) could not adapt software to its data], challenges in accessing computing resources [the New York Police Department (NYPD) connected to an MIT computer through a telephone line], the cost of the machine (IBM System 370 machines with 1 MB of memory and 800 MB of storage cost $25 million in 2017 dollars). In this period, data were too coarse and computing too inaccessible to completely implement the vision of data-driven policing.

Because of numerous contributions from computer scientists, engineers, analysts, and criminologists, many of the technological innovations foreshadowed in the 1967 President's Commission report are now widely available. Data are more refined, software is easier to develop, and computing storage and processing are inexpensive. Policing has unquestionably benefitted from the current (and ongoing) Big Data revolution.

However, now that the Era of Big Data has arrived, the challenge rests in how to use data effectively and provide officers useful new information and not simply data that reiterate what they already know. Data delivered at the wrong time, data being delivered in volumes too large to consume in a timely manner, or data that are too sparse to be helpful may soon become noise to an officer and be ignored altogether. Big Data's potential overshadows these challenges. Police data seemingly contain information that should be able to shed light on unsolved crimes, high-risk places, and police performance if only we can figure out what to collect, analyze, and deliver. In this review, I characterize what Big Data means for policing and discuss the technologies making Big Data possible, how police departments are putting Big Data to use, and how close we are to realizing the vision offered in 1967.

BIG DATA DESCRIBED

Policing's data environment today has certainly changed from 1967, but some key features of Big Data have had a powerful impact on policing. Big Data is a loose description for the general idea of integrating large data sets from multiple sources with the aim of delivering some new, useful insight from those data. Some writers focus on the systems necessary to efficiently store, manage, and query the data (Marz & Warren 2015). Other writers focus on the analytic tools needed to extract meaningful insights from these massive data sources (Dean 2014). The need for a careful definition of Big Data might not even be necessary, as Davenport (2014) wrote in *Big Data @ Work* that he and others "predict a relatively short life span for this unfortunate term."

Rather than wrestle further with producing a rigorous definition of Big Data for policing, I discuss the key Big Data themes affecting policing that will survive past the life span of the term Big Data. Three themes frequently described in Big Data discussions are the three Vs: volume, velocity, and variety (Laney 2001). Police are experiencing all three.

Volume

Data volume is simply the magnitude of the data police collect and store. A 30-minute body-worn camera video produces 800 MB of data, which would have exceeded the capacity of a $25 million computer at the time of the 1967 commission report. A Houston police captain reported that 100 cameras were producing about 1 TB of data per day (Meriwether 2014). Whereas Moore's Law predicts the doubling of transistors on an integrated circuit, and hence a doubling of computing power, every 18 months, Kryder's Law (Walter 2005) predicts a doubling of data storage capacity per square inch of media every 13 months. Even though these are not laws in the scientific sense, they have generally been consistent with trends. Data storage is simply cheaper.

Infrastructure as a Service (IaaS), the provision of computation and data storage over the internet, has put further downward pressure on costs of equipment and costs to manage that equipment. The International Association of Chiefs of Police (IACP) report, "Cost savings, rapid deployment of critical resources, off-site storage and disaster recovery, and dynamic provisioning of new and additional resources when needed are among the tangible benefits that cloud computing potentially offers to law enforcement agencies of all size" (Int. Assoc. Chiefs Police 2015, p. 1). In a survey of 272 law enforcement agencies, 16% of agencies indicated that they are already using cloud services, primarily for email and data storage (Ponemon et al. 2013). The primary drivers of the move to cloud services are cost reduction and software and hardware management. More departments are moving in this direction, with 38% of respondents indicating that they are going to expand their use of cloud computing.

Chaiken & Dormont (1978) indicate that they shrunk their patrol car allocation program and data to 160 kB so that running the program would be cost-effective (about $55 per run in 2017 dollars). Today's vastly lower costs and easier access to data storage have made it possible to analyze large data sets with more complex models, making Big Data feasible even for small agencies.

Velocity

Progress in policing and Big Data is due not only to being able to store more data but also to the greater availability and accessibility of data. The confluence of technological advances in economical storage and high-speed wireless networks means data transfer at high velocity.

The 1967 President's Commission report discusses the inadequacy of bandwidth for radio communication in Chicago, with officers needing to wait their turn for voice communication over the radio. Indeed, networks can still become congested, but police today are pushing enormous volumes through wireless communication networks.

Long-Term Evolution (LTE) wireless networks have had a transformative effect on police information technology. LTE networks offer high data-transfer rates (30 minutes of body camera footage in three minutes) and support for devices moving at speeds up to 220 miles per hour and allow police to use less-expensive equipment with unified standards because of the widespread coverage of LTE (Doumi et al. 2013). With these specifications, police departments have been able to deploy more data devices, aiding in both the collection and distribution of data.

With LTE, police have technologically achieved the Commission's aim of having portable equipment for data collection. High uplink speeds mean that police officers and police equipment in the field can stream data to centralized data storage. Other devices can passively collect and transmit data without officer interaction. AVL reports the location of police vehicles as frequently as every six seconds. Telematics, systems that track the operation of a police vehicle, constantly stream vehicle and driving data back to a centralized system. Other public-safety data collection systems, such as acoustic gunshot detectors, also connect through LTE. Rather than wait for the next day, supervisors and police strategists can have a constant view of police positions and activities as they are occurring.

High downlink speeds mean that police in the field can more readily access information about people and places. To showcase LTE's potential, the US Department of Homeland Security, Purdue University, and the Chicago Police Department conducted a pilot test in which they successfully streamed 720p video over LTE to a moving police car (DHS Cent. Vis. Anal. Command Control Interoper. 2015). The researchers tested the limits of the network, including the range and varying levels of network congestion. Although these affected the quality of the video, the pilot test demonstrated the potential to stream high volumes of information from police data sources to officers in the field. If video is possible, then LTE networks can transmit all other forms of data, such as text, photos, and audio, to officers when and where they need the information.

Variety

With the ability to store and transmit Big Data, police now utilize a wide variety of data sources. Police continue to document the traditional law enforcement transactions, such as stops, arrests, and crime reports. However, police now access unprecedented new data sources.

Telematics transmits (via LTE) data on acceleration, braking, indicators of dangerous driving, seat-belt usage, fuel consumption, and other measurable aspects of vehicle usage. LAPD has outfitted 110 vehicles and has submitted a 2017–2018 budget plan to install 325 more systems, completely outfitting all marked vehicles (Los Angel. Police Dep. 2016).

Community surveillance systems, such as acoustic gunshot detection, cameras, and license plate readers, also generate new sources of data for police. The 1967 Commission report anticipated alarm networks that could relay information to nearby officers automatically. Indeed, some systems tie separate sensors in networks with software piloting the coordination of the technology and calling on human interaction when needed. As an example, one chief described how their acoustic gunshot detection system activated a nearby camera that aimed to capture the license plates of nearby vehicles, one of which the shooter jumped in after leaving a gunshot victim in the street (Police Executive Res. Forum 2012).

