NEW MEXICO LAW REVIEW

Volume 53
Issue 2 *Summer*

Summer 2023

Big Data Policing Capacity Measurement

Ronald J. Coleman

BIG DATA POLICING CAPACITY MEASUREMENT

Ronald J. Coleman[*]

ABSTRACT

Big data, algorithms, and computing technologies are revolutionizing policing. Cell phone data. Transportation data. Purchasing data. Social media and internet data. Facial recognition and biometric data. Use of these and other forms of data to investigate, and even predict, criminal activity is law enforcement's presumptive future. Indeed, law enforcement in several major cities have already begun to develop a big data policing mindset, and new forms of data have played a central role in high-profile matters featured in the Serial and To Live and Die in LA podcasts, as well as in the Supreme Court's leading privacy and criminal procedure case of Carpenter v. United States. Although the ascendancy of big data policing appears inevitable, important empirical questions on local law enforcement agency capacity remain insufficiently answered. For example, do agencies have adequate capacity to facilitate big data policing? If not, how can policymakers best target resources to address capacity shortfalls? Are certain categories of agencies in a comparatively stronger position in terms of capacity? Answering questions such as these requires empirical measurement of phenomena that are notoriously difficult to measure. This Article presents a novel, multidimensional measure of big data policing capacity in U.S. local law enforcement agencies: the Big Data Policing Capacity Index ("BDPCI"). Analysis of the BDPCI provides three principal contributions. First, it offers an overall summary of more than 2,000 local agencies' inadequacy in big data policing capacity using a large-N dataset. Second, it identifies factors that are driving lack of capacity in agencies. Third, it illustrates how differences between groups of Agencies might be analyzed based on size and location, including an illustrative ranking of the fifty U.S. states. This Article is meant to inform stakeholders on agencies' current positions, advise on how best to improve such positions, and drive further research into empirical measurement and big data policing.

[*] Adjunct Professor of Law, Georgetown University Law Center.

INTRODUCTION

Imagine DEA Agent Hank Schrader of *Better Call Saul* and *Breaking Bad* fame suspects that the notorious drug kingpin, Heisenberg, is in fact Schrader's brother-in-law, Walter White.[1] Heisenberg is famous for producing a special brand of methamphetamine known as "blue meth" and is the only one who can successfully produce it. The question for the police in New Mexico tipped off by Schrader: how to go about investigating White further?

The police could choose to take a more traditional approach and, for instance, stake out White's home, search for physical evidence at crime scenes, and question known witnesses. What if, however, they take a different approach and focus on algorithmic and big data techniques to investigate White.[2] Suppose they utilize, among other things, large law enforcement databases, predictive analytics, facial recognition software, biometric technology, phone metadata, and data from social media, websites, and third-party information aggregators.[3] The cell phone data might illuminate White's co-conspirators and place White at the scene of relevant criminal activities.[4] White's presence at such crime scenes might be supported by data from speed cameras, electronic toll records, automated readers of license plates, GPS, and surveillance devices on public transport.[5] Purchasing data from a third-party aggregator or online shopping site might reveal that White or his associates bought the components and items necessary to cook blue meth.[6] Data from biometric identification technology—such as iris, scar, or tattoo information—and facial recognition software might suggest White is Heisenberg, as might activities by White and others on social media or the internet.[7] Searches in datasets from law enforcement or private surveillance collection services might also point to White.[8] In fact, armed with sufficient data and technology, the police might even have been able to stop White or his co-conspirators from becoming involved in criminal activity before they began.[9]

[1] *See Better Call Saul* (Sony Pictures Television Feb. 8, 2015); *Breaking Bad* (Sony Pictures Television Jan. 20, 2008).

[2] *See infra* Part I.A.

[3] *See* Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 330, 353–76 (2015) [hereinafter *Predictive Reasonable Suspicion*].

[4] *Id.* at 355–57; Eric Pait, *Find My Suspect: Tracking People in the Age of Cell Phones*, 2 GEO. L. TECH. REV. 155, 156 (2017).

[5] *Predictive Reasonable Suspicion*, *supra* note 3, at 357.

[6] *Id.* at 357–60, 372; Chad Squitieri, Note, *Confronting Big Data: Applying the Confrontation Clause to Government Data Collection*, 101 VA. L. REV. 2011, 2017–19, 2024–31 (2015).

[7] *Predictive Reasonable Suspicion*, *supra* note 3, at 357–65; Squitieri, *supra* note 6, at 2017–19, 2024–31; *see also* Jeremy Greenberg, Abstract, *What is a Face? Creating a Skin Texture Model for Facial Recognition*, 1 GEO. L. TECH. REV. 214, 214 (2017) ("Facial recognition technology . . . is the general term for a complex series of programs, the aim of which is to have computers recognize and compare images of faces to determine if they match.").

[8] *Predictive Reasonable Suspicion*, *supra* note 3, at 358–65.

[9] *See, e.g.*, *id.* at 351; Andrew Guthrie Ferguson, *The Legal Risks of Big Data Policing*, 33 ABA CRIM. JUST. MAG. 4, 5 (2018) [hereinafter *Legal Risks of Big Data Policing*].

Such big data policing is in ascendancy, and it is the presumptive future of law enforcement.[10] New forms of data have already played a leading role in connection with several high-profile matters, including the homicide case against Adnan Syed made famous by the *Serial* podcast, the disappearance of aspiring model and actress Adea Shabani featured in the *To Live and Die in LA* podcast, and the leading Supreme Court privacy and criminal procedure case of *Carpenter v. United States.*[11] Police administrators have experimented with new surveillance technology and sought out partnerships with private data companies.[12] The big data policing mindset has also been in development in major cities such as New York City, Los Angeles, and Chicago, among others.[13]

Although the ascendancy of big data policing appears inevitable and, despite a large degree of interest in the area, important empirical questions on local law enforcement agency capacity remain insufficiently answered.[14] For example, do agencies have adequate capacity to facilitate big data policing? If not, how can policymakers best target resources to address capacity shortfalls? Are certain categories of agencies in a comparatively stronger position in terms of capacity?[15] Answering questions such as these requires empirical measurement of phenomena that are notoriously difficult to measure.

This Article presents a novel multidimensional measure of big data policing capacity in U.S. local law enforcement agencies: the Big Data Policing Capacity Index ("BDPCI"). Analysis of the BDPCI provides three principal contributions. First, it offers an overall summary of more than 2,000 local agencies' inadequacy in big data policing capacity using a large-N dataset. Second, it identifies factors that are driving lack of capacity in agencies.[16] Third, it illustrates how differences

[10] *See generally Predictive Reasonable Suspicion*, *supra* note 3; *see also Legal Risks of Big Data Policing*, *supra* note 9, at 4–6 (suggesting that "[t]he future of law enforcement is being shaped by new technologies," but stating "these big data policing technologies are in their early stages"). It is also, of course, possible that political or societal pressure could eventually constrain use of big data policing. *See infra* Part II.C.

[11] *See, e.g.*, Jon Swaine, *Serial's Adnan Syed: Doubts Over Cellphone Evidence Central to Retrial*, THE GUARDIAN, https://www.theguardian.com/tv-and-radio/2016/jul/01/serial-adnan-syed-new-trial-hae-min-lee-murder (July 14, 2017, 3:08 PM) [https://perma.cc/BDS2-LKM3]; To Live and Die in LA, *Episode 8: Adea*, *Season 1*, TENDERFOOT TV, (Apr. 5, 2019), https://podcasts.apple.com/us/podcast/adea-8/id1453788965?i=1000434198300 [https://perma.cc/XND9-HLSY] [hereinafter To Live and Die in LA]; *Carpenter v. United States*, 138 S.Ct. 2206, 2211–14 (2018).

[12] Andrew Guthrie Ferguson*, Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1137–43 (2017) [hereinafter *Policing Predictive Policing*]; Andrew Guthrie Ferguson, *Illuminating Black Data Policing*, 15 OHIO ST. J. CRIM. L. 503, 503 (2018) [hereinafter *Black Data Policing*].

[13] *Black Data Policing*, *supra* note 12, at 503; *Legal Risks of Big Data Policing*, *supra* note 9, at 5–6.

[14] *See, e.g.*, ANDREW GUTHRIE FERGUSON, THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT 6 (2017) [hereinafter RISE OF BIG DATA POLICING] ("The big data policing revolution has arrived."); *see infra* Part I; *see also* Aleš Završnik, *Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings*, 18 EUR. J. CRIMINOLOGY 623, 625–27 (2021) (discussing certain prior research); Sarah Brayne, *Big Data Surveillance: The Case of Policing*, 82 AM. SOC. REV. 977, 982 (2017) (same).

[15] *See infra* Parts II.C and III. This Article generally uses "capacity" in the sense of "infrastructure" or "capability." For further discussion, *see infra* Parts II and III.

[16] For instance, with knowledge of such drivers, a policymaker seeking to decrease inadequacy might encourage investment in areas of particular concern, such as perhaps in connection with websites. *See infra* Part III.

between groups of Agencies might be analyzed based on size and location, including an illustrative ranking of the fifty U.S. states.[17]

The remainder of this Article will proceed as follows. Part I will provide background on big data policing. Part II will discuss the methodology employed in this Article and the construction of the BDPCI. Part III will present empirical findings from the BDPCI, including consideration of limitations and sensitivity. This Article is meant to inform stakeholders on agencies' current positions, advise on how best to improve such positions, and drive further research into empirical measurement and big data policing.[18]

I. BACKGROUND ON BIG DATA POLICING

Big data, algorithms, and computing technologies are revolutionizing policing.[19] Cell phone data.[20] Transportation data.[21] Purchasing data.[22] Social media and internet data.[23] Facial recognition and biometric data.[24] Data from databases and

[17] For instance, with knowledge of such differences, a policymaker seeking to decrease inadequacy in Georgia or Texas might make different decisions regarding cameras as one in New York would. *See infra* Part III.

[18] Importantly, this Article takes no position as to the merits or desirability of big data policing. Instead, the focus in this Article is on measuring big data policing capacity and illustrating potential means of increasing capacity if policymakers were to determine that increasing such capacity is desirable.

[19] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 2–3 ("Big data technologies and predictive analytics will revolutionize policing. . . . Behind the data is technology: algorithms, network analysis, data mining, machine learning, and a host of computer technologies being refined and improved every day."). The phrase "big data policing" used throughout this Article is adopted from the work of Andrew Guthrie Ferguson. *See generally id.*; *Predictive Reasonable Suspicion, supra* note 3, at 350; *see also* Kiel Brennan-Marquez, *Big Data Policing and the Redistribution of Anxiety*, 15 OHIO ST. J. CRIM. L. 487, 487 n.2 (2018) ("Here, and throughout the essay, I am adopting Andrew Ferguson's phrase (and drawing inspiration from his work).").

[20] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 11; Pait, *supra* note 4, at 155–56; Stephanie K. Pell & Christopher Soghoian, *A Lot More Than a Pen Register, and Less Than a Wiretap: What the StingRay Teaches Us About How Congress Should Approach the Reform of Law Enforcement Surveillance Authorities*, 16 YALE J.L. & TECH. 134, 134–48 (2013); Paul Ohm, *The Many Revolutions of Carpenter*, 32 HARV. J.L. & TECH. 357, 361–66 (2019); Cal Cumpstone, *Game of Phones: The Fourth Amendment Implications of Real-Time Cell Phone Tracking*, 65 CLEV. ST. L. REV. 77, 84–86 (2016); *Predictive Reasonable Suspicion*, *supra* note 3, at 355–56; *Legal Risks of Big Data Policing*, *supra* note 9, at 6.

[21] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 11; *Legal Risks of Big Data Policing*, *supra* note 9, at 5–6; *Predictive Reasonable Suspicion*, *supra* note 3, at 357.

[22] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 10–11; *Legal Risks of Big Data Policing*, *supra* note 9, at 6.

[23] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 10; Brian Mund, *Social Media Searches and the Reasonable Expectation of Privacy*, 19 YALE J. L. & TECH. 238, 240–58 (2017); Christopher L. Izant, *Equal Access to Public Communications Data for Social Media Surveillance Software*, 31 HARV. J.L. & TECH. 237, 237–44 (2017).

[24] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 10–11; *Predictive Reasonable Suspicion*, *supra* note 3, at 365; Greenberg, *supra* note 7, at 215; Joseph Clarke Celentino, *Face-To-Face with Facial Recognition Evidence: Admissibility Under the Post-Crawford Confrontation Clause*, 114 MICH. L. REV. 1317, 1342–51 (2016); *Fingerprints and Other Biometrics*, FBI, https://www.fbi.gov/services/cjis/fingerprints-and-other-biometrics [https://perma.cc/64S3-HRSV].

other sources.[25] These and other forms of data may be collected, analyzed, and used to arrest and convict suspected criminals or to predict future criminal activity.[26]

A. Defining Big Data Policing

Although there does not appear to be a single, agreed upon definition of "big data," it may refer to "the accumulation and analysis of unusually large datasets."[27] Big data is a shorthand term "that typically means applying the tools of artificial intelligence, like machine learning, to vast new troves of data beyond that captured in standard databases."[28] Big data has become intelligible due to algorithms and large-scale computing power.[29] Powerful computers are now used to sort data and reveal unexpected correlations, with machine learning and predictive analytics permitting educated guesses regarding the meaning of such correlations.[30]

Multiple aspects of modern life may be impacted by big data, algorithms, or machine learning. Algorithms may be used by a social media company to analyze accumulated data, predict what content a user might find interesting, and populate that user's feed with such interesting content.[31] Similarly, algorithms might be used by a video streaming service to recommend programs for future viewing or by a search engine to understand what individuals wish to know.[32] Data analysis might allow a retailer to anticipate what items to stock in specific locations at certain times.[33] Big data and machine learning may be employed in the healthcare industry

[25] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 12–14; *Predictive Reasonable Suspicion*, *supra* note 3, at 330.

[26] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 1–6; *Legal Risks of Big Data Policing*, *supra* note 9, at 5–6. A fulsome exploration of all aspects of big data policing is not possible in the context of this Article. Instead, this Article seeks to provide a very brief introduction as background for its subsequent arguments.

[27] *See Predictive Reasonable Suspicion*, *supra* note 3, at 352 ("[Big data] provides a shorthand term for data collection in a variety of industries and settings."); RISE OF BIG DATA POLICING, *supra* note 14, at 8.

[28] Elizabeth E. Joh, *Policing By Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35, 38, 38 n.24 (2014) [hereinafter *Policing By Numbers*] (citation omitted).

[29] RISE OF BIG DATA POLICING, *supra* note 14, at 18. Algorithm has been defined in various ways, including as a "mathematical or logical procedure for solving a problem" or "a sequence of instructions telling a computer what to do." Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 84 n.1 (2017) (citation omitted); *see also* Kevin Emerson Collins, *The Williamson Revolution in Software's Structure*, 31 BERKELEY TECH. L.J. 1597, 1619 (2016) ("[A]n algorithm is a sequence of steps for performing a task."); Sang Ah Kim, *Social Media Algorithms: Why You See What You See*, 2 GEO. L. TECH. REV. 147, 149 (2017) ("An algorithm is a fancy way to describe a set of steps to reach a goal."); WENDY LEE ET AL., NATIONAL ASSOCIATION OF CRIMINAL DEFENSE LAWYERS, GARBAGE IN, GOSPEL OUT: HOW DATA-DRIVEN POLICING TECHNOLOGIES ENTRENCH HISTORIC RACISM AND 'TECH-WASH' BIAS IN THE CRIMINAL LEGAL SYSTEM 26 (2021) ("[A]n algorithm can broadly be defined as a 'specified sequence of logical operations that provides step-by-step instructions for computers to act on data and thus automate decisions.'").

[30] RISE OF BIG DATA POLICING, *supra* note 14, at 8.

[31] *See* Kim, *supra* note 29, at 149–50.

[32] *See How Netflix's Recommendations System Works*, NETFLIX, https://help.netflix.com/en/node/100639 [https://perma.cc/2QG9-ZJTY]; Andrew L. Beam & Isaac S. Kohane, *Big Data and Machine Learning in Health Care*, 319 J. AM. MED. ASS'N 1317, 1317–18 (2018).

[33] *See, e.g.*, Beth Pearsall, *Predictive Policing: The Future of Law Enforcement?*, 266 NAT'L INST. OF JUST. J. 16, 16 (2010) ("Walmart . . . learned through analysis that when a major weather event is in

with some suggestion that "big data and machine learning can create algorithms that perform on par with human physicians."[34]

Big data is also being used in policing and law enforcement.[35] Algorithms are now being utilized to allocate police resources, notify the police of individuals who may be dangerous, guide attempts to intervene before individuals commit crimes, and impact judicial determinations.[36] "Big data policing," as used in this Article, generally refers to the use of large datasets, algorithms, computing, and related technology in policing and law enforcement.[37]

B. Ascendancy of Big Data Policing

Data has long been central to criminal justice.[38] From early attempts to measure faces, ears, and heads of criminal suspects, to more modern attempts to secure arrestees' DNA, governments have sought to collect data on individuals believed to pose criminal risk.[39] Using big data for criminal justice decision-making is also nothing new.[40] Statistical prediction of recidivism, for instance, may be fairly traced back to at least 1928, when Ernest Burgess designed his parole prediction instrument.[41] Despite this longer history, law enforcement practices have

the forecast, demand for three items rises: duct tape, bottled water and strawberry Pop-Tarts. Armed with this information, stores in the affected areas can ensure their shelves are fully stocked to meet customer needs.").

[34] Beam & Kohane, *supra* note 32, at 1317; *see also* C. Jason Wang, Chun Y. Ng & Robert H. Brook, *Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing*, 323 J. AM. MED. ASS'N 1341, 1341–42 (2020).

[35] *See generally* RISE OF BIG DATA POLICING, *supra* note 14; *see also Legal Risks of Big Data Policing*, *supra* note 9, at 5–6 (discussing predictive policing in Chicago and Manhattan); Elizabeth E. Joh, *The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing*, 10 HARV. L. & POL'Y REV. 15 (2016) [hereinafter *The New Surveillance Discretion*].

[36] Ric Simmons, *Big Data and Procedural Justice: Legitimizing Algorithms in the Criminal Justice System*, 15 OHIO ST. J. CRIM. L. 573, 573 (2018).

[37] *See generally* RISE OF BIG DATA POLICING, *supra* note 14; *see generally Predictive Reasonable Suspicion*, *supra* note 3.

[38] Wayne A. Logan & Andrew Guthrie Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541, 542–49 (2016) ("Warehouses of court files and aging police file cabinets, dating back decades, attest to the practice of assembling vast amounts of detailed personal information."); LEE ET AL., *supra* note 29, at 24 (noting "[e]arly adopters of criminal statistics collected data on indictments, convictions, and acquittals using court records, but this was not a standardized practice across the [United States], with the exception of a few states").

[39] Logan & Ferguson, *supra* note 38, at 542 ("[Governments] have also generated data, recording arrests, issuing warrants, and even creating publicly available lists of individuals thought to raise safety concern. A prime example of the latter is the current profusion of government-created registries targeting specific sub-populations, most notably convicted sex offenders but increasingly others as well.").

[40] Brayne, *supra* note 14, at 981.