Just as declines in cost have facilitated the police's ability to store and transmit data, the emergence of data standards has facilitated the police's ability to absorb data from multiple sources. The Global Justice XML Data Model (GJXDM) and the more general National Information Exchange Model (NIEM), for example, helped establish a common data vocabulary and data grammar so that different systems and different components of the justice system can share information (Ioimo 2016). Technologists have urged police when acquiring new data systems to insist on NIEM compatibility and the ability to export records, permitting the police data operations to be more flexible and not locked into a particular platform (Hollywood & Winkelman 2015). To date, NIEM has been sufficiently flexible to admit new data types, including data from forensic evidence, private alarm monitoring companies, and disaster assistance.

Other Big Data Themes

In addition to Laney's (2001) volume, velocity, and variety, emerging literature on Big Data has noted additional common features (Kitchin 2014). Often the data sources are exhaustive. Unlike statistical collections based on surveys or other samples of populations, Big Data sources typically represent a census of all activity, such as all calls for service, all officer locations at all times, all officer training records, a complete equipment inventory, and all complaints. Although errors exist, the intention for most police data sources is for them to provide a comprehensive accounting in certain domains of people, places, objects, and actions. Policing has obvious exceptions to this; the police do not receive reports of all crimes, and surveillance data produce an exhaustive collection over a limited domain. License-plate recognition (LPR) systems only collect data on cars passing LPR cameras, acoustic gunshot detectors have wide but finite range, and surveillance cameras collect images only where they are pointed.

Big Data is often relational. For example, technically police could link data on incidents to data on officers at the scenes to the officers' training and performance records. In practice, however, the data might exist but not be in a form readily linkable. The design of most police record systems is tactical and transactional. That is, following a police shooting, the police can retrieve information about that incident and the people involved in the incident. The data may be in the form of scanned documents. This is sensible for a transactional organization in which officers focus on a specific incident. However, this design inhibits achieving Big Data strategic aims. For a study of police shootings in New York City, I requested information on which officers were working the same shift on the same beat at the time of a shooting (Ridgeway 2016). The NYPD had logs of who normally worked those hours but did not have data on which officers appeared for their shift. Local supervisors were aware of which officers were out sick, in court, at training, on another assignment, or working their normal role, but the data were not readily available. When an organization's interest is in a single incident or a single individual, the standard record management systems function well. Police departments, however, need to expand to data designs with a Big Data strategy in mind, which means having the ability to filter and link on data features.

Big Data has fine resolution. Typically, for police this means collecting data on the time and place of incidents more precisely. To study the impact of the expansion of the Los Angeles commuter rail system, I needed data on crime in Los Angeles going back to 1988 (Ridgeway & MacDonald 2016). This meant digitizing 2,300 pages of paper records of crime counts at the Los Angeles Public Library and collecting old maps of the LAPD district geography. The archives had quarterly crime counts at the level of the reporting district. However, by 2000 I could get those data directly from the LAPD with dates, times, and addresses, and, in later years, data with latitude and longitude. Devices now automate timestamps and geolocation of incidents, generating data with greater temporal and spatial resolution.

Lastly, Big Data systems are flexible and scalable. As this section has noted, with greater volume, velocity, and variety, Big Data's role in policing will continue to adapt and adjust. In the remainder of this review, I discuss the pressures that are prompting Big Data and policing and how police are using these new resources.

THE ACCESSIBILITY OF DATA HAS INCREASED

For decades, basic policing functions have depended on volume, velocity, and variety of data. Responding to calls, verifying identities, solving crimes, and reporting performance to the public invariably depend on geographic, photographic, and administrative data. The difference between 1967 and today is that volume, velocity, and variety have all increased by orders of magnitude. Two movements, open data and open source, are accelerating this evolution.

Open Government Has Encouraged Open Data

Freedom of information is a common feature of modern democracies. In the United States, the first Freedom of Information Act became effective five months after the 1967 Commission report. However, awareness of the existence of data, whom to ask for the data, and the cost of providing the data were often barriers. In recent years, the open government movement made freedom of information more proactive on the government side, where governments make freely available any information that a member of the public might wish to have. "There is a new compact on the horizon: information produced by and on behalf of citizens is the lifeblood of the economy and the nation; government has a responsibility to treat that information as a national asset. Citizens are connected like never before and have the skill sets and passion to solve problems affecting them locally as well as nationally" (O'Reilly 2010).

On his first day in office, President Barack Obama issued his "Transparency and Open Government" memo stating "Executive departments and agencies should harness new technologies to put information about their operations and decisions online and readily available to the public" (Obama 2009). Although launched primarily as a resource for federal data, data.gov reports that 40 states and 47 cities offer open data portals as of 2017. Numerous other nations have since created open data portals, including the United Kingdom (2010), France (2011), India (2012), and Russia (2014), with greater pressure coming in the form of international agreements (Castro & Korte 2015) and open data ratings (Open Knowl. Int. 2017).

Data portals prominently feature data on public safety. Two of the top three most frequently accessed data sets from New York City's open data portal relate to public safety (the other one concerns restaurants). Other major cities' portals also reflect the strong public interest in raw, incident level data, not merely the aggregated annual statistics. Police departments have noted numerous benefits of open data on policing, including promoting community analyses, possibly yielding innovations and new insights, improving the community's understanding of what the police do, increasing transparency (Police Found. 2015), and reducing the burden of responding to freedom of information requests.

Arguably one of the most influential policing documents of the past ten years was the Final Report of the President's Task Force on 21st Century Policing (Pres. Task Force 21st Century Polic. 2015). The report included nine recommendations encouraging departments to collect and make available data on numerous aspects of policing, including contacts with the public, composition of the police department, officer safety, and use of force. Prompted by the report, the federal government launched the Police Data Initiative, a plan to accelerate the implementation of the data recommendations from the report. The initiative enlisted the commitment of 50 local police departments to release new data sets and partnered those agencies with the International Association of Chiefs of Police, the Police Foundation, Code for America, and ESRI to facilitate posting the data. As of early 2017, the Public Safety Data Portal features data from 129 agencies (Police Found. 2017).

Computing Tools Have Increased Accessibility

At the same time that data were becoming more widely available inside and outside of police departments, the open source software movement increased the availability of tools to make sense of these data. Unlike in 1967, developers create today's software using standard programming languages, making the software executable on standard computing platforms, and software is often released with licenses so that the tools are free and the underlying source code is public and available for modification and improvement. This has created a vibrant open source ecosystem for Big Data (Marz & Warren 2015). Developers might create these as independent applications, but then other developers can harness their functionality and merge them into a common platform. For example, MongoDB is a popular NoSQL database system for Big Data problems (MongoDB Inc. 2017), Geometry Engine–Open Source (GEOS) is a set of open source tools for working with geographic data (GEOS 2017), and R (R Core Team 2017) and Python (Python Softw. Found. 2017) are scripting languages popular among Big Data analysts. Although these are disparate efforts, R and Python have no built-in capacity for managing databases or for manipulating spatial data. However, developers have created interfaces to harness MongoDB, GEOS, and numerous other open source tools to R and Python. For the analyst, gaining access to these tools is typically a single line of code. I have contributed to this effort, creating a generalized boosted modeling (gbm) package for high-dimensional nonparametric regression that R has harnessed as part of its suite of machine-learning tools (GBM Dev. 2017).