[41] J.C. Oleson, *Training to See Risk: Measuring the Accuracy of Clinical and Actuarial Risk Assessments Among Federal Probation Officers*, 75 FED. PROBATION 52, 52 (2011) (noting that the "statistical prediction of recidivism risk has an 80-year history" and that "[e]arly attempts to use actuarial risk assessment in the justice system were often controversial, particularly given high rates of false positives"); Brayne, *supra* note 14, at 981 (noting "actuarial methods have existed in corrections and the courts for

almost a century"); *Policing Predictive Policing*, *supra* note 12, at 1117–18 ("Early adopters such as Ernest Burgess looked at individual risk factors to predict the likelihood of convicted parolees reoffending."). It might also be noted that, starting in the 1970s, sentencing guidelines may have helped embed quantification into legal practices, and it has also been suggested that the last several decades have systemically incorporated data-driven decision-making only more recently.[42] Modern data-driven policing programs may have emerged from repeated attempts by legal scholars, law enforcement, and criminologists to measure and quantify the complicated social processes behind disorder and crime.[43]

Big data policing's ascendancy has arguably resulted from various factors.[44] Such factors include an increase in the volume of collected data,[45] increased capability to connect data networks,[46] increased analytic capabilities due to faster computer processors and greater storage capacity,[47] law enforcement budget pressures,[48] and the desire for objective accountability of law enforcement actions.[49] Currently, big data policing has advanced to the point where it may be considered helpful in: (1) investigating and prosecuting crimes, as well as in (2) predicting future criminal activity.[50] Subsections 1 and 2, below, will treat each of these uses in turn.

seen a shift to "actuarial justice"—whereby actors are employing criteria drawn from risk management to make estimations of criminal risk probabilities. *See* Brayne, *supra* note 14, at 981; *see also Policing Predictive Policing*, *supra* note 12, at 1189 ("Police have entered the age of actuarial justice and, as demonstrated, there is no real hope of going back. The technology exists, is adapting, and is pushing much farther ahead than lawyers, courts, and policymakers.").

[42] *See* Brayne, *supra* note 14, at 981; *see also* LEE ET AL., *supra* note 29, at 26 ("Recent technological advances have since transformed the function and organizational structure of police departments. Between 1990 and 2003, the use of computers by law enforcement personnel skyrocketed from 5% to 56%, and by the early 2000s, the 'vast majority of the nation's police agencies [were] using computerized data systems to monitor the activities of their officers (arrests, citations, calls for service, etc.).'") (citation omitted).

[43] LEE ET AL., *supra* note 29, at 24.

[44] *See Predictive Reasonable Suspicion, supra* note 3, at 353–65; RISE OF BIG DATA POLICING, *supra* note 14, at 4–6, 20–33.

[45] *Predictive Reasonable Suspicion, supra* note 3, at 354 (noting the volume of data collected is "growing exponentially [and] . . . doubling in volume every two years").

[46] *Id.* at 353, 360–62 ("The investigatory utility of standalone databases improves when law enforcement agencies and private companies connect those databases and aggregate their data. Indeed, linking traditional criminal justice data with private data provides a wealth of insights about a person . . . [and] law enforcement and private companies have embraced the idea of networking and sharing personal information.").

[47] *Id.* at 353, 365 ("To solve crimes, law enforcement must not only collect information, but also identify and link individuals to their accumulated data. In short, data must be connected with identifiable human beings.").

[48] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 4, 20–21.

[49] *See id.* at 4–5, 21–33.

[50] *See, e.g.*, *Policing Predictive Policing*, *supra* note 12, at 1112–14; RISE OF BIG DATA POLICING; *supra* note 14, at 1–6; Brayne, *supra* note 14, at 981; *see also* Andrew Guthrie Ferguson, *Big Data Prosecution and* Brady, 67 UCLA L. REV. 180, 182–83 (2020) [hereinafter *Big Data Prosecution*] ("[B]ig data prosecution tools facilitate evidence collection and information sharing, offering the ability to identify suspects by time, place, associations, or other connections. Adding to these types of formalized, structured databases are growing sources of raw, unstructured big data from digital surveillance technologies like video cameras, police body cameras, and automated license plate readers[]. Prosecutors now sit on a wealth of valuable investigative insights—all searchable and potentially relevant for intelligence gathering and criminal prosecution."); *see also* LEE ET AL., *supra* note 29, at 24 (noting contemporary data-driven policing programs "may vary in the types of data and techniques they employ"); Logan & Ferguson, *supra* note 38, at 549 ("The criminal justice system extends from pre-crime surveillance techniques to post-sentencing community supervision. In almost every context, the system has seen a rapid expansion in data collection, generation, storage, and use."). Of course, adoption of differing big data policing technologies has not been uniform. *See infra* Parts I.B.1, I.B.2, and III.

i. Investigating and Prosecuting Crime

Big data policing may be considered useful in investigating and prosecuting crime.[51] Although many techniques and technologies are involved, three will be treated here as illustrative examples: facial recognition, cell phone-related location data, and police body cameras.

One key big data policing technology for investigation is facial recognition.[52] Facial recognition technology seeks to utilize computers in recognizing and comparing facial images to determine whether they match.[53] Most facial recognition systems are now built with the aid of deep learning, a type of machine learning.[54] With myriad potential uses in law enforcement, police have already begun using or experimenting with facial recognition technologies.[55] One

[51] *See, e.g.*, *The New Surveillance Discretion*, *supra* note 35, at 16 ("[Big data] tools are useful in tracking down evidence of past crimes. . . .").

[52] *See* Andrew Guthrie Ferguson, *Facial Recognition and the Fourth Amendment*, 105 MINN. L. REV. 1105, 1107 (2021) [hereinafter *Facial Recognition and the Fourth*] ("Leading the charge of game-changing new surveillance technologies is facial recognition. . . ."); *see also* Barry Friedman et al., *Policing Police Tech: A Soft Law Solution*, 37 BERKELEY TECH. L.J. 701, 709 (2022) ("In 2016, a landmark report on law enforcement use of [face recognition technology] estimated that one in four agencies have access to this tool, with over 117 million American adults already in face recognition databases."); Patrick K. Lin, *How to Save Face & the Fourth Amendment: Developing an Algorithmic Accountability Industry for Facial Recognition Technology in Law Enforcement*, 33 ALB. L.J. SCI. & TECH. (forthcoming 2023) ("Facial recognition appears to be a genie that is not going back in the bottle.").

[53] Greenberg, *supra* note 7, at 214; *see also* WILLIAM CRUMPLER & JAMES A. LEWIS, CENTER FOR STRATEGIC & INT'L STUD., HOW DOES FACIAL RECOGNITION WORK? 1 (2021) ("Facial recognition is a way of using software to determine the similarity between two face images in order to evaluate a claim."); *Facial Recognition and the Fourth*, *supra* note 52, at 1109 ("The simple idea behind facial recognition is to have a computer program automatically match a digital image of a face with a similar digital image of a face in a stored database."); Matthew E. Cavanaugh, *Somebody's Tracking Me: Applying Use Restrictions to Facial Recognition Tracking*, 105 MINN. L. REV. 2443, 2446 (2021) ("At a high level, facial recognition can be understood as a computer generating probabilities that an image of a person matches an image in a database."); RISE OF BIG DATA POLICING, *supra* note 14, at 89 ("Facial-recognition technology offers virtual fingerprinting for anyone in front of a camera.").

[54] CRUMPLER & LEWIS, *supra* note 53, at 1–2; *see also* Greenberg, *supra* note 7, at 215 ("Programmers have broken down the task into four basic steps, each of which entails its own complexities: (1) face detection, which separates the face within an image from its background; (2) normalization, in which the image is adjusted to a standard size, pose, and illumination; (3) feature extraction, in which a mathematical representation of the face is created to use as a reference point for comparison between images; and (4) matching images for identification and verification.").

[55] *See* Christopher Jones, *Law Enforcement Use of Facial Recognition: Bias, Disparate Impacts on People of Color, and the Need for Federal Legislation*, 22 N.C. J. L. & TECH. 777, 781 (2021) ("Over the past twenty years, facial recognition searches by law enforcement have become relatively routine at the state and federal level."); *Facial Recognition and the Fourth*, *supra* note 52, at 1107, 1115; Eldar Haber, *Racial Recognition*, 43 CARDOZO L. REV. 71, 84–85 (2021) ("During the 2001 Superbowl in Tampa, Florida, the police admittedly used facial recognition tools to locate subjects of outstanding warrants, in perhaps the first reported event in America. At roughly the same time, police departments across America were reported to have begun using facial recognition technology. Officially, the New York Police Department [] reported its use of facial recognition since 2011, while the Detroit Police Department has been using it since at least 2017."); Harvey Gee, *Surveillance State: Fourth Amendment Law, Big Data Policing, and Facial Recognition Technology*, 21 BERKELEY J. AFR.-AM. L. & POL'Y 43, 44 (2021) ("[T]he Los Angeles Police Department admitted to using facial recognition nearly 30,000 times since 2009."). It is worth noting that identifying and recognizing a face may be a difficult task for a computer. *See* Greenberg, *supra* note 7, at 215 ("Not only do computers lack instinctive recognition of facial patterns, they lack even the fundamental conception of what a face is, where it is, and how to differentiate between face and not-face in the mass of visual data from a photograph."). For instance, a

commenter recounts the following 2017 story of a Florida man who was fleeing custody:

> After successfully bringing the suspect's car to a halt, the police approached the driver's side of the vehicle to find a man, seemingly unconscious after ingesting an unknown substance, with no identification card and whose fingerprints appeared to have been chewed off. With no other way to identify him, the officers ran a photo of the man through their facial recognition database, a statewide program that [had] been in place for almost twenty years. The database found a likely match for the man's identity, and the police were able to positively identify the suspect despite his unresponsive state.[56]

Facial recognition technologies may be used to confirm identity of an individual for law enforcement (perhaps matching a crime scene image with a photo dataset), conduct targeted tracking (perhaps scanning video feeds to identify a specifically targeted face), or even conduct more generalized surveillance (perhaps using camera video to map activities and movements of individuals over time).[57]

Similarly, cell phone-related location data has been an important big data policing technology.[58] There are hundreds of millions of cell phones in the United States.[59] These cell phones may reveal location information regarding their users, such as through the time-stamped record generated when phones connect to a cell-

computer might first need to learn what a face is before such a computer could differentiate between faces. *Id.* at 216 ("For a computer to understand a face, it must be expressed in a language that the computer can understand: mathematical models."); *see* CRUMPLER & LEWIS, *supra* note 53, at 1 ("Facial recognition is improving rapidly, but while algorithms can achieve very high performance in controlled settings, many systems have lower performance when deployed in the real world.").

[56] Jones, *supra* note 55, at 778–79 (footnotes omitted).

[57] *See Facial Recognition and the Fourth*, *supra* note 52, at 1107, 1115; *see also* Cavanaugh, *supra* note 53, at 2443 ("In the very near future, the technology will be in place for all public movements to be recorded. As you walk down the street, a network of cameras will capture your movements and be able to identify you from your facial features. Facial-recognition-capable cameras will watch from shop windows, from telephone poles, and from body cameras worn by patrolling police officers. Every person will carry at least one such camera with them on their phone.").

[58] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 11; Pait, *supra* note 4, at 155–56 ("With the widespread adoption of cell phones, much of the American population is now carrying a tracking device by default, endowing law enforcement with ample opportunities to locate a suspect."); Pell & Soghoian, *supra* note 20, at 142–48 (discussing the Stingay, a surveillance device "used by law enforcement agencies to obtain, directly and in real time, unique device identifiers and detailed location information of cellular phones[.]"); Ohm, *supra* note 20, at 361–66 (discussing collection of cell-site location information by law enforcement and the Court's opinion in *Carpenter v. United States*); Cumpstone, *supra* note 20, at 84–86 (explaining cellular data tracking); *Predictive Reasonable Suspicion*, *supra* note 3, at 355–56; *Legal Risks of Big Data Policing*, *supra* note 9, at 6 ("As we go about our daily lives, we are tracked by the smartphone in our pocket.").

[59] *See* Carpenter v. United States, 138 S. Ct. 2206, 2211 (2018) ("Cell phones perform their wide and growing variety of functions by connecting to a set of radio antennas called 'cell sites.' . . . Cell phones continuously scan their environment looking for the best signal, which generally comes from the closest cell site. Most modern devices, such as smartphones, tap into the wireless network several times a minute whenever their signal is on, even if the owner is not using one of the phone's features. Each time the phone connects to a cell site, it generates a time-stamped record known as cell-site location information.").

site (called cell-site location information) or via a phone app.[60] Location data facilitated by cell phones has played a role in several high-profile matters. First, it played a role in the homicide case against Adnan Syed for the murder of his girlfriend, Hae Min Lee, made famous by the *Serial* podcast.[61] Hae Min's body was discovered at Leakin Park in Baltimore nearly a month after her disappearance, and prosecutors alleged that records from AT&T placed Adnan's cell phone in or near Leakin Park on the night Hae Min disappeared.[62] Second, it played a role in the Adea Shabani disappearance featured in the *To Live and Die in LA* podcast.[63] In that podcast, it was suggested that Google data had tracked the movements of an individual who allegedly picked up Adea on the day of her disappearance.[64] Third, cell phone-related location data played a central role in the leading Supreme Court privacy and criminal procedure case of *Carpenter*.[65] In that case, a suspect had identified accomplices to the FBI and provided the FBI with certain accomplices' cell phone numbers.[66] The FBI identified additional numbers the suspect had called near the time of certain robberies, and prosecutors used that information to, among other things, obtain Timothy Carpenter's cell phone records.[67] Cell-site data allegedly placed Carpenter near the scene of several robberies.[68]

Lastly, police body cameras have been an important big data policing investigative technology.[69] Body cameras are "small recording devices positioned on an officer's person" that record "what an officer 'sees and hears,'" potentially

[60] *See, e.g.*, Ryan Nakashima, *AP Exclusive: Google Tracks Your Movements, Like It or Not*, AP (Aug. 13, 2018), https://apnews.com/article/north-america-science-technology-business-ap-top-news-828aefab64d4411bac257a07c1af0ecb [https://perma.cc/LB9A-3E7X] ("For the most part, Google is upfront about asking permission to use your location information. An app like Google Maps will remind you to allow access to location if you use it for navigating. If you agree to let it record your location over time, Google Maps will display that history for you in a 'timeline' that maps out your daily movements.").

[61] *See* Swaine, *supra* note 11.

[62] *Id.* ("Those AT&T records said Syed's phone 'pinged' a cellphone tower covering the park and nearby areas during calls he received at 7.09pm and 7.16pm on 13 January 1999. They were described by Syed's current attorney as 'the pillar of the state's case' against him."); *see also* Amelia Mcdonell-Parry, Serial *Subject Adnan Syed: 4 Key Pieces of Evidence, Explained*, ROLLING STONE (July 1, 2016, 7:04 PM), https://www.rollingstone.com/culture/culture-news/serial-subject-adnan-syed-4-key-pieces-of-evidence-explained-240960/ [https://perma.cc/UBH4-4336]; *FBI Agent: Cell Tower Data in* Serial *Case Accurate*, CBS NEWS (Feb. 8, 2016, 5:23 PM), https://www.cbsnews.com/news/fbi-agent-cell-tower-data-in-serial-case-accurate/ [https://perma.cc/Z3GE-VS8F]; Ralph Ellis, *Adnan Syed, Subject of* Serial *Podcast, Will Not Get a New Trial*, CNN (Mar. 8, 2019, 5:16 PM), https://www.cnn.com/2019/03/08/us/serial-adnan-syed-conviction-reinstated/index.html [https://perma.cc/9ES7-W5GH] (discussing cell phone-related evidence "used . . . to place Syed at the site where Lee was buried"); Victoria Saxe, *Junk Evidence: A Call to Scrutinize Historical Cell Site Location Evidence*, 19 U.N.H. L. REV. 133, 146 (2020) ("At trial, cellphone records showing that Syed's phone pinged a cell tower near the park where the victim's body was found 'played a significant role in the State's case and the jury's decision-making process.'").

[63] *See To Live and Die in LA*, *supra* note 11 (Season 1, Episode 8), at 02:17.

[64] *Id.*

[65] *See* 138 S. Ct. 2206, 2212–13 (2018); Ohm, *supra* note 20, at 361–66. Even though this case concerns the FBI rather than local law enforcement, it is still a helpful example for current purposes.

[66] *Carpenter*, 138 S. Ct. at 2212.

[67] *Id.*

[68] *Id.* at 2212–13, 2220 (finding that "[t]he Government's acquisition of the cell-site records [in this case] was a search under [the Fourth] Amendment" and remanding the case to lower court for further proceedings).

[69] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 89 ("The growth of police-worn body cameras provides additional identification capabilities.").

recording the "officer's actions or interactions with others."[70] Their use has "seemingly proliferated" in recent years, and they allow law enforcement to identify when and where officers made contact with a given person.[71] Next-generation cameras incorporating real-time facial-recognition technology may also permit police to know of prior violence, prior warrants, or generally peaceful conduct.[72] Video footage, including from body cameras, has already, for instance, played an important role in connection with the death of George Floyd.[73]

ii. Predicting Criminal Activity

The 2002 Tom Cruise science fiction film, *Minority Report*, depicted a future where a specialized police unit was able to arrest criminals in advance of their crimes.[74] Approximately twenty years later, in numerous American cities, police have used or are using some type of predictive policing for crime deterrence.[75] Predictive policing, as currently theorized, appears largely predicated on the idea that particular varieties of crimes can be identified through the study of past criminal activity.[76] It involves drawing data from a variety of sources, analyzing such data,

[70] Ronald J. Coleman, *Police Body Cameras: Go Big or Go Home?*, 68 BUFF. L. REV. 1353, 1357 (2020) (footnotes omitted) ("Footage from such cameras might, for instance, provide clarity on a disputed incident involving an officer and member of the community."). These cameras "can be small and lightweight such that they may be placed in a variety of areas, including on a uniform, headgear, or even sunglasses." *Id.* at 1358 (footnote omitted).

[71] Coleman, *supra* note 70, at 1363 (footnote omitted); RISE OF BIG DATA POLICING, *supra* note 14, at 89; *see also* Mary D. Fan, *Justice Visualized: Courts and the Body Camera Revolution*, 50 U.C. DAVIS L. REV. 897, 901, 906 (2017) (referring to "the body camera revolution").

[72] RISE OF BIG DATA POLICING, *supra* note 14, at 89; *see also* Katelyn Ringrose, *Law Enforcement's Pairing of Facial Recognition Technology with Body-Worn Cameras Escalates Privacy Concerns*, 105 VA. L. REV. ONLINE 57, 58 (2019) (footnotes omitted) ("Companies are racing to integrate [body-worn cameras] with facial recognition technology, hoping to eventually use artificial intelligence to recognize faces captured in real time, despite privacy concerns. Once equipped with facial-recognition technology, [body-worn cameras] could dramatically increase the number of individuals logged in law enforcement facial-recognition networks, enabling police officers to act as sophisticated surveillance mechanisms.").

[73] *See, e.g.*, Coleman, *supra* note 70, at 1354; Tim Arango, Nicholas Bogel-Burroughs & Jay Senter, *Three Former Officers Were Convicted of Violating George Floyd's Rights*, N.Y. TIMES (last updated Feb. 24, 2022), https://www.nytimes.com/live/2022/02/24/us/george-floyd-trial-verdict#guilty-verdict-george-floyds-rights [https://perma.cc/V8QE-Q89U] ("Prosecutors relied on the mountains of video evidence—from bystanders, from the officers' body-worn cameras, from city surveillance cameras—that provided an excruciating second-by-second record of the killing."); Eric Levenson, *What We Know About the Federal Trial Against 3 Former Minneapolis Officers,* CNN (last updated Jan. 24, 2022, 10:21 AM), https://www.cnn.com/us/live-news/george-floyd-killing-officers-federal-trial-01-24-22/index.html [https://perma.cc/6RDM-WUV3] ("The three ex-officers' actions during Floyd's arrest in May 2020 were shown in detail during Chauvin's state trial in videos from bystanders, police body cameras and surveillance footage.").

[74] *Minority Report*, IMDB, https://www.imdb.com/title/tt0181689/ [https://perma.cc/M36F-4TKD].

[75] *See Legal Risks of Big Data Policing*, *supra* note 9, at 5; Tim Lau, *Predictive Policing Explained*, BRENNAN CTR. FOR JUST. (last updated Apr. 1, 2020), https://www.brennancenter.org/our-work/research-reports/predictive-policing-explained [https://perma.cc/Y28D-WU7V] ("Police departments in some of the largest U.S. cities have been experimenting with predictive policing as a way to forecast criminal activity.").

[76] *See Legal Risks of Big Data Policing*, *supra* note 9, at 5 ("In a predictive policing jurisdiction, the crime numbers are crunched and spit out into usable maps that can identify particular areas of possible crime so that police can patrol those areas. The goal is 'to predict and deter' under the logic that if the risk forecast is accurate, the police presence will deter the potential criminal actor from following through on his criminal plan.").

and then using the results to help anticipate, prevent, and better respond to future criminal activity.[77]

One variety of predictive policing is location-based prediction.[78] This type of predictive policing normally utilizes prior crime data to identify locations and times with a high risk of criminal activity.[79] Volumes of data may be fed into algorithms to produce "hot spots" worthy of additional scrutiny.[80] Armed with this information, officers might, for example, drive through predicted areas of burglaries during lulls in their patrols with the goal of disrupting the potential burglaries before they occur.[81] Certain departments have experimented with predictive software to identify places where criminal activity is likely to occur, and some have partnered with academic enterprises and companies to forecast locations.[82]

Another variety of predictive policing is person-based prediction.[83] This involves departments using predictive analytics for identification of people potentially at risk of becoming involved in criminal activity.[84] This may involve placing an individual on an at-risk "heat list"—perhaps due to the individual's prior connection to violence, associates, and friends.[85] For example, the Chicago Police Department has utilized big data tools in identifying high-risk individuals using a list.[86] Person-based prediction might also involve "focused deterrence"—a theory

[77] *See* Pearsall, *supra* note 33, at 17 (footnote omitted) ("Predictive policing entails becoming less reactive. 'The predictive vision moves law enforcement from focusing on what happened to focusing on what will happen and how to effectively deploy resources in front of crime, thereby changing outcomes,' writes Charlie Beck, chief of the Los Angeles Police Department."); *see also The New Surveillance Discretion*, *supra* note 35, at 16 ("[B]ig data also provides the police with new capabilities to identify ongoing and future threats."); Lau, *supra* note 75 ("Predictive policing involves using algorithms to analyze massive amounts of information in order to predict and help prevent potential future crimes.").

[78] *See* Lau, *supra* note 75 (referring to "[p]lace-based predictive policing" as "the most widely practiced method").

[79] *Id.*

[80] *See* RISE OF BIG DATA POLICING, *supra* note 14, at 63 (footnote omitted) ("While these crime patterns intuitively may be known by police officers, now with advanced data analytics, years' worth of crime patterns can be studied, mapped, and proactively deployed.").

[81] *Id.* (discussing the "risk map" produced by algorithm for the Jennings Police Department in Missouri).

[82] *See Legal Risks of Big Data Policing*, *supra* note 9, at 5 (citation omitted) ("Some predictive policing algorithms only rely on past criminal incidents, day, time, and place, while others add in more complex variables like the time of year, weather, and particular local factors (fairs, football games) and yet other models study fixed structures that might encourage criminal activity (bus stops, liquor stores) providing the cover for loitering and/or the targeting of victims."); *The New Surveillance Discretion*, *supra* note 35, at 16 (footnotes omitted) ("Police departments in Santa Cruz (CA), Seattle, and New York City are experimenting with predictive policing software to identify geographic places where crime is likely to take place. One day the police nationwide may use location-based tweets to inform those same predictions.").

[83] *See* Lau, *supra* note 75.

[84] *Legal Risks of Big Data Policing*, *supra* note 9, at 5; Lau, *supra* note 75 ("Person-based predictive policing . . . attempts to identify individuals or groups who are likely to commit a crime—or to be victim of one—by analyzing for risk factors such as past arrests or victimization patterns.").

[85] RISE OF BIG DATA POLICING, *supra* note 14, at 34.

[86] *See The New Surveillance Discretion*, *supra* note 35, at 16–17 (footnote omitted) ("The Chicago Police Department already uses big data tools to identify high risk persons based on the strength of a person's social networks: a technique borrowed from the military's analysis of insurgent groups."); *Legal Risks of Big Data Policing*, *supra* note 9, at 5 (citation omitted) ("[I]n Chicago, the 'Strategic Subjects List' creates a rank-ordered list of the people in Chicago who are most at risk at being either the perpetrator or victim of a violent crime. Each identified person is given a threat score from 1 to 500+, with the police attention and focus being on those with the highest scores.").

seeking to understand and dismantle criminal actor networks driving violent crime.[87] In theory, police may be aware of the small percentage of individuals involved in crimes, but the challenge is getting the supposed "criminals" to know that police are aware of them.[88] A focused deterrence program might target messages to a small segment of the populace that prosecutors, the community, and police know are engaged in violence.[89]

Predictive policing projects have been attempted in various police departments. For instance, some of the earliest examples were Los Angeles Police Department programs, which came to include identification of likely areas of gun violence (LASER program) and "hot spots" for property-related crime (PredPol program).[90] The New York Police Department also developed in-house algorithms

[87] RISE OF BIG DATA POLICING, *supra* note 14, at 35.

[88] *Id.*

[89] *See id.* at 35–36 ("In 2012, Kansas City, Missouri, implemented a bold data-driven, focused-deterrence experiment. . . . The Kansas City Police Department used advanced social network analysis to visualize the at-risk men responsible for the violence. . . . [T]he focused-deterrence process had three steps: (1) identify criminal actors, (2) give notice to those actors that police are aware of their activities and offer social services, and (3) arrest, prosecute, and punish those individuals who were warned but ignored the warnings."); s*ee also id.* at 37 (Chicago's "heat list" contemplates an algorithmic approach to focused deterrence).

[90] *See* Lau, *supra* note 75 ("[T]he Los Angeles Police Department . . . started working with federal agencies in 2008 to explore predictive policing approaches."); *see also* Brayne, *supra* note 14, at 986–90 (discussing Operation LASER and PredPol); *see also* Eva Ruth Moravec, *Do Algorithms Have a Place in Policing?*, THE ATLANTIC (Sept. 5, 2019), https://www.theatlantic.com/politics/archive/2019/09/do-algorithms-have-place-policing/596851/ [https://perma.cc/Z23Q-YL8U] ("[R]esearchers and the [Los Angeles Police Department] designed LASER and PredPol experiments, and other cities including Chicago, Memphis, Minneapolis, and Dallas followed suit. The predictive-policing trend was becoming so much a part of modern-day life that in 2011, Time magazine called 'pre-emptive policing'—another name for it—one of the 50 best inventions of the year."); *see also* Issie Lapowsky, *How the LAPD Uses Data to Predict Crime*, WIRED (May 22, 2018, 5:02 PM), https://www.wired.com/story/los-angeles-police-department-predictive-policing/ [https://perma.cc/277G-ZX7T] ("The Los Angeles Police Department is one of dozens of cities across the country that's trying to predict where crime will happen—and who those future criminals will be—based on past crime and arrest data."); *see also* WALTER L. PERRY, ET AL., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS 4 (2013) ("Police Chief (ret.) William J. Bratton and the LAPD [Los Angeles Police Department] are credited with envisioning the predictive policing model. By 2008, Chief Bratton had spoken widely in the public arena about the successes of the LAPD, including the department's recent introduction of predictive analytics to anticipate gang violence and to support real-time crime monitoring."); *see also* Mark Puente, *LAPD Ends Another Data-Driven Crime Program Touted to Target Violent Offenders*, L.A. TIMES (Apr. 12, 2019, 4:48 PM), https://www.latimes.com/local/lanow/la-me-laser-lapd-crime-data-program-20190412-story.html [https://perma.cc/KRU8-QAG4] ("LASER, or 'Los Angeles' Strategic Extraction and Restoration,' zones, used data mapping to increase police presence in hot spots and identify specific 'anchor points' such as liquor stores, parking lots and residences connected to certain crimes in an area. Analysts, not computers, identified where many crimes occurred and where to send more officers."); *see also* Johana Bhuiyan, *LAPD Ended Predictive Policing Programs Amid Public Outcry. A New Effort Shares Many of Their Flaws*, THE GUARDIAN (Nov. 8, 2021, 1:00 AM), https://www.theguardian.com/us-news/2021/nov/07/lapd-predictive-policing-surveillance-reform [https://perma.cc/8KMQ-SKSC] ("The Los Angeles police department has been a pioneer in predictive policing, for years touting avant-garde programs that use historical data and software to predict future crime."); *see also* Grace Baek & Taylor Mooney, *LAPD Not Giving Up on Data-Driven Policing, Even After Scrapping Controversial Program*, CBS NEWS (Feb. 23, 2020, 7:00 AM), https://www.cbsnews.com/news/los-angeles-police-department-laser-data-driven-policing-racial-profiling-2-0-cbsn-originals-documentary/ [https://perma.cc/CCT2-XRJE] ("The LAPD . . . is often credited with pioneering data-driven policing programs that are now used across the country.").

for predictive policing and began using them in 2013.[91] The Chicago Police Department was responsible for one of the United States's leading examples of person-based predictive programs.[92] Concerns with predictive policing and independent audits seemingly have led certain police departments to significantly reduce or phase out their programs.[93]

[91] Lau, *supra* note 75 ("According to a 2017 paper by department staff, the [New York Police Department] created predictive algorithms for several crime categories, including shootings, burglaries, felony assaults, grand larcenies, grand larcenies of motor vehicles, and robberies. Those algorithms are used to help assign officers to monitor specific areas."); RISE OF BIG DATA POLICING, *supra* note 14, at 30 ("[William] Bratton [who had taken over as head of the New York Police Department] doubled down on data-driven policing. . . . Predictive policing was in. Bratton ordered tens of thousands of crime-mapping tablet computers for his officers. He oversaw the launch of a cutting-edge, real-time crime command center in Manhattan.").

[92] Lau, *supra* note 75 ("First piloted in 2012, the program, called the 'heat list' or 'strategic subjects list,' created a list of people it considered most likely to commit gun violence or to be a victim of it. The algorithm, developed by researchers at the Illinois Institute of Technology, was inspired by research out of Yale University that argued that epidemiological models used to trace the spread of disease can be used to understand gun violence."); RISE OF BIG DATA POLICING, *supra* note 14, at 37 ("Who gets shot? The algorithm knows. . . . On Memorial Day 2016, 78% of the 64 people shot were on the list. Using the heat list, police have prioritized youth violence to intervene in the lives of the most at-risk men."); *Legal Risks of Big Data Policing*, *supra* note 9, at 5; Jeff Asher & Rob Arthur, *Inside the Algorithm That Tries to Predict Gun Violence in Chicago*, N.Y. TIMES (June 13, 2017), https://www.nytimes.com/2017/06/13/upshot/what-an-algorithm-reveals-about-life-on-chicagos-high-risk-list.html [https://perma.cc/NX2T-E9A7] ("The [Strategic Subject List] is made by an algorithm that tries to predict who is most likely to be involved in a shooting, either as perpetrator or victim."); Kathleen Foody, *Chicago Police End Effort to Predict Gun Offenders, Victims*, AP NEWS (Jan. 23, 2020), https://apnews.com/article/41f75b783d796b80815609e737211cc6 [https://perma.cc/ZY8A-5PNV] ("[The formula] relied on several factors including age during an individual's latest arrest, the number of times someone was the victim of a shooting, battery or assault and their total number of arrests for unlawful weapons. Early versions of the formula also used gang affiliation and narcotics arrests to calculate risk scores."). There are, of course, other examples of experiments or programs in the United States that some might characterize as constituting predictive policing, and the same applies abroad. *See, e.g.*, Ali Winston, *New Orleans Ends its Palantir Predictive Policing Program*, THE VERGE (Mar. 15, 2018, 3:50 PM), https://www.theverge.com/2018/3/15/17126174/new-orleans-palantir-predictive-policing-program-end [https://perma.cc/F3UN-ZRL7]; Olivia Solon & Cyrus Farivar *Predictive Policing Strategies for Children Face Pushback*, NBC NEWS (June 6, 2021, 6:00 AM), https://www.nbcnews.com/tech/tech-news/predictive-policing-strategies-children-face-pushback-n1269674 [https://perma.cc/WD9W-ACZG]; Pearsall, *supra* note 33, at 17; Jean Gil Barroca, *Surveillance and Predictive Policing Through AI*, in URBAN FUTURE WITH A PURPOSE 137 (2021).

[93] Lau, *supra* note 75 ("Critics . . . warn about a lack of transparency from agencies that administer predictive policing programs. They also point to a number of civil rights and civil liberties concerns, including the possibility that algorithms could reinforce racial biases in the criminal justice system. These concerns, combined with independent audits, have led leading police departments, including in Los Angeles and Chicago, to phase out or significantly reduce the use of their predictive policing programs after auditing them."); Ángel Díaz, *Data-Driven Policing's Threat to Our Constitutional Rights*, BROOKINGS INST. (Sept. 13, 2021), https://www.brookings.edu/techstream/data-driven-policings-threat-to-our-constitutional-rights/ [https://perma.cc/3WYF-SW3J]; Puente, *supra* note 90 ("The Los Angeles Police Department has scrapped a second data policing program it once hailed as a way to target violent offenders in neighborhood hot spots, following concerns that the programs unfairly target black and Latino communities."); Bhuiyan, *supra* note 90 ("[N]ewly revealed public documents detail how PredPol and Operation Laser, the department's flagship data-driven programs, validated existing patterns of policing and reinforced decisions to patrol certain people and neighborhoods over others, leading to the over-policing of Black and brown communities in the metropole."); Baek & Mooney, *supra* note 90 ("[I]n a rare reversal for the department, the LAPD shut down LASER."); Foody, *supra* note 92 ("Chicago police have ended a program that sought to predict people most likely to be victims or perpetrators of gun crime."). Certain risks advanced relating to big data policing more generally will be discussed in Part II.C.

C. Risks of Big Data Policing

Notwithstanding big data policing's potential benefits, many have also argued it presents serious challenges—including, in particular, relating to privacy, data security, and fairness.[94] To begin, privacy and data security have been raised as risks of big data policing.[95] Big data policing may present challenges to Fourth Amendment rights of citizens or to the personal privacy of individuals.[96] Facial

[94] *See, e.g.*, *Legal Risks of Big Data Policing*, *supra* note 9, at 5 ("The rise of big data policing creates . . . substantial dangers."); *see* Justin Ye, *The Slippery Slope of Big Data in Policing*, HARV. INT'L REV. (May 27, 2021, 9:00 AM), https://hir.harvard.edu/big-data-in-policing/ [https://perma.cc/8E6L-6VZB]; *see also The New Surveillance Discretion*, *supra* note 35, at 15–19; *see also* Lau, *supra* note 75; *see also* Simmons, *supra* note 36, at 577–78. It is not here possible to itemize and discuss all the risks (or benefits) that have been raised in connection with big data policing, nor is it this Article's intention to weigh risks against benefits. Instead, this Article seeks only to mention some of the most prominent risks that have been advanced.

[95] *See, e.g.*, *Policing By Numbers*, *supra* note 28, at 42 ("While the use of big data in the private sector has raised concerns about consumer privacy, its use by the police raises even bigger questions about the limits of using data to justify surveillance, investigation, and detention by the police."); *see* Lau, *supra* note 75 (discussing predictive policing and the Fourth Amendment); *see also* Ye, *supra* note 94 (discussing security).

[96] *See, e.g.*, *Facial Recognition and the Fourth*, *supra* note 52, at 1109 ("If there is one technological innovation that has gotten the attention of the privacy and civil rights community it is facial recognition."); *see Policing By Numbers*, *supra* note 28, at 38 ("Scholars have widely discussed the shortcomings of applying Fourth Amendment doctrines, once adequate for a world of electronic beepers, physical wiretaps, and binocular surveillance, to rapidly changing technologies. But big data may magnify these concerns considerably.") (footnote omitted); *see also Legal Risks of Big Data Policing*, *supra* note 9, at 6 ("Do current forms of mass surveillance fall outside of the Fourth Amendment?"); *see also The New Surveillance Discretion*, *supra* note 35, at 34 ("In other words, surveillance that does not intrude upon recognized Fourth Amendment interests requires no prior justification by the police. The who, how, and why of police decisions to single out persons for attention is a matter of police discretion.") (footnote omitted); *see also* Lau, *supra* note 75 ("Some legal experts argue that predictive policing systems could threaten rights protected by the Fourth Amendment, which requires 'reasonable suspicion' for a police officer stop—a legal standard that helps protect individuals against 'unreasonable searches and seizures' by the police. Predictive analytics tools may make it easier for police to claim that individuals meet the reasonable suspicion standard, ultimately justifying more stops."); *see also* Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 117 (2018) ("As Andrew Ferguson has observed, the Fourth Amendment's reasonable suspicion requirement is inherently a 'small data doctrine,' rendering it impotent in even its primary uses when it comes to data mining.") (quoting *Predictive Reasonable Suspicion, supra* note 3, at 338); *see also* Michael L. Rich, *Machine Learning, Automated Suspicion Algorithms, and the Fourth Amendment*, 164 U. PA. L. REV. 871, 929 (2016); *see also* Richard M. Re, *Fourth Amendment Fairness*, 116 MICH. L. REV. 1409, 1434–35 (2018); *see also Predictive Reasonable Suspicion*, *supra* note 3, at 329–30 ("The rise of big data technologies offers a challenge to the traditional paradigm of Fourth Amendment law. With little effort, officers can now identify most unknown suspects, not through their observations, but by accessing a web of information containing extensive personal data about suspects."); *see also* Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J. L. & TECH. 106, 126 (2019) ("The federal government has mastered the art of ubiquitous surveillance, some legal and some illegal. Rather than survey the copious types of surveillance, and Supreme Court cases upholding or rejecting them, here we discuss only those forms and doctrines that contribute to AI's erosion of privacy interests."); *see also* Mary Madden et al., *Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans*, 95 WASH. U. L. REV. 53, 55 (2017) ("The privacy harms that poor communities and their residents suffer as a result of pervasive surveillance are especially acute in light of the resulting economic and social consequences and the low likelihood that they will be able to bear the costs associated with remedying those harms. In the 'big data' era, there are growing concerns that low-status Internet users who have lower levels of income or education may be further differentially impacted by certain forms of Internet-enabled data collection, surveillance, and marketing.") (footnote omitted); *see also* Brennan-Marquez, *supra* note 19, at 488 ("[T]he distinctive feature of big data policing—beyond its statistical promise—is that it multiplies, both quantitatively and

recognition, for instance, may raise the specter of unbridled police surveillance, especially when used in a predictive capacity or for tracking a targeted individual for a specified crime.[97] Body cameras might facilitate the capture and sharing of irrelevant, inappropriate, or embarrassing footage of targeted individuals, officers, or even uninvolved bystanders.[98] Use of cell-site location information even led to a

qualitatively, the experience of being subject to 'police presence.' Once all of life is documented and databased, once officials can make use . . . of 'time machines,' officers no longer need to be investigating contemporaneously, let alone physically present, to inspire self-monitoring and behavior modification.") (footnote omitted); *see also* Harvey Gee, *Almost Gone: The Vanishing Fourth Amendment's Allowance of Stingray Surveillance in a Post-Carpenter Age*, 28 S. CAL. REV. L. & SOC. JUST. 409, 410 (2019) ("The Fourth Amendment continues to erode. . . . You can . . . be stopped while you are walking down the street in a 'high crime area' and checked for an active arrest warrant. Cops can also secretly track your location via your smart phone using cell-site simulators, known as Stingrays, that send powerful electronic signals to bait automatic responses from all nearby cell phones.") (footnotes omitted); *see also* Ye, *supra* note 94 ("[T]here are growing concerns about the use of controversial technologies by law enforcement and private companies which, despite good intentions, invade personal privacy."); *see also* Friedman et al., *supra* note 52, at 109 ("[I]n the absence of regulation, tech vendors are enmeshed in a race to the ethical bottom, innovating new and ever more intrusive ways to track and surveil the citizenry. These technologies are marketed aggressively to policing agencies. . . . And agencies use these tools with little in the way of controls that mitigate their civil rights and civil liberties impact.").