Furthermore, sites like GitHub further promote the rapid development and improvement of new tools. Since I posted the gbm code to GitHub, 20 developers from around the world have enhanced its capability by, for example, incorporating survival models, parallelizing the base algorithm, and increasing the reliability through more exhaustive error checking. This same phenomenon is occurring for thousands of other open source tools for processing video, audio, text, and geographic data. The time from conceptualization to proof-of-concept to deployment of a new Big Data module shrinks in this connected developer environment. Although vendors still play a large role in Big Data at police departments, adept analysts can quickly turn a department's Big Data into an app to further some strategic goal.

Standardization of protocols along with the rapid development of new tools has facilitated the growth and potential of Big Data. For example, Javascript Object Notation (JSON) is a standard data markup language that eases the transfer of data between systems (Ecma Int. 2013). Most of the census data, including geographic data, is readily available through a JSON-based specification. Rather than needing to download all the census data to extract the key variables from particular geographies, a JSON query can hit the US census data center and extract the right numbers from the requested data source and return it in a standardized format. Several departments now make their data available through open data platforms, such as Socrata, a platform for open government to facilitate sharing data with the public through a JSON standard.

The computing tools are available, and the public and political leaders exert pressure to release the data. However, the question remains whether the police or the public can put that data to use.

THE VALUE OF INFORMATION TO POLICE

A symposium reviewing criminal justice system progress 30 years after the 1967 Commission affirmed the Commission's prediction: "Widespread use of computers is the most important development in the criminal justice system of the past 30 years" (Office Justice Progr. 1998). Indeed, by 1997 police were putting data to use, mostly digitization of records and computer-aided mapping. But for all the progress in data, it is unclear whether the core functions of police improved? Despite technological advancement, crime clearance rates have remained stubbornly fixed at approximately 45% for violent crimes and just under 20% for property crimes since the 1967 Commission report (Braga et al. 2011). Homicide clearance rates have declined even. Clearly, technological advances lowered the costs of data collection and storage, and those data are more readily available to police and the public. However, for Big Data to have an impact on crime, crime prevention, and crime clearance, we must figure out how to translate the data into knowledge about crime, better-informed tactics, or improved strategies. We need to assess whether the information gathered has value as the 1967 Commission anticipated.

Value of Information

The year before the 1967 Commission released its report, Howard (1966) laid out a theory of the value of information, describing the thought process and path for computing expected gains with clairvoyance versus uncertain information. In more standard statistical notation, Howard expressed the expected profit, v, given information, ξ, as $E(v|\xi)$. If instead we were able to have clairvoyance about a particular additional piece of information, x, then the expected profit would be a presumably higher value $E(v|C_x, \xi)$. If eliminating uncertainty about x incurs costs (e.g., data collection, data storage, labor costs for a Big Data–savvy criminologist), then those costs make sense only if they are less than the value of clairvoyance about x, $E(v|C_x, \xi) - E(v|\xi)$. That is, the costs to eliminate uncertainty in x must be smaller than the profit achievable by eliminating uncertainty. We can use randomized trials to estimate $E(v|C_x, \xi) - E(v|\xi)$ by randomizing the information offered to experimental units.

The medical community has conducted numerous randomized controlled trials on the value of information. Liu & Wyatt (2011) cite several meta-analyses involving hundreds of randomized control trials (RCTs) that tested whether the access and presentation of additional medical information systems influence practice and outcomes. For example, Goud et al. (2009) conducted an RCT and found that providing electronic medical records along with the associated cardiac rehabilitation guidelines (versus only the electronic medical record) decreased overtreatment (saving expenses on unnecessary treatments) and undertreatment (saving through prevention). Evaluating the value of information systems in law enforcement should follow this same rigorous model.

Compared with medical informatics, there have been few studies looking at the effect of information on police. Garicano & Heaton (2010) studied the link between information technology investments and record keeping, crime rates, clearance rates, and officer safety using data from Law Enforcement Management and Administrative Statistics (LEMAS), the Uniform Crime Reporting (UCR) program, and the US Census. They do find some modest officer safety benefits, but they "seem insufficiently large to justify the high cost of deploying an IT system" (Garicano & Heaton 2010, p. 187), especially when contrasted with cheaper more direct safety measures like body armor and seat belts. They find that general IT investments result in improvements in record keeping but on their own produce no reduction in crime or improvement in clearance rates. However, when IT investments were coupled with adoption of data-driven management processes (e.g., CompStat) ". . .police agencies, like firms, appear likely to enjoy the benefits of computerization only when they identify the specific ways in which new information and data availability interact with existing organizational practices and make adjustments accordingly" (Garicano & Heaton 2010, p. 196).

The following applications of Big Data to policing suggest potential, but the evidence so far is mixed on whether police can effectively use Big Data. This does not mean that Big Data cannot improve policing, only that we have some distance to travel before we can figure out how best to use Big Data to generate impact.

Pushing Real-Time Information to Officers

Perhaps the most obvious Big Data application is the effort to put information on people and places in front of the right officers at the right time. Traditionally, police have had to be proactive about getting information, explicitly making requests either over the radio or through mobile terminals for data. This includes running a license plate or checking for warrants. Instead, systems that push data aim to anticipate the informational needs of an officer even if the officer does not know that they need it. For example, on the way to a call, officers should be made aware of known prior issues at the location, warrants on people associated with the location, orders of protection, and photos of individuals believed to be involved at the location.

Intelligence-led policing has been both a cause and a consequence of Big Data. The demand for intelligence on offenders incentivized the creation of information technology that could feed the demand (Ratcliffe 2002). At the same time, with ready access to information, police have an opportunity to make links and connections that used to be more labor intensive (Ratcliffe 2008).

NYPD's Domain Awareness System (DAS) exemplifies this kind of Big Data policing. The system integrates numerous data sources within the department and makes that information available to frontline officers for tactical purposes as well as to commanders developing crime prevention strategies (Levine et al. 2017). The system includes data from "audio gunshot detectors. . .environmental sensors. . .two billion readings from license plates (with photos), 100 million summonses, 54 million 911 calls, 15 million complaints, 12 million detective reports, 11 million arrests, 2 million warrants, and 30 days of video from 9,000 cameras" (Levine et al. 2017, p. 73). Officers can customize alerts for issues related to their responsibilities, such as 911 calls within their beat, an LPR reading for a witness they seek, or a camera detecting an abandoned package. Through DAS, officers can connect directly with 911 callers rather than work through the dispatcher as an intermediary. The officer can also pull up historical information about the address where they are headed so that they can know in advance about restraining orders, mental health issues, and warrants.

Police will use this information if given access. In a study of the use of information technology in police, Koper et al. (2016) reported that when responding to a call involving a particular address or person as many as 81% of officers in a department pulled up data on the police response history to that address or person in advance. Police appear to be comfortable using the Big Data infrastructure to support their familiar activities, such as responding to calls and making arrests. However, the same research team found that police have not transitioned to a phase of using their data between calls, such as to inform a problem-oriented policing strategy.

For information to be effective, consumers need to know it exists, know where it is located, have access to it, and know how to use it (Moody & Walsh 1999). Taniguchi & Gill (2013) reported on a technology-needs assessment and an RCT of mobile applications for pushing recent crime information down to frontline officers at the Redlands (CA) Police Department. They evaluated CrimeView NearMe, a mobile application that gives officers the ability to filter and view details about cases in their area. Although they found a tech-savvy police force, the largest challenge was that most officers with access to NearMe never bothered to open the application during the study period. Those that used the application felt that it did not offer any additional information beyond what they already knew from experience or daily briefings. Other research on officer use of technology has found similar results (Koper et al. 2015). NYPD tracks usage of DAS in terms of the number of unique officer logins (3,000 unique users per month) and the number of information searches (120,000 per month) to affirm that officers are making use of the data.

Information needs to be new, relevant, timely, and impactful. It is possible that information pushed to frontline officers provides no new insight for those officers. One Redlands police officer noted, "We know where the crime is, because we're taking the reports and making the arrests" (Taniguchi & Gill 2013, p. 69). The challenge for Big Data in policing is to figure out how to provide something previously unknown to the officers, clarify something obscure, or remind them of people and places that they otherwise might neglect. Despite the technologies' weak effect in

Redlands, another officer there noted, "This area also isn't a big crime area… maybe something like this would be useful on the streets of Los Angeles, but it's not necessary here" (Taniguchi & Gill 2013, p. 69). Researchers still need to determine how much information an officer can absorb and how to prioritize that information.

Unlike the Redlands experience, a similar RCT in Lincoln, Nebraska, found the officers accessed and acted on the data more often (Casady et al. 2015). The Lincoln Police Department conducted an RCT of P3i, the predecessor to CrimeView NearMe. They found that those officers randomized to have access to the system had significantly more citation arrests per month, warrant arrests per month, and reports per month than a control group with no access to the system. Officers in Lincoln appeared to engage the technology more intensely than did officers in Redlands, and the results of the RCT indicate that this use likely caused an increase in productivity. When officers can readily see that a technology is pushing actionable information, they might more readily engage with it. For example, in Mesa, Arizona, police officers randomized to use LPR scanned eight times as many license plates as comparison officers who were manually entering license plates and had 2.7 times as many hits on stolen cars (Taylor et al. 2012).

SPECIFIC APPLICATIONS OF BIG DATA IN POLICING

Certainly, Big Data, if used effectively, can improve the fairness and quality of policing. Notable specific applications of Big Data to policing include predictive policing, tracking police locations, recording video, and measuring performance. These applications adeptly illustrate both the immense potential, as well as the challenges, of Big Data for policing.

Predictive Policing

Predictive policing is probably the most anticipated product in the era of Big Data. A 2014 survey of 200 police departments found that predictive policing was in use at 38% of responding departments, but 70% of respondents stated they will be using predictive policing by 2017 (Police Executive Res. Forum 2014). Predictive policing is the use of data in predictive models to anticipate and prevent crime. The predictions are often about times and places at which crime is likely to occur but can also be about who is likely to be the offender or victim. With Big Data, we should be able to track police and offender activity at great resolution and with more precision, and, as a result, we should be able to use modern analytical methods to devise optimal strategies, anticipating where and when police are needed. At the same time, Big Data increases the risk of apophenia, i.e., believing that we see a pattern in a data set when no pattern truly exists (boyd & Crawford 2012).

Translating a prediction into a prevention action remains undeveloped and minimally evaluated (Ridgeway 2013). Even if we assume that the data have predictive value and our statistical models have reasonable precision, what should the police do in response? Police in Chicago developed a strategic suspect list (SSL) to predict the likelihood that an individual would be a homicide victim based on whether they were coarrested in the past five years with someone who became a homicide victim. The prediction model also included data on second-degree coarrests, i.e., whether the individual was coarrested with someone who was coarrested with someone who became a homicide victim. An evaluation of the SSL found that, except for one district, "SSL subjects received relatively little attention" (Saunders et al. 2016, p. 356), to the point that the evaluation team questioned whether to even call it a prevention strategy (Saunders et al. 2016). Analysis of a quasi-experimental design found that being on the SSL did not change the risk of being a homicide or shooting victim, but it did nearly triple their odds of being arrested for a shooting. Even though prior research has found that connections to other shooting victims are associated with becoming a shooting victim (Papachristos et al. 2012), turning that into a prevention strategy seems to have been elusive in Chicago.

Other predictive policing interventions have shown that Big Data, modern statistical methods, and prevention strategies might affect crime. Mohler et al. (2015) put a crime analyst and an epidemic-type aftershock sequence (ETAS) crime forecasting model in a head-to-head prediction competition. During a silent test period in which officers in the field received no predictions, the crime analyst and the ETAS model would select areas of fixed size (typically representing less than 1% of the entire patrol area) forecasted to be likely to contain a large share of future crime. The ETAS model outperformed the crime analyst by a factor of 1.4 to 2.2. Mohler et al. (2015) also conducted an RCT and found that police spending 1,000 minutes in ETAS-predicted hot spots eliminated one crime on average, whereas police would need to spend 2,000 minutes in hot spots generated from LAPD's standard practice to have the same one-crime reduction. This difference is not statistically significant, but signals that some gains, although modest, are possible with predictive policing.

Crime prediction does not need to be policing's only prediction focus. In 2009, I, along with colleagues, worked with the LAPD to predict which police applicants were most likely to successfully complete the recruiting process and join the LAPD (Lim et al. 2009). Some candidates were very likely to join the police force, such as those living in Los Angeles County with some college experience. Other candidates never joined, such as those living outside California and reporting problematic answers on a background questionnaire. We anticipated that LAPD data on recruits should help recruiters identify viable candidates. We decided on principle that we were willing to forego some predictive accuracy if we could produce a predictive model that was simple to understand, explain, and implement. Such sparse linear scoring systems are common in medicine (Ustun & Rudin 2016). We developed a point system in which each new recruit would score a number of points depending on how close to Los Angeles they lived, their education, and the number of issues on their Preliminary Background Assessment. LAPD recruiters simply added up the points scored in each of these dimensions to produce a priority score for each recruit. They then reallocated effort away from those unlikely to ever join and toward those with a higher chance of joining.

Predictive policing remains promising rather than proven. We know that data and prediction models can improve hot-spot prediction, but as with most aspects of Big Data and policing, the open question is how to convert this into something operationally useful with a crime-reduction benefit.

Tracking Police Locations

AVL is a technology that transmits at regular, and typically frequent, intervals the location of a vehicle. Early designs of AVL systems specifically focused on policing applications, noting that AVL systems would provide for "extremely efficient dispatching of mobile police units" (Chisholm 1968, p. 1) and "would permit intelligent redeployment of the forces so as to insure that adequate forces are dispatched to the scene of an emergency without inadvertently depleting large sections of a city of all police protection" (Knickel 1971, p. 1). The first law enforcement use of AVL in the United States in Schaumburg, Illinois began in 1992 (McLaughlin 1994). Almost all new dispatch systems have built-in AVL capabilities, and it is becoming more prevalent. Consistent with the inventors' original image of AVL usage, to date police departments have used AVL largely tactically, for example, dispatching vehicles nearest to a call for service. However, greater strategic opportunities for AVL data are appearing.

Weisburd et al. (2015) used AVL data from the Dallas Police Department to determine how much patrol time was unallocated, when police are available to handle a call, and how much time was assigned to designated hot spots. They concluded that giving commanders information about the quantity of unallocated time and the fraction of that time spent in hot spots helps "commanders to reallocate free time to crime prevention activities via directed patrol" (Weisburd et al. 2015, p. 381). Telep et al. (2014) used AVL data to verify that officers randomly assigned to hot spots were in fact spending the right amount of time in the right hot spots. Collection is not limited to vehicles. Researchers are tracking patrols of individual officers in Bogotá as a part of a broken windows and hot-spot policing intervention (Blattman et al. 2016). They determined that areas responsible for 25% of the city's crime received 10% of patrol's attention. Big Data makes it possible for police executives to measure with greater precision whether police activities in the field reflect their strategies as designed.