[97] *See* Susan Pratt, *From the Eyes of a Machine: Image Recognition Technologies*, 5 GEO. L. TECH. REV. 201, 209 (2021) ("[T]here are concerns that some [image recognition technology] applications may lead to privacy violations and concerns that minority populations may be targeted. These concerns relate primarily to facial recognition technology. For example, alarms were raised when a statement was made that a researched facial recognition tool could 'predict criminals.'") (footnotes omitted); *see also* Cavanaugh, *supra* note 53, at 2473–74 ("[T]he rise of modern industrial societies and systems of government have changed the dynamics of surveillance practices. As smaller units of social organization have been replaced by massive governments and more efficient, industrialized systems, concerns about surveillance that were once confined to person-to-person practices like 'Peeping Toms' gave way to the dystopian visions of Orwell's *1984* and Huxley's *Brave New World*. This change was driven by the development of ever more efficient surveillance systems. The development of panvasive surveillance technologies like facial recognition tracking is a continuation of this trend.") (footnotes omitted); *see also Facial Recognition and the Fourth*, *supra* note 52, at 1107 (Suggesting experimentation by police with facial recognition technology is "causing great public concern, because the scope and scale of these new surveillance systems threatens to upend the existing power relationship between police and the people"); *see also* Gee, *supra* note 55, at 76 ("[F]acial recognition and facial surveillance technology are the latest threats to associational privacy and personal security."); *see generally* Kashmir Hill, *The Secretive Company That Might End Privacy As We Know It,* N.Y. TIMES (last updated Nov. 2, 2021), https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html [https://perma.cc/D8E6-9HKP] (discussing Clearview AI's facial recognition app and privacy); *see also* Greg Bensinger, *How Illinois is Winning in the Fight Against Big Tech,* N.Y. TIMES (May 30, 2022), https://www.nytimes.com/2022/05/30/opinion/illinois-biometric-data-privacy.html [https://perma.cc/CWL2-XUS7] (discussing Clearview AI and a lawsuit in Illinois). "Face tracking" is a term that has been used to describe officers actively targeting a specific person for a specific crime using facial recognition. *See Facial Recognition and the Fourth*, *supra* note 52, at 1122 ("[With face tracking, officers] are not just passively monitoring for generalized surveillance purposes but actively investigating a particular crime with an identifiable suspect using facial recognition matching software. As a general matter, police might use what I am terming 'face tracking' in three different ways: (1) scanning stored video footage to identify a targeted face in the crowd; (2) scanning real-time video feeds to identify a targeted face; and (3) scanning image databases from private third-party platforms to identify a targeted face."); *see also* Cavanaugh, *supra* note 53, at 244–50 ("'Face surveillance' refers to the generalized monitoring of a public space. Face identification matches a particular person (individualized suspicion is present). Facial recognition tracking, or 'face tracking,' combines the two—it describes the practice of obtaining information about an individual's movements using aggregated data obtained via facial identification.") (footnotes omitted).

[98] *See* Coleman, *supra* note 70, at 1371–72; Danielle Evans, *Police Body Cameras: Mending Fences and How Pittsburgh is a Leading Example*, 16 PITT. J. TECH. L. & POL'Y 76, 83 (2015); Johnathan M.

recent Supreme Court case on the Fourth Amendment.[99] In terms of data security, if a greater quantity of sensitive personal information—such as biometrics—were collected and stored to support big data policing programs, and if such collection and storage were not accompanied by adequate security measures, data breaches could lead to identity theft and financial loss for impacted citizens.[100]

Fairness has also been raised as a risk of big data policing.[101] Critics charge that big data policing techniques lack sufficient transparency and may effectively be a "black box."[102] Similarly, critics point out the risks associated with bad data, biases,

Nixon, *Eye Spy Injustice: Delving into the Implications Police Body Cameras Will Have on Police Officers and Citizens*, 60 HOW. L. J. 719, 733 (2017). Combining body cameras and facial recognition could compound the privacy concerns. *See, e.g.*, MICHAEL D. WHITE, POLICE OFFICER BODY-WORN CAMERAS: ASSESSING THE EVIDENCE 27–28 (2014).

[99] *See* Carpenter v. United States, 138 S. Ct. 2206, 2211 (2018); *see also* Alan Z. Rozenshtein, *Fourth Amendment Reasonableness After Carpenter*, 128 YALE L. J. FORUM 943, 945–47 (2019) (discussing *Carpenter*); Gee, *supra* note 96, at 423-29 (2019).

[100] *See, e.g.*, Ye, *supra* note 94 ("A breach of personal data, such as the biometric data, contained in an individual's face could lead to problems with identity theft, which in turn could lead to financial losses with credit and bank accounts."); *Policing Predictive Policing*, *supra* note 12, at 1185 ("The real problem arises with personal data in big data systems. While police benefit from collecting, aggregating, and sharing that individualized and sometimes-sensitive data, concerns about data security exist."); *see also* PETER SWIRE & DEBRAE KENNEDY-MAYO, U.S. PRIVATE SECTOR PRIVACY: LAW AND PRACTICE FOR INFORMATION PRIVACY PROFESSIONALS 422–23 (Julia Homer, 3d ed. 2020) (discussing risk of big data breach and importance of information security).

[101] *See, e.g.*, Lau, *supra* note 75 ("[C]ivil rights organizations, researchers, advocates from overly policed communities, and others have expressed concerns that using algorithmic techniques to forecast crime, particularly by relying on historical police data, could perpetuate existing racial biases in the criminal justice system."); Simmons, *supra* note 36, at 577–78; LEE ET AL., *supra* note 29, at 45–46.

[102] *See* Simmons, *supra* note 36, at 578 ("Unfortunately, big data algorithms are notoriously opaque and incomprehensible, sometimes even to those who are applying them."); *see also Black Data Policing*, *supra* note 12, at 504 ("[B]ig data policing is opaque, lacking transparency because most of the magic happens as a result of 'black box' proprietary and mathematically complex algorithms."); *see also* Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J. L. & TECH. 103, 107–08 (2018) ("Because the designing entities typically do not disclose their predictive models or algorithms, there is a growing literature criticizing the 'black box' opacity of these processes. These black boxes are impervious to question, and many worry that they may be discriminatory, erroneous, or otherwise problematic. Journalists and scholars who have begun to seek details from public entities about these algorithms generally come up short as their freedom of information requests are denied or go unanswered."); *see also* Deven R. Desai & Joshua A. Kroll, *Trust But Verify: A Guide to Algorithms and the Law*, 31 HARV. J. L. & TECH. 1, 6–7 (2017) ("Software and algorithms have gained much attention under the premise that they 'exercise power over us' because they 'select[] what information is considered most relevant to us, a crucial feature of our participation in public life,' are 'powerful entities that govern, judge, sort, regulate, classify, influence, or otherwise discipline the world,' and are 'black boxes.'"); *see also* Katherine Kwong, *The Algorithm Says You Did It: The Use of Black Box Algorithms to Analyze Complex DNA Evidence*, 31 HARV. J. L. & TECH. 275, 298 (2017) ("Some courts have already been willing to exclude evidence produced by black-box DNA analysis."); *see also* Renata M. O'Donnell, *Challenging Racist Predictive Policing Algorithms Under the Equal Protection Clause*, 94 N.Y.U. L. REV. 544, 546–47 (2019) ("Officers are beginning to delegate decisions about policing to the minds of machines. Programmers endow predictive policing algorithms with machine learning . . . With each use, algorithms automatically adapt to incorporate newly perceived patterns into their source codes via machine learning and become better at discerning patterns that exist in the additional swaths of data to which they are exposed. In this way, machine learning creates a 'black-box' conundrum, wherein the algorithm learns and incorporates new patterns into its code with each decision it makes, such that the humans relying on the algorithm do not know what criteria the algorithm might have relied on in generating a certain decision."); *see also* Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, 26 WM. & MARY BILL RTS. J. 287, 288 [hereinafter *Feeding the Machine*] ("[A]n emerging body of scholarship and journalism has already begun to question the presumed neutrality, efficiency, and quality

and inequalities adversely impacting inferences and outcomes.[103] For instance, if a predictive policing algorithm considered past drug arrests in predicting future

of the big data analysis used in policing and other criminal justice institutions. Some have called for greater transparency regarding the 'black box' algorithms that can influence decisions about suspicion, bail, sentencing, and parole. Still others have asked whether the private companies responsible for developing these big data programs should be permitted to invoke intellectual property rights to keep some information from defendants, judges, and researchers."); *see also* Erik Bakke, *Predictive Policing: The Argument For Public Transparency*, 74 N.Y.U. ANN. SURV. AM. L. 131, 133 (2018) ("While real-time predictions of crime locations must be withheld for the technology to provide any real benefit, police should at the very least reveal algorithm inputs, algorithms, and obsolete predictions whenever possible if employing predictive policing."); *see also* Ye, *supra* note 94; *see also* Selbst, *supra* note 96, at 189 ("Every algorithmic accountability proposal (accountability proposals for any technology, really) eventually meets the question of how to handle the trade secret problem. In short, many companies are protecting their algorithms by claiming that they are trade secrets and, therefore, cannot be disclosed. Despite often being of questionable legal merit, such claims are being treated credulously by courts and are given great weight in public debates about 'black box' technologies."); *see also* Samuel R. Wiseman, *The Criminal Justice Black Box*, 78 OHIO ST. L. J. 349, 357 (2017) ("identif[ying] the problem of the black box of criminal justice data"). In connection with algorithms, a "black-box tool" might, for instance, refer to "an algorithmic risk instrument which is not transparent about what is input into the software program and/or how the outputs are generated and quantified." *See also* Melissa Hamilton, *The Biased Algorithm: Evidence of Disparate Impact on Hispanics*, 56 AM. CRIM. L. REV. 1553, 1558 (2019).

[103] *See, e.g.*, Simmons, *supra* note 36, at 577 ("[O]ne of the primary criticisms of big data algorithms is that they can reinforce existing biases and partialities that are already built into the system. If certain groups in society are more likely to be convicted because of their race or the neighborhood in which they live, then a sentencing algorithm that uses, as key factors, prior convictions or home address to determine the length of the sentence will exacerbate existing inequalities."); *Legal Risks of Big Data Policing*, *supra* note 9, at 6–7 ("What if the algorithms demonstrate a racial bias?"); Ye, *supra* note 94 ("Predictive policing is flawed because historical data about who has committed crimes in the past have inherent biases stemming from who and what police prioritize. . . . [T]o give another example, drug arrests in the USA disproportionally target Black individuals, which creates biased historical data."); Lau, *supra* note 75 ("A 2019 study by the AI Now Institute, for example, describes how some police departments rely on 'dirty data'—or data that is 'derived from or influenced by corrupt, biased, and unlawful practices,' including both discriminatory policing and manipulation of crime statistics—to inform their predictive policing systems. Relying on historical crime data can replicate biased police practices and reinforce over-policing of communities of color, while manipulating crime numbers to meet quotas or produce ambitious crime reduction results can give rise to more policing in the neighborhoods in which those statistics are concentrated."); LEE ET AL., *supra* note 29, at 45 ("A growing body of research and journalism has shown that use of predictive algorithms in policing—which primarily use and are trained on historical crime data—replicate and amplify existing systemic biases, often with little to no thought given to how 'different crime-reduction policies, crime legislation, profiling tendencies, or sentencing biases influence the patterns found by [such] algorithms in the data.'"); *Black Data Policing*, *supra* note 12, at 504 ("[B]ig data policing is racially encoded, colored by the history of real-world policing that disproportionality impacts communities of color. Police data comes from the real world, and all of the long-standing discriminatory impacts of implicit and explicit bias color that data. Black data is black, brown, and marked by disproportionate impacts on communities of color."); O'Donnell, *supra* note 102, at 548 ("[W]hen input data—like historical crime data and dragnet data searches—contains information about race, a machine learning algorithm becomes biased by parsing the patterns that exist between race and criminality, regardless of whether the developer explicitly wrote that its source code ought to find such a pattern."); Selbst, *supra* note 96, at 194 ("If they remain unregulated, predictive policing systems will harden and perpetuate the racial discrimination that pervades the criminal justice system. . . . Given the history of discriminatory policing, no technology or police practice should ever be adopted without investigating how it impacts minority populations. Society cannot afford to let the allure of new technologies blind people to the systemic inequalities they can perpetuate."); *Feeding the Machine*, *supra* note 102, at 289 ("To be sure, there have already been concerns raised that the inputs for policing algorithms reflect racial biases."); Bakke, *supra* note 102, at 139 ("Because predictive algorithms rely on preexisting data, biased data can generate biased predictions"); *Policing Predictive Policing*, *supra* note 12, at 1145–50; *see also* Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L. J. 2218, 2225 (2019) ("Actuarial risk assessment, in other words, has revealed the racial inequality inherent in *all* crime

criminals and the drug arrest history skewed in the direction of a given social group, a critic might point out the risk of bias against such social group.[104] Similarly, in the case of facial recognition, critics have charged that the "technology is much less accurate in identifying people of color," and this could lead to false accusations of criminal activity.[105] Alternatively, plain human error might impact the accuracy of data inputs, such as when an officer records or transposes the incorrect address of a crime.[106] Critics may also question whether juries are able to adequately evaluate accuracy, reliability, and fairness of analyses and outcomes.[107] For example, in the case of body cameras, jurors could "reach unjust conclusions" based on body camera footage or "the perceived 'objectivity' of [such] footage could lead to overreliance," even if "biases may impact viewers."[108]

prediction in a racially unequal world, forcing us to confront a much deeper problem than the dangers of a new technology."); Sheri B. Pan, *Get To Know Me: Protecting Privacy and Autonomy Under Big Data's Penetrating Gaze*, 30 HARV. J. L. & TECH. 239, 258 (2016) ("If people increasingly encounter big data generating inferences and helping to form decisions that accord with their preconceived notions of others, big data may deepen the prejudices that exist in society."); Caryn Devins, et al., *The Law and Big Data*, 27 CORNELL J. L. & PUB. POL'Y 357, 361 (2017) ("Because all relevant facts cannot be defined, let alone included, the models used to interpret Big Data are inherently biased in unknown and arbitrary ways."). On the other hand, proponents may argue that big data policing has the potential to combat bias and discrimination. *See, e.g.*, Brennan-Marquez, *supra* note 19 at 490 ("For one thing, data can discourage police from relying on bias, conscious or unconscious, to guide their decisions. To borrow an example from Bennett Capers, 'the increased use of public surveillance cameras and facial recognition technology, coupled with access to Big Data and perhaps terahertz scanners capable of distance scanning for firearms, could do much [to] tackl[e] the . . . problem [of] racialized policing,' and disrupt the 'young plus black equals probable cause' equation that stands a shameful hallmark of much contemporary policing."); Brayne, *supra* note 14, at 982; RISE OF BIG DATA POLICING, *supra* note 14, at 131–36.

[104] *See* Bakke, *supra* note 102, at 139–40.

[105] *See, e.g.*, Jones, *supra* note 55, at 779 (discussing an example of "a thirty-three-year-old Black man from New Jersey, who spent ten days in jail after being falsely accused. . . ."); Clare Garvie & Jonathan Frankle, *Facial-Recognition Software Might Have a Racial Bias Problem*, THE ATLANTIC (Apr. 7, 2016), https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991 [https://perma.cc/8WUM-9KKF] ("Facial-recognition systems are more likely either to misidentify or fail to identify African Americans than other races, errors that could result in innocent citizens being marked as suspects in crimes. And though this technology is being rolled out by law enforcement across the country, little is being done to explore—or correct—for the bias."); Clare Garvie, *Garbage In, Garbage Out: Face Recognition on Flawed Data*, GEO. L. CTR. ON PRIV. & TECH. (May 16, 2019), https://www.flawedfacedata.com [https://perma.cc/TV2D-KZ6N] (discussing potential for misidentification); Haber, *supra* note 55, at 73 ("[Recognition] technology, most notably facial recognition, is constantly and systematically proven to be erroneous—making many inaccurate identifications (false positives). Such inaccuracy, as researchers continuously prove, is not equally spread between cohorts, making dramatically more false identifications for women than for men and . . . for Black people than for white people.").

[106] *See Policing Predictive Policing*, *supra* note 12, at 1145 ("To be used, data must be collected, and much of that collection is done by human beings. Human beings make mistakes.").

[107] *Legal Risks of Big Data Policing*, *supra* note 9, at 7. There may also be risks for other court-related stakeholders, such as lawyer and judges. *Id.*

[108] Coleman, *supra* note 70, at 1369–70; *see also* Mitch Zamoff, *Assessing the Impact of Police Body Camera Evidence on the Litigation of Excessive Force Cases*, 54 GA. L. REV. 1, 18–19 (2019). Big data policing may, of course, come with other potential risks not discussed here. For example, it might have implications for the confrontation rights of criminal defendants. *See, e.g.*, Andrea Roth, *Machine Testimony*, 126 YALE L.J. 1972, 2039–48 (2017) (arguing machine sources may trigger confrontation right, and noting issue of whether machine source could be a "witness" against a defendant "deserves [future] Article-length treatment"); Squitieri, *supra* note 6, at 2013, 2049 (focusing on Confrontation Clause in connection with "mass data collection" by government); Celentino, *supra* note 24, at 1342–51 (arguing Confrontation Clause would "apply to most facial recognition evidence"); *see also* Paul F.

Notwithstanding the ongoing debates on privacy, security, and fairness, big data policing—at least in some form—still appears in ascendancy.[109] Accordingly, it remains important to consider whether United States local law enforcement agencies are adequately prepared for big data policing's rise.

II. METHODOLOGY

This Article creates the Big Data Policing Capacity Index ("BDPCI"), which seeks to measure the inadequacy of big data policing capacity in United States local law enforcement agencies. This Part describes the data and methodology employed, as well as construction of the index.[110]

A. Data

The BDPCI utilizes data from the 2016 Law Enforcement Management and Administrative Statistics study ("2016 LEMAS"), produced by the University of Michigan's Inter-university Consortium for Political and Social Research and authored by the United States Department of Justice's Office of Justice Programs, Bureau of Justice Statistics.[111] The 2016 LEMAS sample was drawn from a 2016 law

Rothstein & Ronald J. Coleman, *Confrontation's Multi-Analyst Problem*, 9 TEX. A&M L. REV. 165 (2021) (providing background on forensic reports and the Confrontation Clause); Paul F. Rothstein & Ronald J. Coleman, *Confronting Memory Loss*, 55 GA. L. REV. 95 (2020) (discussing Confrontation Clause background). Similarly, depending on the specifications of the program, purchasing of equipment such as body cameras might be an expensive undertaking for police departments. *See, e.g.*, Coleman, *supra* note 70, at 1372–74 ("In addition to the initial costs of implementing a [body camera] program, long-term usage requires substantial ongoing expenditures, in particular for data storage and manipulation, as well as for producing a 'courtroom-ready' product. The dollar cost of managing and storing data can be 'staggering' and may run into the hundreds of thousands or millions."); Karson Kampfe, *Police-Worn Body Cameras: Balancing Privacy and Accountability Through State and Police Department Action*, 76 OHIO ST. L.J. 1153, 1178 (2015).