Blue force tracking, the general term for systems that monitor locations of security personnel, permits commanders, supervisors, and officers themselves to perceive which areas have police coverage. LAPD, for example, uses blue force tracking at sports stadiums. Supervisors and officers can constantly view each other's location and visualize where officers are clumping and which areas of the stadium have insufficient coverage. They also use the technology to improve officer safety, for example, when sending officers to evacuate homes when the periodic wildfire season comes to Los Angeles.

These examples show that location data can be a strategic asset for making more efficient use of police resources. Current uses of tracking foreshadow the next stages of force monitoring, moving beyond just time and location to what officers are doing. Several police departments have deployed more comprehensive vehicle telematics in their fleets, including the LAPD (Aguilar 2014). The LAPD's 2017–2018 budget proposal ties telematics to three key department objectives (Los Angel. Police Dep. 2016). First, they aim to use the real-time location data to determine whether the planned amount of patrol resources is in targeted hot-spot areas. Second, they expect that deployment of telematics will cause officers to be more thoughtful about their driving behavior, that analysis of driving behavior can identify careless drivers and assign them to a corrective training program, and that they can reduce damage to vehicles and payouts from accidents. Lastly, with LAPD currently using 10,000 gallons of gasoline per day, they expect telematics to result in greater fuel conservation.

Telematics is new to policing, so no evidence has emerged about its value, but findings from other sectors are promising. For example, fuel consumption costs are nearly 6% of UPS's revenue. Researchers noted from telematics data that UPS trucks idling more than 10 seconds consumed more energy and produced more emissions than turning off the vehicle and restarting it (Watson et al. 2010). Pushing that information to drivers changed behavior.

Video

The 1967 Commission report underestimated the importance of video to police, mentioning video as frequently as electronic cocktail olives (only once in passing as a technique to explore). Cameras have become cheaper, more compact, and more easily connected through LTE networks, giving police access to more video sources. The President's Task Force on 21st Century Policing recommended the expansion of video collection, particularly body-worn cameras (Pres. Task Force 21st Century Polic. 2015). Evidence suggests that cameras can have an effect. Body-worn cameras reduce the risk of use of force and the rate of complaints (Ariel et al. 2015). CCTV causes modest crime reduction, although reduction in auto theft and theft from autos primarily drives the decrease (Welsh & Farrington 2009).

The question remains about what to do with all this video. Is the effect of video limited to deterrence based on the fear of being caught on camera? Or can we learn something from the video data? Indeed, those incidents that rise to the attention of public interest become preserved and dissected. However, the thousands of hours produced that receive no attention likely have information on police performance issues about which police executives and the public care.

Going forward, turning video and audio data into information needs advanced analytics. For example, video likely contains information on police performance, police professionalism, and how often police officers fail to comply with department policy. Dixon et al. (2008) report on a labor-intensive study of more than 300 in-car camera videos of traffic stops in Cincinnati, OH. Using human coders, they documented numerous aspects of the traffic stops, including objective measures such as time of the stop and whether the officers conducted a search but also subjective measures such as the politeness of the driver and the police officer. To be of practical usage in the long term, video and audio analytics could conceivably score interactions based on how heated the conversations become. Supervisors could then oversample those interactions that have the markings of a more complex interaction rather than reviewing a random sample of recorded interactions. Regular identification and review of these complex interactions could lead to improved training and strategies for reducing the use of force.

Aside from the use of video in after-action reviews, real-time analytics could eventually be useful to flag incidents as they are happening, detecting shouting commands rather than conversation or detecting foot pursuits, an incident category that often ends with a use of force. Humans are not well-suited for monitoring CCTV cameras for sustained periods of time. Boredom leads to "inattentional blindness" (Näsholm et al. 2014). Computer vision researchers have produced promising methods for detecting aggressive behavior (Gracia et al. 2015), and researchers meet regularly for the International Conference on Imaging for Crime Detection and Prevention. For video to contribute any further to crime reduction and crime clearance, automation of the monitoring of video streams will be essential.

Performance Measurement

As shown in the previous examples in this section, Big Data in policing is only in part about crime fighting. Big Data is equally important to police executives and the public for monitoring the performance of the police. The 1931 Wickersham Commission highlighted the need for data on police activities so that the public and local leaders could have proper oversight of their police (Harmon 2012). Early warning systems are a frequent component of court-ordered police reforms (Ridgeway & MacDonald 2009, Walker 2001) and are a Big Data opportunity. These systems receive streams of data on police activities, with the aim of flagging potential problematic officers as soon as possible. Supervisors can track officers on performance measures, including rates of force, injuries, contact with minorities, and complaints, flagging those with unusual rates.

Accounting for statistical confounding and setting thresholds for unusualness are essential for these systems' effectiveness and legitimacy. For example, Ridgeway & MacDonald (2009) discuss a case in which 86% of a particular officer's stops involved black pedestrians. The time, place, and context of this officer's patrol should inform us whether 86% is potentially reasonable. We can compare this 86% to the percentage of black pedestrians among stops that other officers made in similar circumstances. In practice, police departments have had limited data or limited analytic capability to carefully craft a collection of stops made in similar circumstances. They typically lump officers together by a coarse geography and shift and look for outliers. However, any supervisor with an officer found to be an outlier is left wondering whether observed differences in a performance measure are due to a genuine problem with the officer or to the specific times and places in which the officer was working.

Big Data's higher resolution allows for much finer comparisons. We collected data on the exact date and time (not just the shift), the longitude and latitude (not just district), and other aspects of the contact (e.g., responding to a call for service). Using propensity scores and doubly robust estimation, we precisely customized a set of comparison stops for each officer. We gathered stops other officers made such that the distribution of their features matched the distribution of features for the officer in question. For example, in the case of the officer who stopped black pedestrians in 86% of total stops, black pedestrians were only 55% of stops other officers made at the same time of day, same day of week, same month, and same latitude and longitude. This officer cannot claim that the reason for stopping black pedestrians so frequently had anything to do with time, place, or context. All officers at these times and places should have encountered the same kind of suspicious activities, had similar calls for service and dealt with other similarities tied to time and place. Big Data makes it possible to closely tailor an officer's benchmark to their specific patrol or assignment patterns.

Ridgeway & MacDonald (2014) present a more general approach applicable to many aspects of the justice system. We showed that the same methodology can measure the performance of policing in a neighborhood. Advances in Big Data collection and analysis could lead to a persistent performance measurement system with which police, city officials, and the public could monitor police performance.

BIG DATA AND CONCERNS ABOUT SURVEILLANCE AND DISPARATE IMPACT

Fears of exacerbating racial bias and chipping away at privacy are often by-products of most technological innovations in policing. The 1967 Commission expressed concern about the numerous surveillance methods available at the time and the lack of regulation of their use. They also foreshadowed the concerns with Big Data: "The inherent inefficiencies of manual files containing millions of names have provided a built-in protection. Accessibility will be greatly enhanced by putting the files in a computer; so that the protection afforded by inefficiency will diminish, and special attention must be directed at protecting privacy" (Pres. Comm. Law Enforc. Adm. Justice. 1967, p. 268).

Scholars point out that Big Data in policing might largely overlap with some form of surveillance. Joh (2014, p. 38) notes that "Fourth Amendment doctrines, once adequate for a world of electronic beepers, physical wiretaps, and binocular surveillance" are not adapting at the same speed as technology and that Big Data "may magnify these concerns considerably." Until recently, costs and technical capability limited the degree to which police could synthesize data to address a crime problem or solve a case. When a crime occurs, detectives go on foot to survey the area for cameras that might have caught the perpetrator in transit or in the act. This cost limits the police's ability to have comprehensive photosurveillance, but this will change and video analytics can make sifting through video for "interesting" frames feasible. As of January 2017, the Combined DNA Index System (CODIS) contained nearly 13 million offender profiles and 2.6 million arrestee profiles, almost 5% of the United States population. This represents a massive data set with tremendous power to personally identify individuals. Fears of massive DNA surveillance are premature because of legal barriers and the cost to collect and process DNA evidence, which ranges from $800 to $2,500 (Roman et al. 2009). However, costs of DNA analysis are plummeting faster than Moore's Law or Kryder's Law (Farmera & Lafonda 2016).

Part of the concern over New York City's stop-and-frisk policy was that the NYPD was retaining data on individuals who police stopped but did not arrest, cite, or give any official warning, prompting the New York State Legislature to pass a law applicable only to cities of more than one million people (i.e., New York City) that barred the retention of such records [New York Assem. Bill A11177 (2010); New York Senate Bill S07945 (2010)]. We do not know how long police should retain records before purging them. Some research has suggested that the public might support long-term record retention (Merola et al. 2014). We need research on the time between a data collection and when police realized that they needed the information. Scholars have urged legislatures to pass laws limiting record retention without good information on suitable retention durations (Etzioni & Etzioni 2016).

In addition to surveillance concerns, predictive policing models can imitate existing patterns of discrimination in the data used to create the models (Barocas & Selbst 2016, Lum & Isaac 2016). For example, if a racially biased police force were to fit a new forecasting model to predict from last week's drug arrests the times and places of next week's drug arrests, this would create a feedback loop (Lum & Isaac 2016). If last week's drug arrests were the product of racially biased police activity, that racial bias is now encoded into a prediction model. Those locations with a high predicted likelihood of drug arrests would receive a greater police force allocation. New prediction models would learn from these new drug arrests and the cycle would continue, always reinforcing the original disproportionate allocation of the police force.

The fix is simple. Prediction models aiming to predict community needs do not directly produce feedback loops like models that predict police outputs. Fortunately, the vast majority of predictive policing applications involve police departments aiming to anticipate when and where the public will request police services. For example, predictive analytics built into NYPD's DAS predicts shootings, burglaries, felony assaults, grand larcenies, grand larcenies of motor vehicles, and robberies, none of which would be susceptible to feedback loops (Levine et al. 2017). Anticipating the next bar fight, commercial burglary, or the next shooting only helps the police better position themselves to be responsive to the community.

FUTURE RESEARCH DIRECTIONS

The 1967 Commission's statement that "modern information technology now permits a massive assault on these problems at a level never before conceivable" (Pres. Comm. Law Enforc. Adm. Justice 1967, p. 266) is a timeless assertion. Police have at their disposal technology that can give them eyes and ears where they need them, a real-time view of officer positions and activities, and information on people and places when they need it. Modern programming languages, standardized data formats, and economical storage and computing power have overcome the challenges of the 1967 environment. Everything the 1967 Commission members envisioned is technically possible, but what should police do with it?

Although we have achieved much of the technological aims that the 1967 Commission envisioned, this achievement has not necessarily translated into the imagined improvements. Evidence from RCTs of LPR demonstrates that data on stolen cars and license plates of suspects has value. LPR's efficient processing of license plates improves the police's ability to recover stolen vehicles (Taylor et al. 2012), although it seems to have no deterrent effect (Lum et al. 2010). Pushing data to police can make them more effective if they use it (Casady et al. 2015) but not if they never access it (Taniguchi & Gill 2013). Predictive models can produce better predictions than human crime analysts but have not yet demonstrated a crime prevention benefit (Mohler et al. 2015).

Over the coming decades, criminologists need to rigorously test new information interventions. We are far from figuring out how to tap the crime prevention and crime clearance potential of these data sources. Criminologists have made substantial advancements in studying what officers should do in place-based interventions, such as when data indicate an area as a hot spot (Ariel et al. 2016, Braga & Bond 2008, Eck & Guerette 2012, Ratcliffe et al 2011, Weisburd & Green

1995). For other questions, we have little or no evidence base. For example, what should police do if data indicate a particular person is at high risk of victimization or offending? How can Big Data advise officers how best to use unallocated time? What should police managers do when data signal leading indicators of a problematic officer?

Clearance rates are stubbornly unchanged since 1967 and homicide clearances have even declined. Police may still be underutilizing the technologies most likely to clear a crime, such as cameras and forensic evidence, to produce a crime clearance benefit. Cameras might deter some from offending, but they might make the motivated offender simply wear a hat.

In 2015, the National Institute of Justice solicited research on the value of information, asking researchers to "develop quantifiable measures that reflect the importance of the different kinds of data, offset by the cost to collect, process, analyze, distribute, and retain that data" (Natl. Inst. Justice 2015). The National Institute of Justice correctly identified the present challenge, the one Howard (1966) described, figuring out whether policing's Big Data has value and how to extract that value from our technological advances. The 1967 Commission correctly predicted the widespread adoption of Big Data in policing. They would be pleased that we are no longer in an environment where "1,000 fingerprint clerks at the FBI process about 23,000 such fingerprint records each day" (Pres. Comm. Law Enforc. Adm. Justice 1967, p. 268) and that numerous other policing operations have been digitized and automated. Perhaps they would be underwhelmed that the innovations have not caused equivalently impressive crime declines and crime clearance improvements. Criminologists and technologist collaborators have accomplished much, but surely we can wring more knowledge, insight, and innovations out of these technologies to achieve the Commission's vision.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Discussions with numerous colleagues helped shape this review. I particularly appreciate conversations with Maggie Goodrich, former Chief Information Officer for the Los Angeles Police Department, and Rebecca Neusteter, former Director of Research, Policy, and Planning for the New York Police Department, as well as my colleagues at the University of Pennsylvania, Dr. Nelson Lim, Professor Charles Loeffler, Professor John MacDonald, and Ms. Ruth Moyer.