[109] *See generally Predictive Reasonable Suspicion*, *supra* note 3; *see also Legal Risks of Big Data Policing*, *supra* note 9, at 4–6. Some also argue big data policing holds great promise for law enforcement. *See, e.g.*, *id.* at 5 (noting that big data policing's rise "creates real opportunities"). Specifically, proponents may point to possible improvements in law enforcement efficiency and accountability. Brayne, *supra* note 14, at 981; RISE OF BIG DATA POLICING, *supra* note 14, at 28. In terms of efficiency, potential benefits of using big data include quicker investigations, smarter policing, predictive deterrence and lowered crime rates, and an ability to visualize criminal issues in new ways. *Black Data Policing*, *supra* note 11, at 503; Brayne, *supra* note 14, at 981–82 ("It may improve the prediction and preemption of behaviors by helping law enforcement deploy resources more efficiently, ultimately helping prevent and intercept crimes, thus reducing crime rates."). In terms of accountability, utilization of big data offers at least the appearance of more objective and less discretionary decision-making. Brayne, *supra* note 14, at 982 ("[I]t remains an open empirical question to what extent the adoption of advanced analytics will reduce organizational inefficiencies and inequalities, or serve to entrench power dynamics within organizations."). For instance, some may view data-driven policing practices as a partial means of responding to the calls for police reform in the wake of movements like Black Lives Matter. *Id.*; RISE OF BIG DATA POLICING, *supra* note 14, at 28 ("Out of the tension of black lives' frustration with police officers and blue lives' frustration with police administration, the lure of technology to add objectivity to policing and to do more with less began to grow.").

[110] The measurement methodology utilized in this Article is broadly similar to that previously recounted in Ronald J. Coleman, *Measuring Police Body Camera Infrastructure*, 62 SANTA CLARA L. REV. 273, 291 (2022) and Ronald J. Coleman & Ana Vaz, *Law and Multidimensional Measurement*, 44 S. ILL. U. L. J. 253, 255–59 (2020).

[111] *See generally* U.S. DEP'T OF JUSTICE, OFF. OF JUSTICE PROGRAMS, BUREAU OF JUSTICE STAT., LAW ENFORCEMENT MANAGEMENT AND ADMINISTRATIVE STATISTICS (LEMAS), 2016 (2016)

enforcement database that contains 15,810 law enforcement agencies, including sheriffs' offices (3,066 total), primary state police departments (49 total), and local or county police departments (12,695 total).[112] The ultimate sample size was 3,471 agencies, with a survey questionnaire sent to each.[113] Data was collected via hardcopy (25%) or the web (75%), with 2,779 agencies having completed the survey (80% response rate).[114] The present Article focuses solely on the 2,135 local police departments within the final database, which this Article will refer to as the "Local Agencies" or simply the "Agencies."[115]

B. Adjusted Headcount Ratio

The BDPCI utilizes the functional form of James Foster and Sabina Alkire's Adjusted Headcount Ratio (M_0), a multidimensional measure initially developed in an effort to study poverty.[116] The Adjusted Headcount Ratio permits identification of inadequate units via the analysis of insufficiencies in such units across a number of chosen indicators.[117] For instance, the Adjusted Headcount Ratio might be used to

[hereinafter 2016 LEMAS]. Although the present Article cites to the codebook for the 2016 LEMAS study, the data is primarily drawn from the study's Stata dataset. *See* Coleman, *supra* note 70, at 1375 n.76 (taking same approach to similar dataset).

[112] 2016 LEMAS, *supra* note 111, at 5 ("Local police departments and sheriffs' offices were chosen for the 2016 LEMAS using a stratified sample design based on number of full- and part-time sworn officers (part-time officers were counted as 0.5 full-time equivalents) and agency type. The sample represented all general purpose state and local law enforcement agencies in the [United States] with the equivalent of at least one full-time sworn officer with separate samples drawn of local police departments and sheriffs' offices. All 49 primary state law enforcement agencies (state police and highway patrol) and all local departments and sheriffs' offices with 100 or more full-time sworn officers were included. Agencies serving special jurisdictions (such as schools, airports, or parks), or with special enforcement responsibilities (such as conservation laws or alcohol laws), were considered out of scope for the LEMAS."). *Id.* Certain agencies were removed from the database's universe, such as if they lacked sworn staffing counts.

[113] *Id.* at 5–6.

[114] *Id.* at 6 ("Mode used was based on agency preference. Sixty-nine agencies responded using both web and hardcopy; in such cases, the data obtained via web were used.").

[115] *Id.* (noting that, in addition to the local police departments, the final database also includes 44 state agencies and 600 sheriffs' offices, after certain exclusions). The terminology "Agency" or "Local Agency" may also be used. This Article chooses to focus exclusively on the Local Agencies in an effort to aid uniformity in the units measured. *See* Coleman & Vaz, *supra* note 110, at 262 ("We suspect that, given the different nature of the three types of agencies, their capacities should be evaluated with reference to different criteria. Thus, for purposes of our measure, we focused exclusively on the local police departments."); *see also* Coleman, *supra* note 110, at 292 ("Since the positions of the three different types of agencies—county/local, state, and sheriffs' offices—may be distinct and since this Article is most interested in the position of the more local agencies, this Article focuses solely on data from the [] county and local police departments."). Please note, observations with relevant information missing are excluded from the BDPCI and its reported findings. *See generally* 2016 LEMAS, *supra* note 111; *see also* Coleman, *supra* note 110, at 293 n.93. As such, the final number of BDPCI observations—and the final number of studied Agencies—is 2,067, rather than 2,135. *See* Coleman, *supra* note 110, at 293 n.93 (taking similar approach). Further, decomposed results by Agency groups may not be statistically significant and the group-level data is not representative at such group levels. Accordingly, the group-level analysis is illustrative only.

[116] *See generally* Sabina Alkire & James Foster, *Counting and Multidimensional Poverty Measurement*, 95 J. PUB. ECON. 476, 482 (2011); SABINA ALKIRE ET AL., MULTIDIMENSIONAL POVERTY MEASUREMENT AND ANALYSIS (2015).

[117] *See* Sabina Alkire, Jose Manuel Roche & Ana Vaz, *Changes Over Time in Multidimensional Poverty: Methodology and Results for 34 Countries*, 94 WORLD DEV. 232, 233 (2017); Coleman, *supra*

determine which in a set of institutions (the studied units) should be deemed inadequate pursuant to established criteria (which criteria are set by the researcher).[118]

Construction of an Adjusted Headcount Ratio measure may be viewed as a five-step process.[119] First, define the measure's purpose.[120] Second, establish the relevant unit of identification.[121] Third, select statistical indicators and assign indicator weights.[122] Fourth, set indicator insufficiency cutoffs and the overall inadequacy cutoff for studied units.[123] Fifth, calculate the Adjusted Headcount Ratio,

note 110, at 293–94 ("It has been called a 'high-resolution lens' and is particularly suited to informing policy, since it produces an overall measure, may be decomposed for targeting particular subgroups, permits identification of inadequacy drivers, and is suited to both ordinal and cardinal data. The M_0 may be preferable to dashboards and composite indices—other multidimensional measurement techniques which have more commonly been featured in legal scholarship—since these other techniques 'focus on each factor individually, and so fail to reveal how different factors are interdependent.' The M_0, for instance, permits identification of units that experience insufficiency in a larger share of indicators simultaneously.") (citations omitted). The present Article utilizes the terms "inadequate" and "insufficient," instead of "deprived" and "poor" as in the poverty literature, since the present Article adapts the measurement methodology for the law enforcement context. *See* Coleman & Vaz, *supra* note 110, at 255 n.17 (taking similar approach); Coleman, *supra* note 110, at 293 n.94 (same). A unit is deemed "insufficient" when it fails to meet the specified sufficiency threshold for an indicator and "inadequate" when it fails to meet the overall specified inadequacy threshold. *See* Coleman, *supra* note 110, at 293 n.95.

[118] *See* Coleman & Vaz, *supra* note 110, at 254–58.

[119] *Id.* at 256 ("Although there are not necessarily defined 'steps' for creating a measure, for convenience, we present the process of creating a measure as consisting of several distinct steps."); Coleman, *supra* note 110, at 294–96 (discussing five steps).

[120] Coleman & Vaz, *supra* note 110, at 256 ("It is necessary to isolate the phenomenon one is seeking to study and the rationale for such study. For instance, a measure of quality of justice may be created to help guide budget or policy decisions, aid particularly underserved populations, monitor improvements over time, or complement other collected statistics. It is critical to clearly establish the purpose for the measure at the outset, since this decision guides subsequent steps in constructing the measure.") (citations omitted); Coleman, *supra* note 110, at 294 n.99 (noting "[t]his requires determining 'why the measure is being created.'").

[121] *See* Coleman, *supra* note 110, at 294 ("The unit of identification will be the entity under study."); Coleman & Vaz, *supra* note 110, at 256, 257 ("The unit of identification is the entity identified by the measure as inadequate or adequate. For instance, one might seek to analyze cities or states, certain courts, police precincts, or individuals. What should guide the choice of appropriate unit of identification is the measure's purpose.").

[122] *See* Coleman & Vaz, *supra* note 110, at 257 ("When conceptualizing the phenomenon under analysis, a researcher might identify certain variables—such as income, educational attainment, or number of arrests—as most relevant to capture the phenomenon and select such variables as indicators. . . . In order to combine indicators into a measure, it is necessary to weight the indicators. Accordingly, each indicator is given a weight based on the importance of that indicator as compared to other indicators in the measure."); *see also* ALKIRE ET AL., *supra* note 116, at 197 (noting an indicator's assigned weight "reflects the value that an insufficiency in such indicator has for inadequacy, relative to insufficiencies in other indicators."). Indicators might be defined as data elements representing "statistical data for a specified time, place, and other characteristics." Coleman & Vaz, *supra* note 110, at 257 n.27.

[123] *See* Coleman & Vaz, *supra* note 110, at 257 ("Here, we are concerned with what *minimums* must be met. The insufficiency cutoff for an indicator reflects the minimum attainment required so as not to be insufficient in such indicator. The inadequacy cutoff reflects what minimum share of weighted insufficiencies would be necessary to identify a unit as inadequate."); *see also* ALKIRE ET AL., *supra* note 116, at 197. There are a number of approaches for assigning cutoffs, each with respective tradeoffs. *See* Coleman & Vaz, *supra* note 110, at 257 n.33 ("When building a measure . . . one could deem a unit inadequate if such unit were insufficient in at least one indicator (called the union approach) . . . [, which] approach would generally identify a large group of units as inadequate [], potentially including some which are only insufficient in a single indicator and whose performance may not be impaired by such

which is also sometimes referred to as the "multidimensional index."[124] This final step requires identifying inadequate units,[125] and then calculating the "incidence of inadequacy" (that is, the "proportion of inadequate units, also called the headcount ratio"),[126] "intensity of inadequacy" (that is, the "average share of insufficiencies among the inadequate units, also called breadth of inadequacy"),[127] and Adjusted Headcount Ratio (that is, "a measure of overall inadequacy—considering 'incidence' and 'intensity[]'"—that constitutes the relevant inadequacy index of interest).[128]

After calculation of the Adjusted Headcount Ratio, several further analyses may be made.[129] First, estimation of the "uncensored" and "censored" headcount ratios "reveals the pattern of insufficiencies in the population."[130] Uncensored headcount ratios will "summarize the prevalence of the different insufficiencies

insufficiency. An alternative option might be to deem a unit inadequate only if it were insufficient in all indicators (called the intersection approach) . . . [, which] approach generally identifies as inadequate a very small group of units, perhaps leaving out units with many insufficiencies [] whose performance might be hindered even though they are not insufficient in all indicators. Where appropriate, it is helpful to select an inadequacy cutoff between these two extremes, [] potentially permitting one to identify as inadequate only those units with enough insufficiencies as might compromise a unit's performance."); Coleman, *supra* note 110, at 294; ALKIRE ET AL., *supra* note 116, at 152.

[124] Coleman, *supra* note 110, at 294–95 n.107; Coleman & Vaz, *supra* note 110, at 258.

[125] *See* Coleman & Vaz, *supra* note 110, at 258; *see also* Coleman, *supra* note 110, at 295 n.108 ("Suppose you have a population of n units and information on their attainments in d indicators. Let x_{ij} represent the attainment of unit i on indicator j. Assume w_j stands for the relative weight of indicator j, and the weights of the d indicators sum to one: $\sum_{j=1}^{d} w_j = 1$. Then, let z_j reflect the insufficiency cutoff for indicator j, and k denote the overall inadequacy cutoff. Unit i is identified as insufficient in indicator j if its attainment on that indicator is below the respective insufficiency cutoff: $g_{ij} = 1$ if $x_{ij} < z_j$ and $g_{ij} = 0$ if $x_{ij} \geq z_j$. The inadequacy score of unit i, denoted c_i, is the weighted sum of its insufficiencies: $c_i = \sum_{j=1}^{d} w_j g_{ij}$. Unit i is identified as inadequate if its inadequacy score is equal to or greater than the inadequacy cutoff: $c_i \geq k$.").

[126] Coleman, *supra* note 110, at 295 n.109 ("The incidence of inadequacy . . ., denoted by H, is the proportion of inadequate units: $H = \frac{q}{n}$, where q is the number of inadequate units."); Coleman & Vaz, *supra* note 110, at 255 n.14 (noting "incidence of inadequacy" refers to "the percentage of analyzed units that are inadequate").

[127] Coleman, *supra* note 110, at 295 n.110 ("The intensity . . . [, denoted A,] is the average inadequacy score among the inadequate units: $A = \frac{1}{q}\sum_{i=1}^{n} c_i I\,(c_i \geq \ k)$ where $I(.)$ is an identification function that assumes the value one if the condition between parentheses is true for unit i, and zero otherwise."); *see also* Coleman & Vaz, *supra* note 110, at 255 n.15 (noting "intensity of inadequacy" refers to "the average proportion of insufficiencies faced by inadequate units simultaneously").

[128] Coleman, *supra* note 110, at 295 n.111 ("The . . . adjusted headcount ratio[], denoted M_0, reflects the incidence of inadequacy adjusted for the intensity: $M_0 = HA$ or $M_0 = \frac{1}{n}\sum_{i=1}^{n} c_i I(c_i \geq \ k)$."); *see also* Alkire, Roche & Vaz, *supra* note 117, at 233. ("More intuitively, the M_0 can also be expressed as the product of two intuitive partial indices incidence and intensity."); *see also* Coleman & Vaz, *supra* note 110, at 258-59 ("The multidimensional index corresponds to the insufficiencies experienced by inadequate units expressed as a proportion of all possible insufficiencies (if all units were insufficient in all indicators). In being sensitive to both the incidence *and* intensity of inadequacy, the multidimensional index can capture the effects of policies that either reduce the number of inadequate units or improve the position of inadequate units. For example, suppose a policy was successful at reducing the number of insufficiencies experienced by a set of highly inadequate units, but such policy failed to make any inadequate unit adequate. A measure focused only on incidence would fail to reveal the value of such policy, but the multidimensional index would capture it."). A summary table of the five steps for measure construction is available in Coleman & Vaz, *supra* note 110, at 259.

[129] *See generally* Coleman & Vaz, *supra* note 110, at 259–61 (discussing possible calculations and presenting table of same). Calculations not relevant to this Article are not presented.

[130] Coleman, *supra* note 110, at 296; *see also* Coleman & Vaz, *supra* note 110, at 259, 261; *see also* Alkire, Roche & Vaz, *supra* note 117, at 233.

among the population."[131] Censored headcount ratios will "summarize the prevalence of insufficiencies experienced by *only* the inadequate units."[132] Second, calculation of the percentage contribution allows one to "investigate the drivers of inadequacy."[133] This involves breaking the measure down by contribution of each indicator.[134] An indicator having "a large relative contribution could become a policy priority."[135] Third, one might analyze decomposed results.[136] The Adjusted Headcount Ratio is decomposable by subgroups, including by unit location or size.[137] Calculation of subgroup-level results might "permit targeting resources to those groups most in need."[138]

C. Constructing the Big Data Policing Capacity Index

Construction of the BDPCI followed the five-step process described in Part II.B. Those steps were: (1) define the measure's purpose; (2) establish the relevant unit of identification; (3) select statistical indicators and assign indicator weights; (4) set indicator insufficiency cutoffs and the overall inadequacy cutoff for studied units; and (5) calculate the Adjusted Headcount Ratio.[139] In connection with the first two steps, the purpose of the BDPCI was to measure inadequacy in United States local law enforcement agencies, and the Local Agencies were adopted as the relevant unit of identification.

[131] Coleman, *supra* note 110, at 296 n.116 ("The uncensored headcount ratio of indicator *j*, denoted h_j, is the proportion of units that are insufficient in that indicator: $h_j = \frac{1}{n}\sum_{i=1}^{n} g_{ij}$."); Coleman & Vaz, *supra* note 110, at 259.

[132] Coleman, *supra* note 110, at 296 n.117 ("The censored headcount ratio of indicator *j*, denoted $h_j(k) = \frac{1}{n}\sum_{i=1}^{n} g_{ij} I(c_i \geq k)$."); Coleman & Vaz, *supra* note 110, at 259 n.46 ("[B]y definition, for any given indicators, the censored headcount ratio is always smaller than, or equal to, the uncensored headcount ratio.").

[133] Coleman & Vaz, *supra* note 110, at 259; ALKIRE ET AL., *supra* note 116, at 166, 185.

[134] *See* Coleman, *supra* note 110, at 296 n.119 ("Since the Adjusted Headcount Ratio 'can be written as the weighted sum of the censored headcount ratios ($M_0 = \sum_{j=1}^{d} w_j h_j(k)$), the relative contribution of an indicator is obtained by multiplying the indicator's censored headcount ratio by the indicator's weight and dividing by the [Adjusted Headcount Ratio].'"); ALKIRE ET AL., *supra* note 116, at 166, 185; *see also* Coleman & Vaz, *supra* note 110, at 259–60.

[135] Coleman, *supra* note 110, at 296; *see also* Coleman & Vaz, *supra* note 110, at 260 ("For example, suppose a policymaker aims to reduce inadequacy in access to justice, and the created measure reflects that 40% of the inadequacy in the measure derives from an indicator for cost of legal services, while the other eight indicators account for 10% or less each. In such circumstances, a policymaker might most easily reduce inadequacy by taking actions targeted at the cost of legal services, such as improving dissemination of information regarding pro bono legal services.").

[136] *See* Coleman & Vaz, *supra* note 110, at 260; ALKIRE ET AL., *supra* note 116, at 185.

[137] *See* Coleman, *supra* note 110, at 296 n.122 ("Suppose the population can be divided into *m* exhaustive and mutually exclusive subgroups, M_0^l is the [Adjusted Headcount Ratio] for subgroup *l* and v^l denotes the population share of such group. Then, the [Adjusted Headcount Ratio] can be expressed as the weighted sum of the subgroups' [Adjusted Headcount Ratios]: $M_0 = \sum_{l=1}^{m} v^l M_0^l$."); *see also* Coleman & Vaz, *supra* note 110, at 260 ("This feature allows one to analyze the situation of particular subgroups or draw comparisons between the performances of different subgroups.").

[138] Coleman, *supra* note 110, at 297; *see also* Coleman & Vaz, *supra* note 110, at 260 ("Subgroup results may inform policy. . . . Returning to the example of access to justice, suppose the measure reflected that an indicator for linguistic barriers . . . drives inadequacy among a specific ethnic minority. In such circumstances, a legal aid organization seeking to increase access to justice for that ethnic minority might be best served by, for instance, increasing use of interpretation and translation services.").

[139] *See supra* Part II.B. Please note, the fifth step is considered in Part III, along with other calculations.

In connection with steps three and four, seven indicators were selected based on 2016 LEMAS questions. The indicators were chosen based on 2016 LEMAS's functional data and review of the big data policing literature referenced in Part I. Each indicator received an insufficiency cutoff and an equal weight. Table 1 summarizes these adopted parameters.

First, the "Cameras" indicator identified as insufficient Local Agencies not regularly operating any of the enumerated video camera types.[140] The enumerated types were: "Fixed-site surveillance in public areas," "Mobile surveillance," "In patrol cars," "On police officers (e.g., body-worn cameras)," "On weapons," and "On aerial drones."[141]

Second, the "Website" indicator considered insufficient those Local Agencies either not maintaining a website at all or those not using their website to perform any of an enumerated set of tasks.[142] The enumerated tasks were: "Providing direct access to crime statistics/data," "Providing direct access to stop (i.e., motor vehicle or street/field) statistics/data," "Providing direct access to arrest statistics/data," "Enabling citizens to report crimes or problems," "Enabling citizens to ask questions and/or provide feedback," and "Enabling citizens to file complaints about police behavior or actions."[143]

Third, the "Computer Use" indicator deemed insufficient Local Agencies not using computers to perform at least two of a set of four relevant tasks.[144] These relevant tasks were: "Crime analysis (including crime mapping or hotspot identification)," "Social network analysis," "Intelligence gathering," and "Inter-agency information transmission."[145]

Fourth, the "Criminal Incident Reports" indicator found insufficient Local Agencies using paper reports as the primary "method for transmitting criminal incident reports from the field to [the] agency's record management system."[146] Other enumerated primary methods, such as "In-car fixed laptop/tablet" or "Mobile laptop/tablet or phone" were identified as sufficient.[147]

Fifth, the "Technologies Use" indicator considered insufficient Local Agencies not regularly using any of nine relevant technologies.[148] The relevant

[140] Based on question 36 ("During the fiscal year including June 30, 2016, how many of the following types of video cameras were operated by your agency on a REGULAR basis? If none, enter '0'.") in 2016 LEMAS. *See* 2016 LEMAS, *supra* note 111, at 321–31 (capitalized in original survey).

[141] *See id.* Of course, regular usage of video cameras does not necessarily imply that captured footage is analyzed or that camera use is optimized. A similar caveat may apply to other selected indicators.

[142] Based on question 37 ("As of June 30, 2016, did your agency maintain a website for any of the following?") in 2016 LEMAS. *See id.* at 332–36.

[143] *See id.*

[144] Based on question 39 ("As of June 30, 2016, did your agency use computers for any of the following functions?") in 2016 LEMAS. *See id.* at 339–42.

[145] *See id.* Question 39 also sought information on computer usage for "Automated booking," but this was not adopted as part of the indicator, since it seems comparatively less relevant for big data policing. *See id.*; *see supra* Part I.

[146] Based on question 40 ("As of June 30, 2016, what was the PRIMARY method for transmitting criminal incident reports from the field to your agency's record management system? Mark only one response.") in 2016 LEMAS. *See* 2016 LEMAS, *supra* note 111, at 342–43 (capitalized in original survey).

[147] *See id.*

[148] Based on question 41 ("As of June 30, 2016, did your agency use any of the following technologies on a REGULAR basis?") in 2016 LEMAS. *See id.* at 343–50 (capitalized in original survey).

technologies were: “Automated Fingerprint Identification System (“AFIS”),” “Facial recognition,” “License plate readers (“LPR”),” “Infrared (thermal) imagers,” “Stolen vehicle tracking (e.g., LoJack),” “Gunshot detection (e.g., Shotspotter),” “Firearm tracing (e.g., eTrace),” “Ballistic imaging (e.g., NIBN, IBIS),” and “Global Positioning System (“GPS”).”[149]

Sixth, the “In-Field Computers” indicator deemed insufficient Local Agencies not using in-field computers.[150] It was initially hoped that this indicator could include information regarding different uses of in-field computers; however, such information was ultimately excluded due to missing observations.[151]

Seventh, and finally, the “Computerized Files” indicator identified as insufficient Local Agencies not maintaining their own computerized files with at least nine of the enumerated seventeen types of information.[152] The enumerated types of information were: “Arrests,” “Calls for service,” “Civilian complaints,” “Criminal incident reports,” “Firearms recovered, seized or found,” “Gangs,” “Informants,” “Intelligence related to terrorist activity,” “Motor vehicle stops,” “Motor vehicle accidents,” “Pawn shop data,” “Protective orders,” “Stolen property,” “Street/field stops,” “Use of force incidents,” “Video surveillance,” and “Warrants.”[153]

[149] *See id.* Question 41 also asked for data regarding usage of “Electrical/engine disruption” and “Tire deflation devices,” but these were excluded from the indicator, since they appeared comparatively less relevant for big data policing. *See supra* Part I.

[150] Based on question 42 (“As of June 30, 2016, did your agency’s field/patrol officers have direct access to the following types of information using in-field vehicle-mounted or mobile computers?”) in 2016 LEMAS. *See* 2016 LEMAS, *supra* note 111, at 350–56.

[151] *See id.*

[152] Based on question 44 (“As of June 30, 2016, did your agency maintain its own computerized files with any of the following information?”) in 2016 LEMAS. *See id.* at 357–67.

[153] *See id.* Here, the assumption is that there is a share or number of types of information that should be digitized. This indicator does not specify exactly what types of information must be digitized, since that might vary across Agencies and types of investigations.

Table 1 – BDPCI: Indicators, Cutoffs, and Weights[154]

Indicator	Cutoff ("Insufficient if")	Weight (%)
Cameras	Does not regularly operate **any** of following video camera types: - "Fixed-site surveillance in public areas" - "Mobile surveillance" - "In patrol cars" - "On police officers (e.g., body-worn cameras)" - "On weapons" - "On aerial drones"	14.29
Website	Does not maintain website, or does not use its website to perform **any** of following tasks: - "Providing direct access to crime statistics/data" - "Providing direct access to stop (i.e., motor vehicle or street/field) statistics/data" - "Providing direct access to arrest statistics/data" - "Enabling citizens to report crimes or problems" - "Enabling citizens to ask questions and/or provide feedback" - "Enabling citizens to file complaints about police behavior or actions"	14.29
Computer Use	Does not use computers to perform at least **two** of following tasks: - "Crime analysis (including crime mapping or hotspot identification)" - "Social network analysis" - "Intelligence gathering" - "Inter-agency information transmission"	14.29
Criminal Incident Reports	Uses paper reports as primary method for transmittal of criminal incident reports from field to its record management system	14.29
Technologies Use	Does not regularly use **any** of following technologies: - "Automated Fingerprint Identification System (AFIS)" - "Facial recognition" - "License plate readers (LPR)" - "Infrared (thermal) imagers" - "Stolen vehicle tracking (e.g., LoJack)" - "Gunshot detection (e.g., Shotspotter)" - "Firearm tracing (e.g., eTrace)" - "Ballistic imaging (e.g., NIBN, IBIS)" - "Global Positioning System (GPS)"	14.29
In-Field Computers	Does not make use of in-field computers	14.29
Computerized Files	Does not maintain own computerized files with at least **nine** of following types of information: - "Arrests" - "Calls for service" - "Civilian complaints" - "Criminal incident reports" - "Firearms recovered, seized or found" - "Gangs" - "Informants" - "Intelligence related to terrorist activity" - "Motor vehicle stops" - "Motor vehicle accidents" - "Pawn shop data" - "Protective orders" - "Stolen property" - "Street/field stops" - "Use of force incidents" - "Video surveillance" - "Warrants"	14.29

[154] Quotations in the table are drawn from questions described above in 2016 LEMAS. *See id.* at 321–36, 339–50, 357–67; Part II.C.

III. EMPIRICAL RESULTS

This Part presents findings from the BDPCI analysis. Study limitations and sensitivity are also discussed.[155]

A. Findings

Analysis of the BDPCI provides an overall view of big data policing inadequacy, as well as a more granular picture of factors driving inadequacy.[156] To start, estimations are made of the incidence, intensity, and the BDPCI (that is, the Adjusted Headcount Ratio).[157] Table 2 presents these values with respective 95% confidence intervals. The incidence reveals that 37% of the Agencies are inadequate.[158] The intensity reflects that inadequate Agencies are insufficient, on average, in 59% of indicators—meaning in more than four indicators.[159] The BDPCI is 0.221, reflecting that "the total insufficiencies experienced by inadequate agencies correspond" to 22% of all possible insufficiencies.[160] These aggregated values present a broad summary of big data policing inadequacy among Agencies, and policymakers deciding on countrywide policy might find them particularly useful.[161]

Table 2 – Incidence, Intensity, and BDPCI

Cutoff (k) = 3 out of 7	**Value**	**Confidence Interval (95%)**	
Incidence (H, %)	37.2	35.2	39.2
Intensity (A, %)	59.4	58.1	60.6
BDPCI (M_0)	0.221	0.209	0.233

Next, it is helpful to analyze the types of insufficiencies more prevalent among Agencies.[162] This begins with analysis of uncensored and censored headcount ratios.[163] Figure 1 reflects the uncensored headcount ratios in the lighter color and the censored headcount ratios in the darker color. As previously noted, uncensored headcount ratios "summarize the prevalence of the different insufficiencies among the population," while censored headcount ratios "summarize the prevalence of insufficiencies experienced by *only* the inadequate units."[164] The uncensored headcount ratios show that, for example, half of Agencies are insufficient in Website, but only 17% are insufficient in Cameras. As such, a policymaker seeking to decrease inadequacy in big data policing capacity across Agencies might consider helping Agencies create websites rather than encouraging Agencies to purchase or utilize

[155] Presentation of this Article's empirical results largely follows the format set out in Coleman & Vaz, *supra* note 110, at 265–71 and Coleman, *supra* note 110, at 300–15.

[156] *See* Coleman & Vaz, *supra* note 110, at 265.

[157] Coleman, *supra* note 110, at 300.

[158] Coleman & Vaz, *supra* note 110, at 265.

[159] Coleman, *supra* note 110, at 300.

[160] *See* Coleman & Vaz, *supra* note 110, at 265.

[161] Coleman, *supra* note 110, at 300.

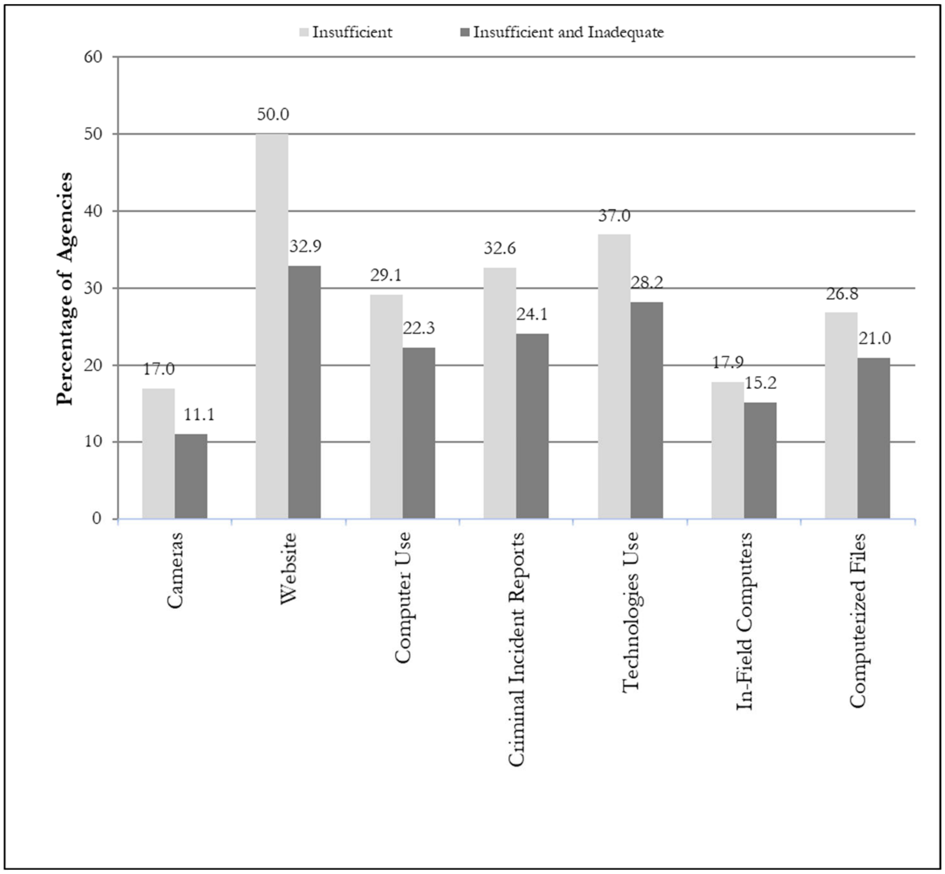
[162] Coleman & Vaz, *supra* note 110, at 266.

[163] Coleman, *supra* note 110, at 300–01.

[164] *Id.* at 296; Coleman & Vaz, *supra* note 110, at 259.

cameras. When viewing the censored headcount ratios, for instance, although Website still remains the most prevalent type of insufficiency, Technologies Use is quite close and might also be worthy of policymaker attention.

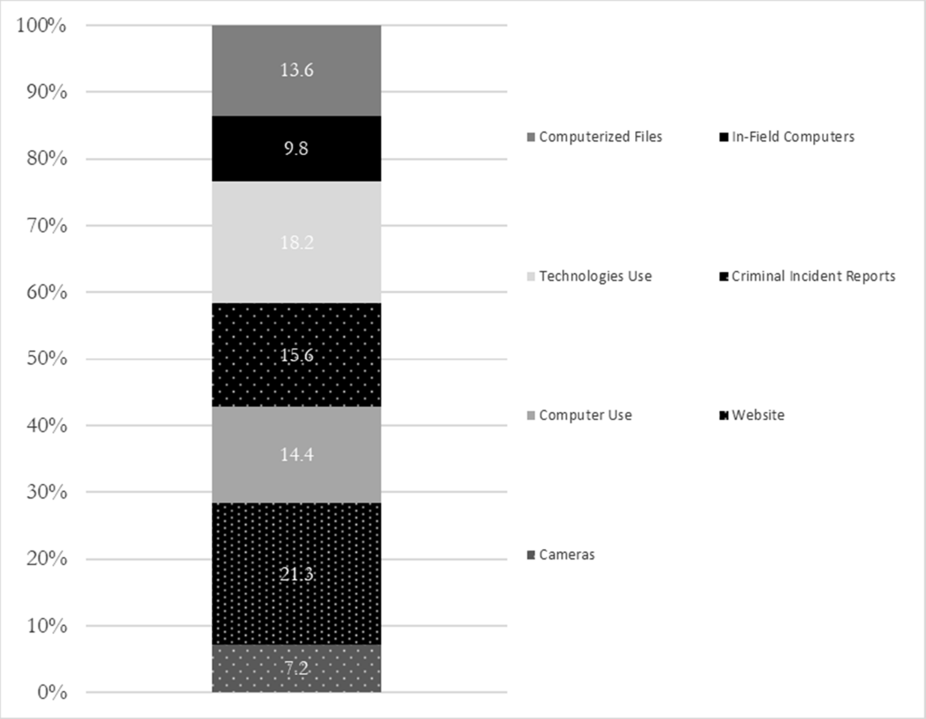
Figure 1 – Uncensored and Censored Headcount Ratios



Another mechanism for isolating inadequacy drivers among Agencies is calculating the relative contribution of indicators to the BDPCI.[165] These values are summarized in Figure 2. It is again apparent that the Website and Technologies Use indicators are key inadequacy drivers among Agencies. Each of these two indicators accounts for approximately 1/5 of the overall index. On the other hand, the Cameras and In-Field Computers indicators do not appear to be important inadequacy drivers, and these areas are perhaps less worthy of policymaker attention.

[165] Coleman, *supra* note 110, at 302; Coleman & Vaz, *supra* note 110, at 267.

Figure 2 – Percentage Contribution of Indicators to BDPCI[166]



The BDPCI analyses so far have focused on all Agencies, but countrywide averages might "mask significant differences *across* groups of agencies."[167] The BDPCI could also reveal the positions of Agency groups.[168] Figures 3 and 4 reflect the BDPCI and incidence, respectively, by Local Agency size.[169] Figures 3 and 4 suggest that, on average, smaller Agencies suffer much greater big data policing capacity shortfalls than larger Agencies. For instance, more than 60% of Agencies with one to nine officers are inadequate, while this percentage reduces to less than 25% among Agencies with ten to twenty-four officers and to almost 0% among Agencies with fifty or more officers. Although it might be assumed that larger agencies would have better big data policing capacity, it would be useful for policymakers to have actual evidence of that to support resource allocation.[170]

[166] In Figure 2, and in several other figures in this Article, percentages may not always sum to 100% due to rounding.

[167] Coleman & Vaz, *supra* note 110, at 268.

[168] *Id.* at 268–69 ("[W]e can use the framework to zoom in on the situation of particular groups.").

[169] Figures 3 and 4 also reflect 95% confidence intervals.

[170] Coleman, *supra* note 110, at 303 (noting "economies of scale" might help larger agency body camera infrastructure).