LITERATURE CITED

Aguilar E. 2014. LAPD to track how safely officers are driving patrol cars. *KPCC*. Dec. 22, **http://www.scpr.org/news/2014/12/22/48820/lapd-to-track-how-safely-officers-are-driving-patr/**

Ariel B, Farrar WA, Sutherland A. 2015. The effect of police body-worn cameras on use of force and citizens' complaints against the police: a randomized controlled trial. *J. Quant. Criminol.* 31(3):509–35

Ariel B, Weinborn C, Sherman LW. 2016. "Soft" policing at hot spots—Do police community support officers work? A randomized controlled trial. *J. Exp. Criminol.* 12(3):277–317

Barocas S, Selbst AD. 2016. Big Data's disparate impact. *Calif. Law Rev.* 104(3):671–732

Blattman C, Green D, Ortega D, Tobon S. 2016. *Pre-Analysis Plan for "The Impact of Hotspot Policing and Municipal Services on Crime: Experimental Evidence from Bogotá"*. Nashville, TN: AEA RCT Regist. **https://www.socialscienceregistry.org/trials/1156/history/10745** boyd d, Crawford K. 2012. Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. *Inform. Commun. Soc.* 15(5):662–79

Braga AA, Bond BJ. 2008. Policing crime and disorder hot spots: a randomized controlled trial. *Criminology* 46:577–608

Braga AA, Flynn EA, Kelling GL, Cole CM. 2011. *Moving the Work of Criminal Investigators Towards Crime Control*. Washington, DC: Natl. Inst. Justice

Casady TK, Cottingham I, Ramírez JP, Samal A, Tomkins AJ, et al. 2015. *A Randomized-Trial Evaluation of a Law Enforcement Application for Smartphones and Laptops that Uses GIS and Location-Based Services' to Pinpoint Persons-of-Interest*. Lincoln, NE: Univ. Neb. **https://www.ncjrs.gov/pdffiles1/nij/grants/248593.pdf**

Castro D, Korte T. 2015. *Open Data in the G8: A Review of Progress on the Open Data Charter*. Washington, DC: Cent. Data Innov. **http://www2.datainnovation.org/2015-open-data-g8.pdf**

Chaiken JM, Dormont P. 1978. A patrol car allocation model: background. *Manag. Sci.* 24(12):1280–90

Chisholm JP. 1968. Mobile emergency unit location system. *US Patent No. 3419865*

Crowther RF. 1964. *The Use of a Computer System for Police Manpower Allocation in St. Louis, Missouri*. Bloomington, IN: Dep. Police Adm.

Davenport TH. 2014. *Big Data @ Work: Dispelling the Myths, Uncovering the Opportunities*. Boston, MA: Harv. Bus. Sch. Publ. Corp.

Dean J. 2014. *Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners*. Hoboken, NJ: Wiley

DHS Cent. Vis. Anal. Command Control Interoper. 2015. *Chicago LTE Video Pilot Final Lessons Learned and Test Report*. West Lafayette, IN: Purdue. **https://www.dhs.gov/sites/default/files/publications/Chicago%20LTE%20Video%20Pilot%20Lessons%20Learned%20Test%20Report.pdf**

Dixon TL, Schell TL, Giles H, Drogos KL. 2008. The influence of race in police-civilian interactions: a content analysis of videotaped interactions taken during Cincinnati police traffic stops. *J. Commun.* 58:530–49

Doumi T, Dolan MF, Tatesh S, Casati A, Tsirtsis G, et al. 2013. LTE for public safety networks. *IEEE Commun. Mag.* 51(2):106–12

Eck JE, Guerette RT. 2012. Place-based crime prevention: theory, evidence, and policy. In *The Oxford Handbook of Crime Prevention*, ed. DP Farrington, BC Welsh, pp. 354–83. New York: Oxford Univ. Press

Ecma Int. 2013. *The JSON Data Interchange Format*. Geneva: Ecma Int.

Etzioni A, Etzioni O. 2016. Keeping AI legal. *Vanderbilt J. Entertain. Technol. Law* 19(1):133–46

Farmera JD, Lafonda F. 2016. How predictable is technological progress? *Res. Policy* 45(3):647–65

Garicano L, Heaton P. 2010. Information technology, organization, and productivity in the public sector: evidence from police departments. *J. Labor Econ.* 28(1):167–201

GBM Dev. 2017. *Generalized Boosted Modeling* (source code). **https://github.com/gbm-developers/gbm**

GEOS. 2017. *Geometry Engine Open Source*. **https://trac.osgeo.org/geos/**

Goud R, de Keizer NF, ter Riet G, Wyatt JC, Hasman A, et al. 2009. Effect of guideline based computerised decision support on decision making of multidisciplinary teams: cluster randomised trial in cardiac rehabilitation. *BMJ* 338:b1440

Gracia IS, Suarez OD, Garcia GB, Kim T-K. 2015. Fast fight detection. *PLOS ONE* 10(4):e0120448

Harmon R. 2012. Why do we (still) lack data on policing? *Marquette Law Rev.* 96(4):1119–46

Hollywood JS, Winkelman Z. 2015. *Improving Information-Sharing Across Law Enforcement: Why Can't We Know?* Santa Monica, CA: RAND Corp.

Howard RA. 1966. Information value theory. *IEEE Trans. Syst. Sci. Cybern.* 2(1):22–26

Int. Assoc. Chiefs Police. 2015. *Guiding Principles on Cloud Computing in Law Enforcement*. Alexandria, VA: Int. Assoc. Chiefs Police. **http://www.theiacp.org/portals/0/documents/pdfs/cloudcomputingprinciples.pdf**

Ioimo R. 2016. *Introduction to Criminal Justice Information Systems*. Boca Raton, FL: CRC Press

Joh EE. 2014. Policing by numbers: Big Data and the Fourth Amendment. *Washington Law Rev.* 89(1):35–68

Kitchin R. 2014. The real-time city? Big Data and smart urbanism. *GeoJournal* 79(1):1–14

Knickel ER. 1971. Vehicle locator system. *US Patent No. 3568161*

Koper CS, Lum C, Hibdon J. 2016. The uses and impacts of mobile computing technology in hot spots policing. *Eval. Rev.* 39(6):587–624

Koper CS, Lum C, Willis JJ, Woods DJ, Hibdon J. 2015. *Realizing the Potential of Technology in Policing: A Multisite Study of the Social, Organizational, and Behavioral Aspects of Implementing Policing Technologies*. Fairfax, VA: Cent. Evid.-Based Crime Policy George Mason Univ. Police Executive Res. Forum

Laney D. 2001. 3D data management: controlling data volume, velocity, and variety. *Appl. Deliv. Strateg.* 949:1–3. **https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf**

Levine ES, Tisch J, Tasso A, Joy M. 2017. The New York City Police Department's domain awareness system. *Interfaces* 47(1):70–84

Lim N, Matthies C, Ridgeway G, Gifford B. 2009. *To Protect and to Serve: Enhancing the Efficiency of LAPD Recruiting*. Santa Monica, CA: RAND Corp.

Liu JL, Wyatt JC. 2011. The case for randomized controlled trials to assess the impact of clinical information systems. *J. Am. Med. Inform. Assoc.* 18(2):173–80

Los Angel. Police Dep. 2016. *2017–18 Fiscal Year Proposed Budget*. Los Angeles, CA: City Los Angel. **http://www.lapdpolicecom.lacity.org/112216/BPC_16-0417.pdf**

Lum C, Merola L, Willis J, Cave B. 2010. *License Plate Recognition Technology (LPR): Impact Evaluation and Community Assessment, Final Report*. Fairfax, VA: George Mason Univ Cent. Evid.-Based Crime Policy

Lum K, Isaac W. 2016. To predict and serve? *Significance* 13(5):14–19

Marz N, Warren J. 2015. *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. Shelter Island, NY: Manning Publ.