Figure 3 – BDPCI by Local Agency Size

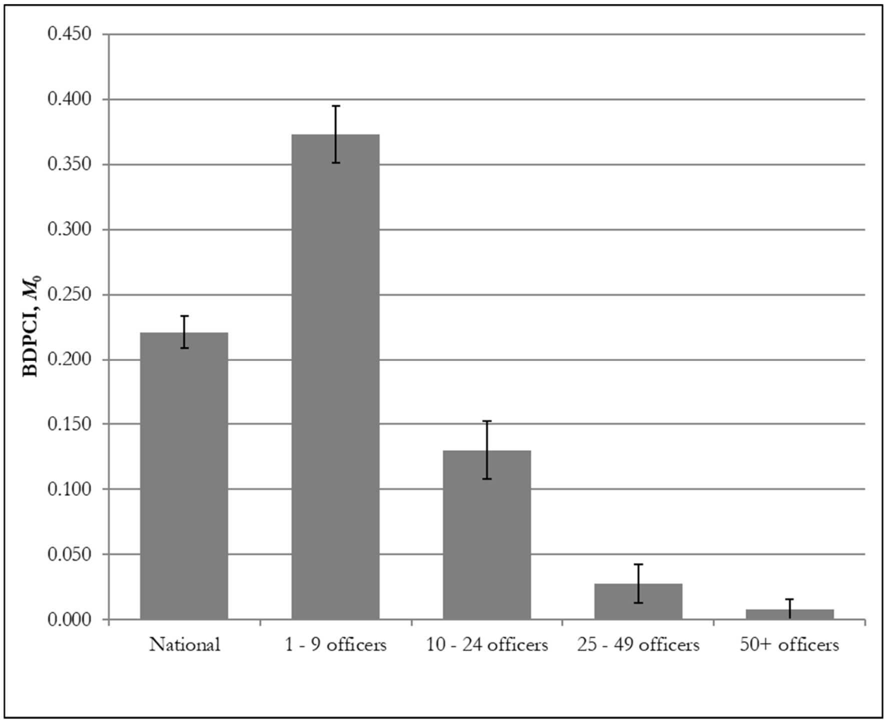
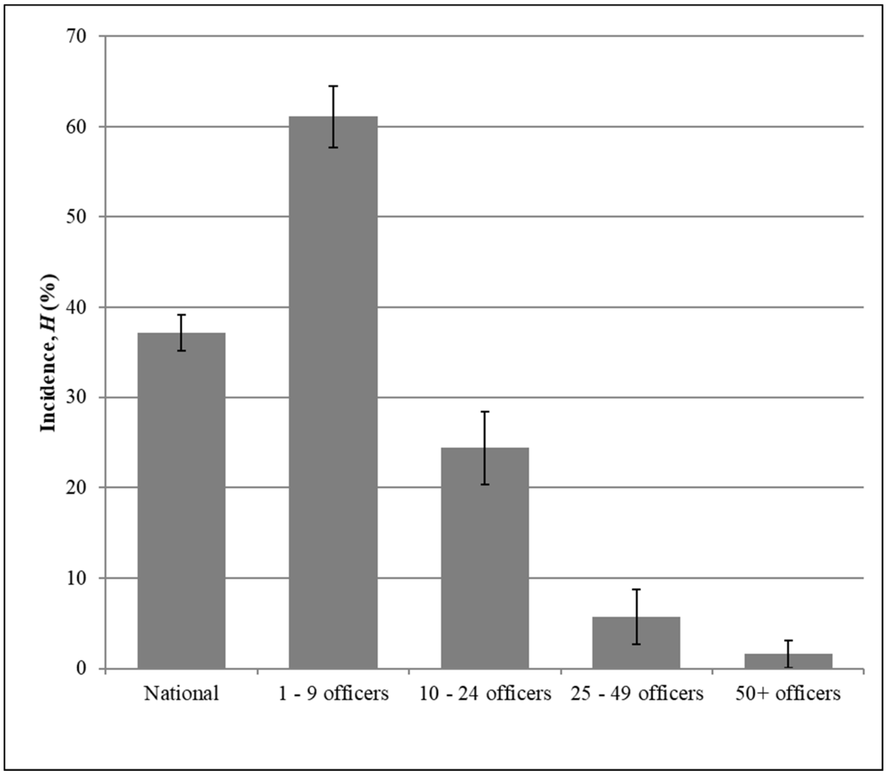


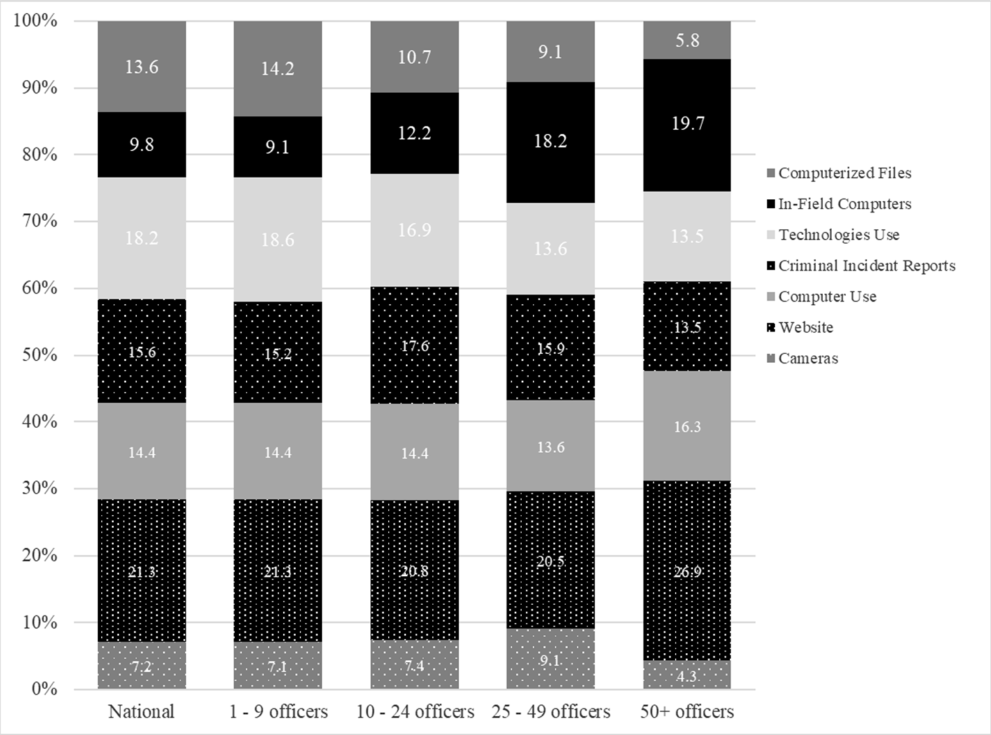
Figure 4 – Incidence of Inadequacy by Local Agency Size



Percentage contribution of indicators to the BDPCI by Local Agency size may also be calculated.[171] These values appear in Figure 5. As Figure 5 shows, there is a decent degree of consistency in drivers across Local Agency size groups, with the Website indicator again being a key inadequacy driver across these groups. Policymaker investment in improving the Technologies Use, Criminal Incident Reports, and Computer Use indicators might also help across Agency groups. However, there are still variations across groups that a policymaker could consider. For instance, policymaker investment in computerized files would be comparatively more helpful for Agencies with one to nine officers (for whom the Computerized Files indicator accounts for 14% of the measure) than it would for Agencies with fifty or more officers (for whom the Computerized Files indicator accounts for only 6% of the measure). The opposite might be true in the case of policymaker investment in improving the In-Field Computers indicator.

[171] *Id.* at 305.

Figure 5 – Percentage Contribution of Indicators to BDPCI by Local Agency Size



Big data policing capacity may also vary based on location of Agencies.[172] Table 3 presents an illustrative ranking of U.S. states by each state's BDPCI, and Figure 6 provides an illustrative visualization of U.S. states by BDPCI.[173] The gray line in Table 3—running along Montana—reflects where the BDPCI national average falls within the state ranking. Six states in the ranking—Alaska, Arizona, Delaware, Hawaii, Rhode Island, and Utah—have no inadequacy among Agencies sampled, but each has a low observation count. New Jersey, California, and Massachusetts each have a decent number of observations and remain well above most other states in big data policing capacity. Toward the bottom of Table 3, states such as Mississippi and Louisiana may need to make heavier investments in improving big data policing capacity in order to catch up with other states. Figure 6 permits visual comparison of state capacity performance. A local policymaker in Texas, for instance, might be much more comfortable with big data policing capacity in her state than a comparable policymaker in Missouri.

[172] *Id.* at 306.

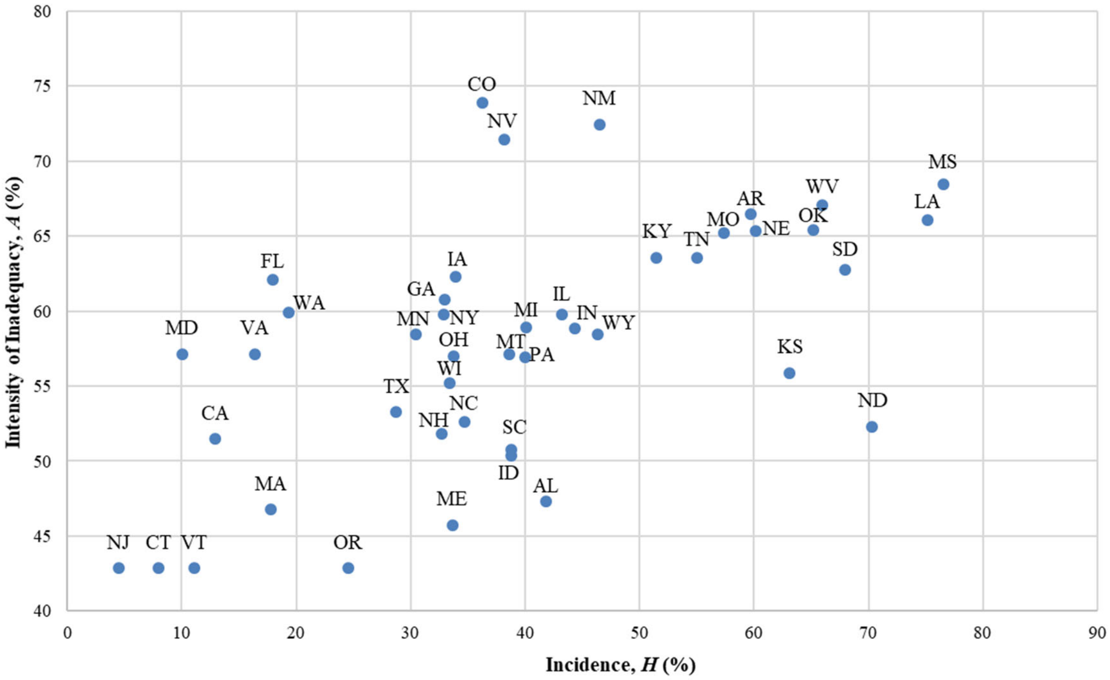
[173] The terminology "illustrative" is utilized for the ranking and visualization as a reminder that 2016 LEMAS was not constructed so as to be representative at a state level. *See generally* 2016 LEMAS, *supra* note 111.

Table 3 – Ranking of U.S. States by BDPCI[174]

U.S. State	No. Agencies Sampled	BDPCI (M_0)	Incidence (H, %)	Intensity (A, %)
National	2067	0.221	37.2	59.4
AK	4	0.000	0.0	0.0
AZ	11	0.000	0.0	0.0
DE	6	0.000	0.0	0.0
HI	1	0.000	0.0	0.0
RI	11	0.000	0.0	0.0
UT	14	0.000	0.0	0.0
NJ	82	0.019	4.5	42.9
CT	15	0.034	8.0	42.9
VT	12	0.047	11.1	42.9
MD	16	0.057	10.0	57.1
CA	48	0.067	12.9	51.5
MA	57	0.083	17.8	46.8
VA	24	0.094	16.4	57.1
OR	23	0.105	24.5	42.9
FL	42	0.111	17.9	62.1
WA	27	0.116	19.3	59.9
TX	129	0.153	28.7	53.3
ME	19	0.154	33.7	45.7
NH	33	0.170	32.7	51.8
MN	59	0.178	30.5	58.4
NC	67	0.183	34.7	52.6
WI	88	0.184	33.4	55.2
OH	112	0.192	33.7	57.0
ID	7	0.195	38.8	50.4
NY	68	0.196	32.8	59.7
SC	25	0.197	38.8	50.8
AL	39	0.198	41.8	47.3
GA	52	0.200	32.9	60.8
IA	40	0.211	33.9	62.3
MT	7	0.221	38.6	57.1
PA	155	0.228	40.0	56.9
MI	78	0.236	40.1	58.9
IL	139	0.258	43.2	59.8
IN	78	0.261	44.3	58.8
CO	27	0.268	36.3	73.9
WY	10	0.271	46.3	58.4
NV	3	0.273	38.2	71.4
KY	46	0.327	51.5	63.6
NM	8	0.337	46.5	72.5
TN	37	0.350	55.0	63.6
KS	37	0.353	63.1	55.9
ND	6	0.368	70.3	52.3
MO	81	0.374	57.3	65.2
NE	21	0.393	60.1	65.3
AR	44	0.397	59.7	66.4
OK	59	0.426	65.2	65.4
SD	17	0.427	68.0	62.8
WV	23	0.442	65.9	67.1
LA	37	0.497	75.2	66.1
MS	24	0.524	76.6	68.5

[174] This ranking focuses exclusively on the fifty U.S. states. The District of Columbia was considered for inclusion in the ranking but ultimately excluded because it had no sampled Agencies.

Figure 6 – BDPCI by U.S. State[175]



Continuing with analysis of big data policing capacity by location of Agencies, Figures 7, 8, and 9, depict the BDPCI, incidence, and intensity, respectively, in the largest ten U.S. states by population.[176] Each figure reflects states in descending order of the depicted index.[177] As these figures show, California and Texas appear to perform comparatively well in terms of BDPCI, incidence, and intensity. Similarly, Illinois generally appears to perform the poorest across these three metrics. However, there are differences across Figures 7, 8, and 9. For instance, Florida is an interesting case of variance, in that it has the second lowest BDPCI and incidence among the ten states but the highest intensity. Florida may have a comparatively lower proportion of inadequate Local Agencies but a comparatively greater share of indicator insufficiencies among its inadequate Local Agencies.[178]

[175] Figure 6 excludes Alaska, Arizona, Delaware, Hawaii, Rhode Island, and Utah since, as reflected in Table 3, these states have zero values for BDPCI, Incidence, and Intensity.

[176] *See* Coleman, *supra* note 110, at 306–08; *US States - Ranked by Population 2022*, WORLD POPULATION REV. (June 16, 2022), https://worldpopulationreview.com/states [https://perma.cc/VC82-E5RD]. These three figures also reflect 95% confidence intervals. Please note, the standard errors in state-level results may be much higher, and as a result, the confidence intervals may be very wide.

[177] Coleman, *supra* note 110, at 306 n.141 (noting "[t]he inability to determine statistical significance is primarily due to the small size of the states' samples").

[178] *Id.* at 307.

Figure 7 – BDPCI in Largest Ten U.S. States

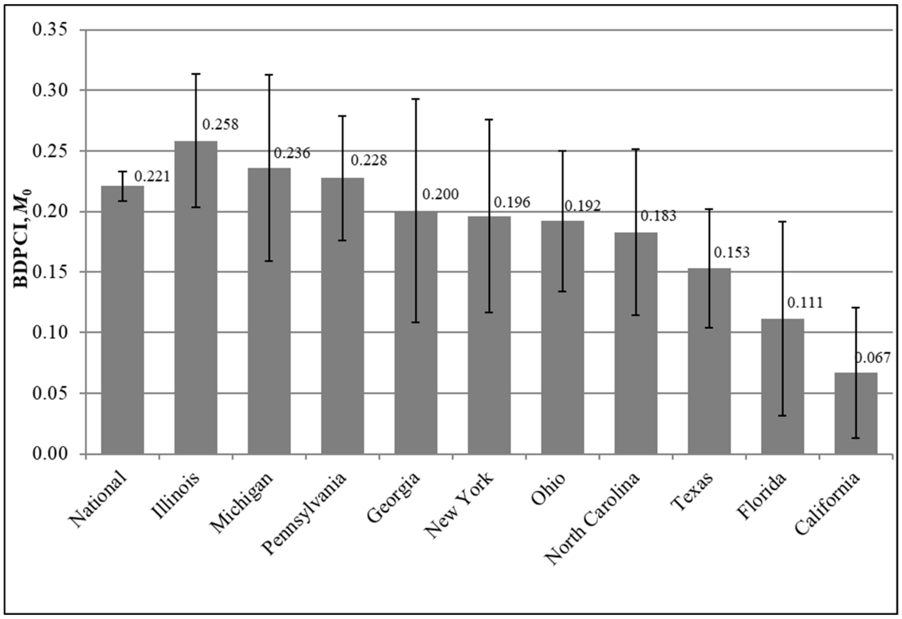


Figure 8 – Incidence of Inadequacy in Largest Ten U.S. States

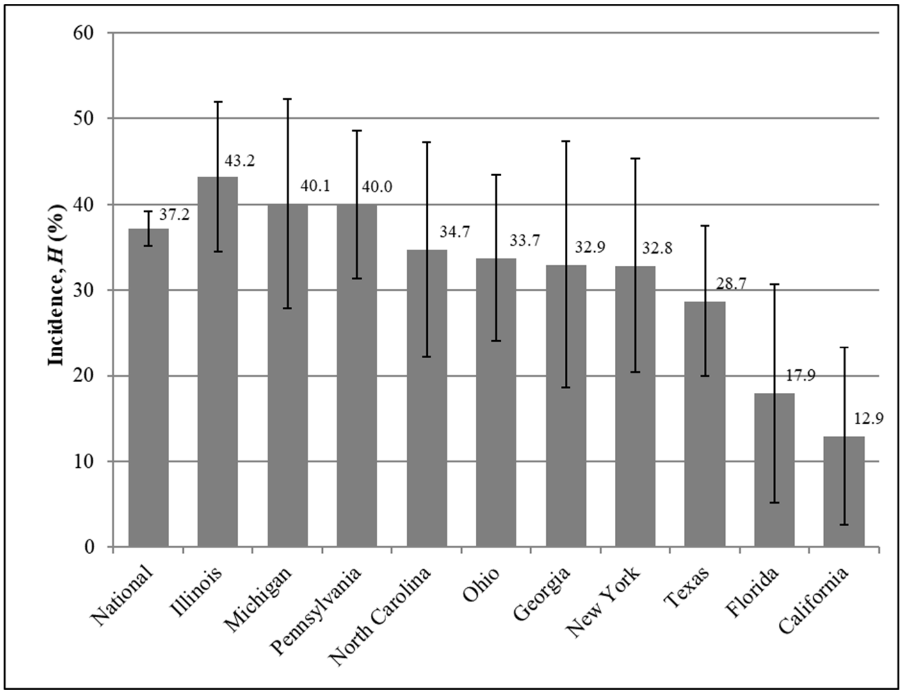
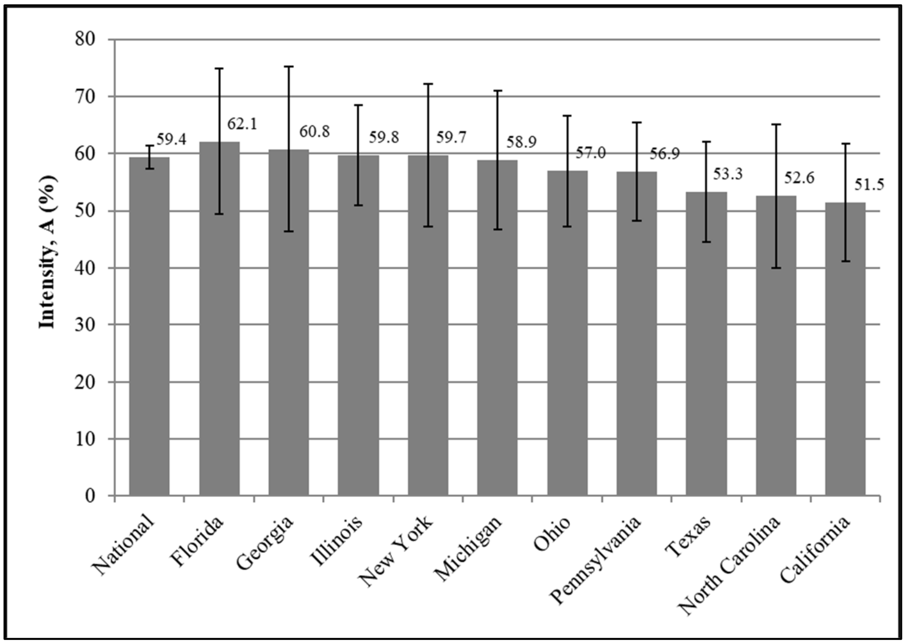


Figure 9 – Intensity of Inadequacy in Largest Ten U.S. States

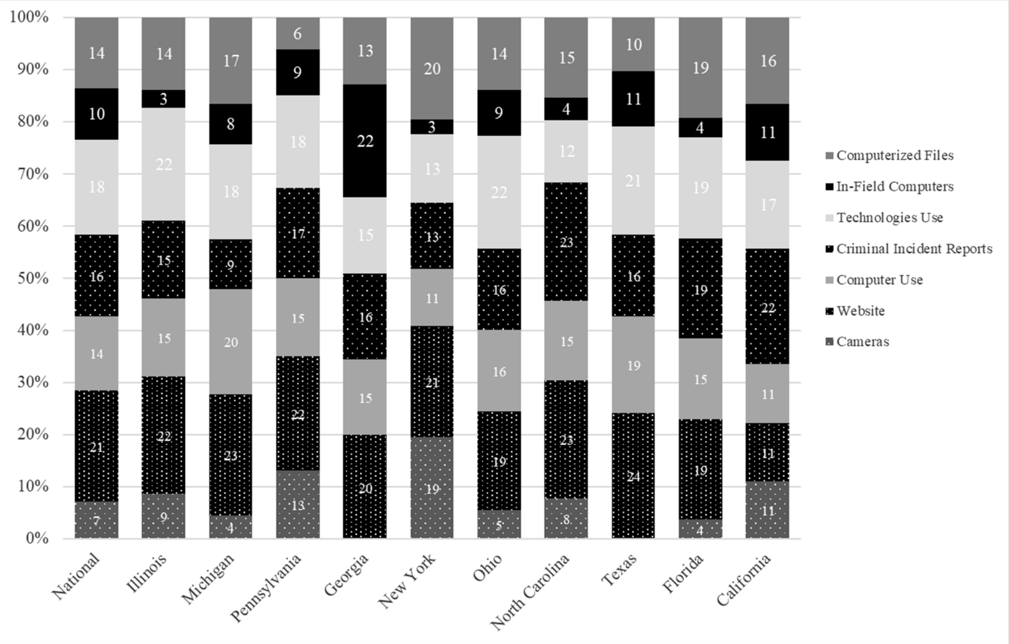


The relative contributions of indicators to the BDPCI in the largest ten U.S. states are depicted in Figure 10.[179] This figure might help a policymaker efficiently allocate resources in a given state to improve big data policing capacity.[180] For instance, a Texas policymaker might be interested to know that in Texas, the Website, Computer Use, and Technologies Use indicators make up the vast majority of the measure. The Texas policymaker might be able to effectively invest in those areas to increase capacity. Figure 10 also shows that the Cameras indicator plays no role in inadequacy in Texas or Georgia, but it constitutes nearly one-fifth of the measure in New York. This means that investments in cameras might make sense for a policymaker in New York but not for one in Georgia or Texas. Policymakers might also consider that the In-Field Computers indicator makes up a relatively small portion of the BDPCI in Florida, North Carolina, New York, and Illinois but a much larger relative portion in Georgia. Finally, investment in improving the Website indicator could be a good general priority across most of these ten states, but such investment might be comparatively less of a priority in California.