McLaughlin M. 1994. *Automatic vehicle location for law enforcement*. SEARCH Tech. Bull. 3, SEARCH, Sacramento, CA.

Meriwether K. 2014. Privacy, storage questions loom as NYPD begins camera pilot. *Gotham Gaz.*, Dec. 3. **http://www.gothamgazette.com/index.php/government/5455-privacy-storage-questions-loom-as-nypd-begins-camera-pilot**

Merola LM, Lum C, Cave B, Hibdon J. 2014. Community support for license plate recognition. *Polic. Int. J. Police Strateg. Manag.* 37(1):30–51

Mohler GO, Short MB, Malinowski S, Johnson M, Tita GE, et al. 2015. Randomized controlled field trials of predictive policing. *J. Am. Stat. Assoc.* 110(512):1399–411

MongoDB Inc. 2017. *MongoDB*. **https://github.com/mongodb**

Moody DL, Walsh P. 1999. Measuring the value of information: an asset valuation approach. *Proc. Eur. Conf. Inf. Syst., 7th, Copenhagen*, June 23–25, pp. 496–512. Guimarães, Port.: ECIS

Näsholm E, Rohlfing S, Sauer JD. 2014. Pirate stealth or inattentional blindness? The effects of target relevance and sustained attention on security monitoring for experienced and naïve operators. *PLOS ONE* 9(1):e86157

Natl. Inst. Justice. 2015. *Developing a Method to Valuate Law Enforcement Data*. Washington, DC: Natl. Inst. Justice. **https://www.ncjrs.gov/pdffiles1/nij/sl001135.pdf**

Obama B. 2009. *Transparency and Open Government: Memorandum for the Heads of Executive Departments and Agencies*. Washington, DC: White House. **https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government**

Office Justice Progr. 1998. *The Challenge of Crime in a Free Society: Looking Back…Looking Forward: An Overview for the Symposium on the 30th Anniversary of the President's Commission on Law Enforcement and Administration of Justice*. Washington, DC: US Dep. Justice

Open Knowl. Int. 2017. *Global Open Data Index*. **http://index.okfn.org/**

O'Reilly T. 2010. Government as a platform. In *Open Government: Collaboration, Transparency, and Participation in Practice*, ed. D Lathrop, L Ruma, pp. 11–39. Sebastopol, CA: O'Reilly Media

Papachristos AV, Braga AA, Hureau DM. 2012. Social networks and the risk of gunshot injury. *J. Urban Health Bull. N. Y. Acad. Med.* 89(6):992–1003

Police Executive Res. Forum. 2012. *How Are Innovations in Technology Transforming Policing?* Washington, DC: Police Executive Res. Forum. **http://www.policeforum.org/assets/docs/Critical_Issues_Series/how%20are%20innovations%20in%20technology%20transforming%20policing%202012.pdf**

Police Executive Res. Forum. 2014. *Future Trends in Policing*. Washington DC: Police Executive Res. Forum

Police Found. 2015. *5 Things You Need to Know About Open Data in Policing*. Washington, DC: Police Found.

Police Found. 2017. *Police Data Initiative*. Washington, DC: Police Found. **https://publicsafetydataportal.org/**

Ponemon L, Gould J, Roberts DJ. 2013. *Leveraging the Cloud for Law Enforcement*. Alexandria, VA: Int. Assoc. Chiefs Police. **https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=263661**

Pres. Task Force 21st Century Polic. 2015. *The Final Report of the President's Task Force on 21st Century Policing*. Washington, DC: Office Community Oriented Polic. Serv.

Pres. Comm. Law Enforc. Adm. Justice. 1967. *The Challenge of Crime in a Free Society*. Washington, DC: Gov. Print. Office

Python Softw. Found. 2017. *Python Language Reference, version 3.6.1.* **http://www.python.org**

R Core Team. 2017. R: A language and environment for statistical computing. Vienna: R Found. Stat. Comput. **https://www.r-project.org/**

Ratcliffe JH. 2002. Intelligence-led policing and the problems of turning rhetoric into practice. *Polic. Soc.* 12(1):53–66

Ratcliffe JH. 2008. *Intelligence-Led Policing*. Cullompton, UK: Willan Publ.

Ratcliffe JH, Taniguchi T, Groff ER, Wood JD. 2011. The Philadelphia Foot Patrol Experiment: a randomized controlled trial of police patrol effectiveness in violent crime hotspots. *Criminology* 49(3):795–831

Ridgeway G. 2013. Linking prediction and prevention. *Criminol. Public Policy* 12(3):545–50

Ridgeway G. 2016. Officer risk factors associated with police shootings: a matched case–control study. *Stat. Public Policy* 3(1):1–6

Ridgeway G, MacDonald JM. 2009. Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *J. Am. Stat. Assoc.* 104(486):661–68

Ridgeway G, MacDonald JM. 2014. A method for internal benchmarking of criminal justice system performance. *Crime Delinq*. 60(1):145–62

Ridgeway G, MacDonald JM. 2016. Effect of rail transit on crime: a study of Los Angeles from 1988–2014. *J. Quant. Criminol.* 33(2):277–91

Roman JK, Reid SE, Chalfin AJ, Knight CR. 2009. The DNA field experiment: a randomized trial of the cost-effectiveness of using DNA to solve property crimes. *J. Exp. Criminol.* 5:345–69

Saunders J, Hunt P, Hollywood JS. 2016. Predictions put into practice: a quasi-experimental evaluation of Chicago's predictive policing pilot. *J. Exp. Criminol.* 12(3):347–71

Taniguchi T, Gill C. 2013. *The Mobilization of Crime Mapping and Intelligence Gathering: Evaluating Smartphone Deployment & Custom App Development in a Mid-Size Law Enforcement Agency*. Washington, DC: Police Found. **http://www.policefoundation.org/content/mobilization-crime-mapping**

Taylor B, Koper C, Woods D. 2012. Combating vehicle theft in Arizona: a randomized experiment with license plate recognition technology. *Crim. Justice Rev.* 37(1):24–50

Telep CW, Mitchell RJ, Weisburd D. 2014. How much time should the police spend at crime hot spots? Answers from a police agency directed randomized field trial in Sacramento, California. *Justice Q*. 31(5): 905–33

Ustun B, Rudin C. 2016. Supersparse linear integer models for optimized medical scoring systems. *Mach. Learn.* 102(3):349–91

Walker S. 2001. Searching for the denominator: problems with police traffic stop data and an early warning system solution. *Justice Res. Policy* 3(2):63–95

Walter C. 2005. Kryder's law. *Sci. Am.*, Aug. 1. **https://www.scientificamerican.com/article/kryders-law/#**

Watson RT, Boudreau M-C, Li SS, Levis J. 2010. Telematics at UPS: en route to energy informatics. *MIS Q. Exec.* 9(1):1–11

Weisburd D, Green L. 1995. Policing drug hot spots: the Jersey City drug market analysis experiment. *Justice Q*. 12(4):711–35

Weisburd D, Groff ER, Jones G, Cave B, Amendola KL, et al. 2015. The Dallas patrol management experiment: Can AVL technologies be used to harness unallocated patrol time for crime prevention? *J. Exp. Criminol.* 11(3):367–91

Welsh BC, Farrington DP. 2009. Public area CCTV and crime prevention: an updated systematic review and meta-analysis. *Justice Q*. 26(4):716–45