[179] Coleman & Vaz, *supra* note 110, at 269–70.

[180] Coleman, *supra* note 110, at 309.

Figure 10 – Percentage Contribution of Indicators to BDPCI in Largest Ten U.S. States



The BDPCI's aggregated and group-level results may be used by policymakers to view overall or decomposed Agency inadequacy in big data policing capacity, isolate drivers of inadequacy, and identify priority areas for potential investment.[181] The BDPCI framework is also very flexible, such that stakeholders could adapt the measure to their needs by opting for different indicators, weights, or cutoffs.[182]

B. Limitations & Sensitivity

This Article's findings are subject to certain limitations, and two are emphasized here.[183] First, data limitations exist.[184] Such limitations restricted available variables for adaptation as indicators.[185] For example, the utilized dataset did not permit in-depth measurement of certain items of interest such as algorithms,

[181] *Id.* at 309.

[182] Coleman, *supra* note 110, at 309.

[183] Several limitations are similar to those previously noted in Coleman, *supra* note 110, at 310 and Coleman, *supra* note 70, at 1389–90.

[184] Coleman, *supra* note 70, at 1389.

[185] *See* Coleman, *supra* note 110, at 310; Alkire, Roche & Vaz, *supra* note 117, at 232 (describing related measure in poverty context as "data constrained").

different types of predictive policing, and selected other emerging technologies.[186] Such limitations also restricted the type of analyses that could be performed. For example, as previously noted, the utilized dataset was not representative at the state level, which limited relevant state comparisons.[187] Second, measurement limitations exist.[188] The measure utilized—like other related measures—may ultimately fail to measure what it purports to.[189] Specifically, normative choices were involved in its creation.[190] Construction of the measure necessitated choosing indicators and assigning relevant weights and cutoffs, each of which is subjective to some degree.[191] As such, sub-optimal values and parameters might have been incorporated.[192]

Since selecting parameters implies normative choices, it can be important to check the sensitivity of such chosen parameters, and two illustrative examples are provided here.[193] First, Figures 11 and 12 depict the BDPCI and incidence, respectively, by Local Agency size for varied inadequacy cutoffs.[194] These figures suggest that distinctions across Agencies of different sizes may be relevant at various possible cutoffs.[195]

[186] *See generally* 2016 LEMAS, *supra* note 111. Note that, "this is in no way a criticism of the data collection, as there may be good reason to limit the number of questions posed, so as to avoid a burdensome and lengthy survey." Coleman, *supra* note 70, at 1389 n.128; *see also* Scott W. Phillips et al., *The Impact of General Police Officer Outlooks on Their Attitudes Toward Body-Worn Cameras*, 43 POLICING: AN INT'L J. 451, 462 (2020). In some instances, more observations for existing questions would have also helped. *See* Coleman, *supra* note 110, at 310.

[187] Other data limitations from 2016 LEMAS might also limit findings here. *See* 2016 LEMAS, *supra* note 111, at 5–8; Coleman, *supra* note 110, at 310.

[188] *See* Coleman, *supra* note 110, at 310; Coleman, *supra* note 70, at 1389.

[189] *See* Coleman, *supra* note 110, at 310; Alkire, Roche & Vaz, *supra* note 117, at 232 (describing related measure in poverty context as "imperfect").

[190] Coleman, *supra* note 110, at 310.

[191] Coleman & Vaz, *supra* note 110, at 270; Coleman, *supra* note 110, at 310.

[192] *See* Coleman, *supra* note 110, at 310 ("Similarly, the indicators were constructed from the responses of the Local Agencies only, so the indicators may merely reflect the subjective beliefs of the responding law enforcement employees, rather than that of the full agencies or other important societal stakeholders."); *see also* Jordan C. Pickering, *Officers' Perceptions Regarding the Unexpected Effects of Body-Worn Cameras*, 43 POLICING: AN INT'L J. 390, 400 (2020) (noting a similar limitation).

[193] Coleman, *supra* note 110, at 310–11; *see also* Coleman & Vaz, *supra* note 110, at 270 ("Since so many normative decisions are required when constructing a measure, it is important to analyze how sensitive the calculated results are to changes in the selected parameters.").

[194] Coleman, *supra* note 110, 310–12.

[195] It should be noted that, at higher cutoffs, certain groups may be effectively "tied" at zero.

Figure 11 – BDPCI by Local Agency Size for Varied Inadequacy Cutoffs

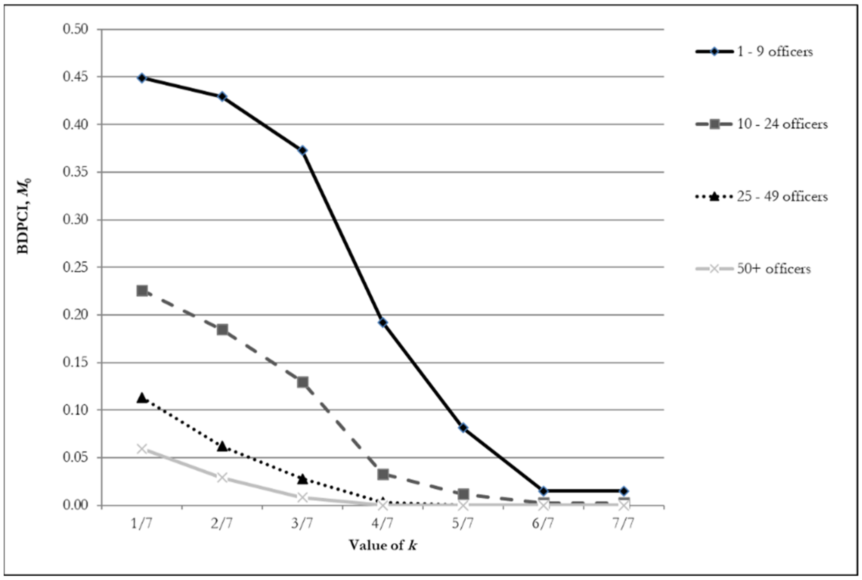
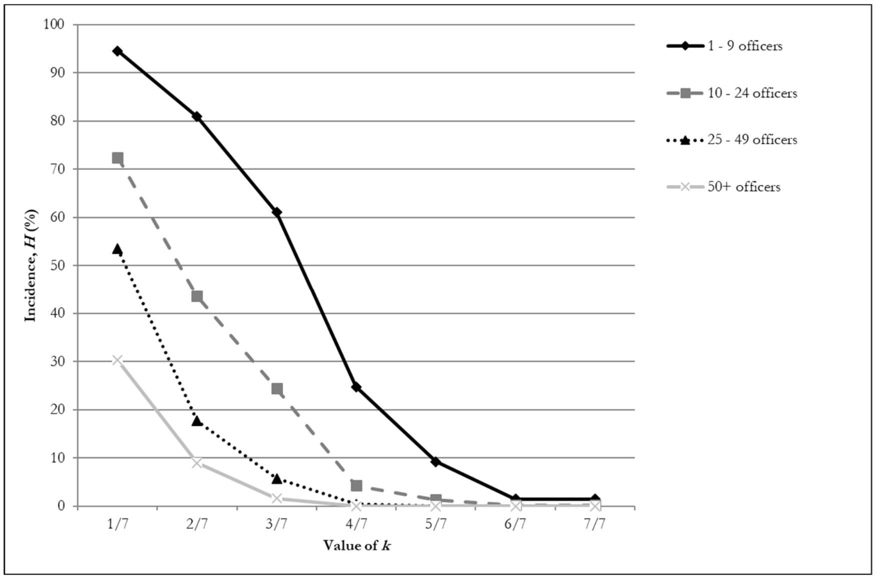
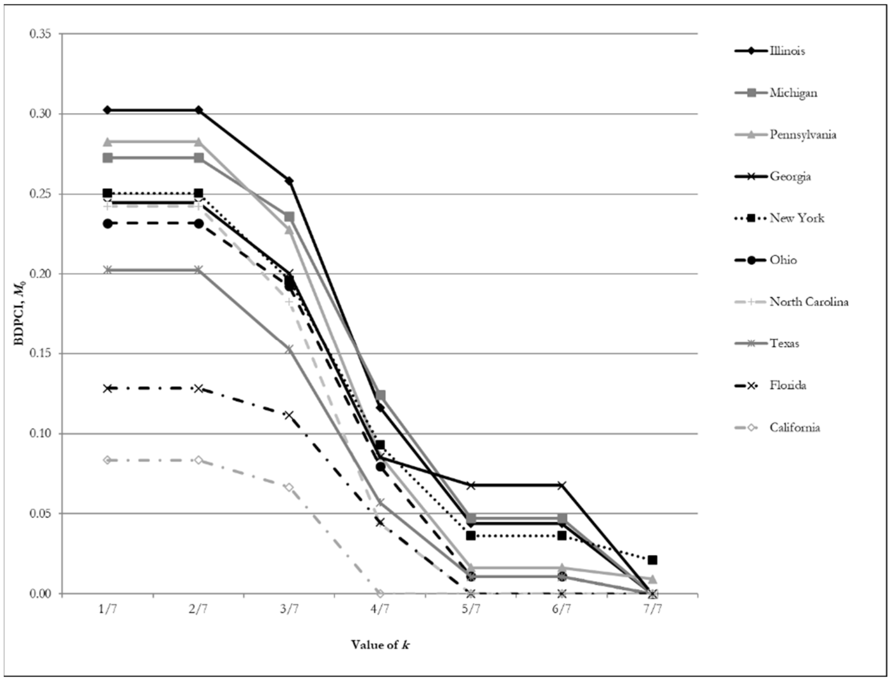


Figure 12 – Incidence of Inadequacy by Local Agency Size for Varied Inadequacy Cutoffs



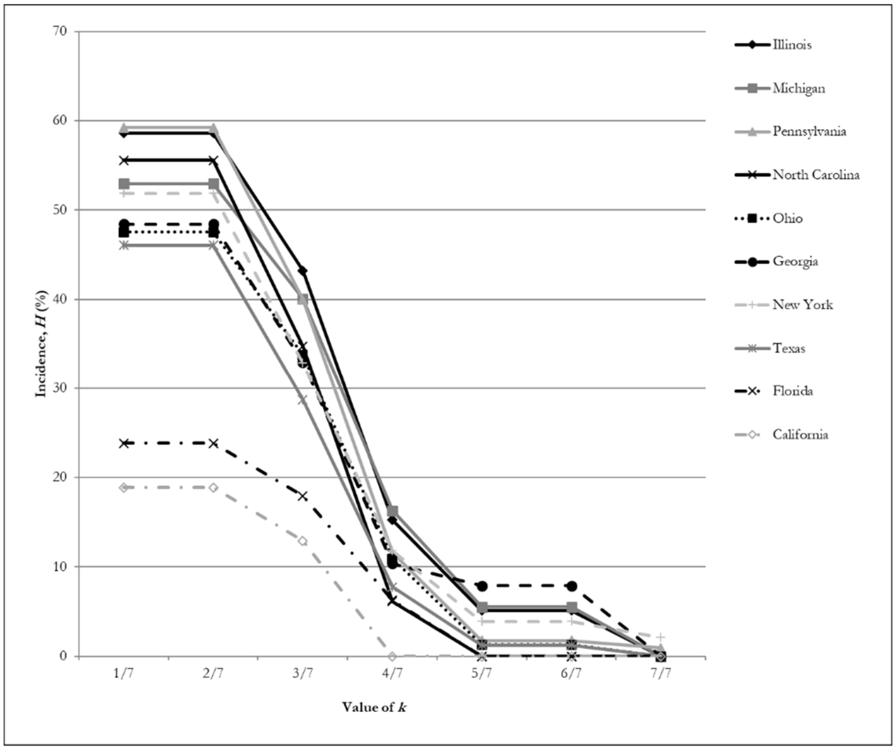
Second, Figures 13 and 14 depict the BDPCI and incidence, respectively, in the largest ten U.S. states for varied inadequacy cutoffs. Here, it is possible to see certain sensitivity to selected cutoff. Florida is an interesting case of sensitivity with respect to BDPCI and Georgia is an interesting case with respect to incidence. A policymaker should consider state-level results with this sensitivity in mind and could also consider calculating the measure with several cutoffs if helpful and feasible.[196]

Figure 13 – BDPCI in Largest Ten U.S. States for Varied Inadequacy Cutoffs



[196] Coleman, *supra* note 110, at 312.

Figure 14 – Incidence of Inadequacy in Largest Ten U.S. States for Varied Inadequacy Cutoffs



CONCLUSION

This Article presents and analyzes a novel multidimensional measure of U.S. local law enforcement big data policing capacity: the BDPCI. Notwithstanding important limitations, analysis of the BDPCI offers an overall picture of big data policing inadequacy across more than 2,000 local agencies, an illustration of how differences between Agency groups might be investigated, and information on the factors driving inadequacy.

There are numerous areas for future research, including certain areas specifically discussed here. First, it would be useful for researchers to create further large-N datasets with additional questions relating to big data policing.[197] For instance, how many agencies are using different types of predictive policing? How many are conducting training on the use of facial recognition technology with body camera footage? How many have big data policing-specific policies and procedures? How many are using artificial intelligence, machine learning, and/or algorithms, and can this data be further decomposed? Are agencies happy with these new techniques and technologies? Even better than additional questions in a larger survey would be

[197] *Id.* at 314.

comprehensive, large-N datasets specifically targeted to big data policing.[198] A greater supply of data—and more targeted data—would allow for better measurement indicators.[199] Second, new datasets could be constructed that seek answers from non-law enforcement personnel, such as from the public or officers of the courts.[200] For example, perhaps researchers could gather additional nationwide empirical data on non-law enforcement perceptions of big data policing's perceived benefits and risks.[201] Third, it would be nice to have additional nationwide data that is representative at the state level. Such data would help confirm or refute illustrative state-level results drawn from the BDPCI. Fourth, additional measurement of big data policing capacity could be conducted, ideally using a variety of measurement methodologies.[202] Further measurement would help either confirm or refute this Article's findings.[203] Fifth, if local policymakers or stakeholders take actions based on the BDPCI, impact evaluations could be conducted on such programs to help test their utility.[204] Sixth, and finally, given sufficient data, researchers could seek to isolate changes in big data policing capacity over time.[205] For example, have local law enforcement agencies been building capacity since data was collected for the 2016 LEMAS dataset, and will big data policing's continued ascendancy be accompanied by advances in capacity in the years ahead?

Chief Justice Roberts, in *Carpenter v. United States*, noted "the seismic shifts in digital technology" and suggested that the cell phone had become "almost a 'feature of human anatomy.'"[206] As the technological revolution continues to impact policing, it is critical for policymakers and stakeholders to understand the extent to which local law enforcement agencies are prepared for the changes. It is hoped that the BDPCI is a good incremental step toward that goal and that this Article drives further research into empirical measurement and big data policing.

[198] The authors and producers of 2016 LEMAS have, for instance, already produced a large, targeted dataset for police body cameras. *See generally* U.S. DEP'T OF JUSTICE, OFF. OF JUST. PROGRAMS, BUREAU OF JUST. STAT., LAW ENFORCEMENT MANAGEMENT AND ADMINISTRATIVE STATISTICS BODY-WORN CAMERA SUPPLEMENT (LEMAS-BWCS), 2016 (2016) [hereinafter LEMAS-BWCS].

[199] Coleman, *supra* note 110, at 314. As noted above, this is not a criticism of the utilized dataset. *See* Coleman, *supra* note 70, at 1389 n.128; Phillips et al., *supra* note 186, at 462.

[200] *See* Coleman, *supra* note 70, at 1391 ("Having perceptions of others—such as community members, prosecutors, defense attorneys, and jurors—might help reduce bias and present a broader perspective."); Pickering, *supra* note 192, at 400 ("[T]he results from this study represent the perceptions and opinions of participating police officers. While such information is undoubtedly informative and relevant, it is important that future research assess the accuracy of these perceptions by collecting additional data from community members, local prosecutors, and jurors.").

[201] The body-worn camera supplement authored and produced by the same entities as 2016 LEMAS contained some perception-based questions. *See generally* LEMAS-BWCS, *supra* note 198. These questions offered "four [answer] choices[,] basically analogous to a four-point Likert-type scale." *See* Coleman, *supra* note 70, at 1376, 1376 n.84 (noting choices were "'Strongly disagree', 'Disagree', 'Agree', and 'Strongly agree'"); Phillips et al., *supra* note 186, at 456. Although such perception-based questions sought law enforcement perceptions, similar questions could be drafted for use with other constituencies. *See, e.g.*, Coleman, *supra* note 70, at 1391.

[202] Coleman, *supra* note 115, at 315.

[203] *Id.*

[204] *Id.*

[205] *See generally* Alkire, Roche & Vaz, *supra* note 117; *see also* Coleman, *supra* note 110, at 315.

[206] 138 S.Ct. 2206, 2218–19 (2018).