

# Hope, Hype, and Fear: The Promise and Potential Pitfalls of Artificial Intelligence in Criminal Justice

William S. Isaac\*

## I. INTRODUCTION

Over the past decade, algorithmic decision systems (ADSs)—applications of statistical or computational techniques designed to assist human-decision making processes—have moved from an obscure domain of statistics and computer science into the mainstream.<sup>1</sup> The rapid decline in the cost of computer processing and ubiquity of digital data storage have created a dramatic rise in the adoption of ADSs using applied machine learning algorithms, transforming various sectors of society from digital advertising to political campaigns,<sup>2</sup> risk modeling for the banking sector,<sup>3</sup> health care and beyond.<sup>4</sup> In particular, many practitioners in the public sector have begun turning to ADSs as a means to stretch limited public resources amidst growing public demands for equity and accountability.<sup>5</sup> Advocates of these “intelligence-led” or “evidence-based” policy approaches assume big data tools will allow government agencies to use *objective* data to overcome historical inequalities to better serve underrepresented groups.<sup>6</sup>

---

\* William Isaac is a Doctoral Candidate in Political Science at Michigan State University and a 2017 Open Society Foundation Fellow. For helpful comments and insights on earlier drafts, I would like to thank Bobbi Isaac, Eric Juenke, Matt Grossmann, Cory Smidt, and Josh Sapotichne. Research for this article was supported in part by the Open Society Foundations. The opinions expressed herein are the author’s own and do not necessarily express the views of the Open Society Foundations.

<sup>1</sup> EXEC. OFFICE OF THE PRESIDENT, BIG DATA: A REPORT ON ALGORITHMIC SYSTEMS, OPPORTUNITY, AND CIVIL RIGHTS 5 (2016), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf) [<https://perma.cc/2FMG-Z744>]; CATHY O’NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016).

<sup>2</sup> EITAN D. HERSH, HACKING THE ELECTORATE: HOW CAMPAIGNS PERCEIVE VOTERS (2015); David W. Nickerson & Todd Rogers, *Political Campaigns and Big Data*, 28 J. ECON. PERSP. 51, 53 (2014).

<sup>3</sup> Yanhao Wei et al., *Credit Scoring with Social Network Data*, 35 MARKETING SCI. 234 (2016).

<sup>4</sup> Julia Powles & Hal Hodson, *Google DeepMind and Healthcare in an Age of Algorithms*, 7 HEALTH & TECH. 351 (2017).

<sup>5</sup> Bruno Lepri et al., *The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good*, in TRANSPARENT DATA MINING FOR BIG AND SMALL DATA 3 (Tania Cerquitelli et al. eds., 2017).

<sup>6</sup> EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 53 (2014), [https://obamawhitehouse.archives.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf) [<https://perma.cc/Q6L5-FYKC>]; Bhagwan Chowdhry et al., *How Big Data Can*

However, the assumption of objective data is flawed. All human behavior or social phenomenon that machine learning algorithms attempt to predict come from a data-generation process (DGP) which is comprised of trillions of complex interactions between the roughly seven billion people that inhabit our planet. The DGP is often unseen to the analyst, but we make assumptions regarding this process by the choice of statistical models or the inferences we derive from our analysis. Machine learning algorithms are often theorized and developed in cases of simulated data or data with outcome variables with little ambiguity in interpretation or method of collection. However, if we assume incorrectly about the DGP, the predictions and conclusions we generate will be highly inaccurate.<sup>7</sup> Furthermore, because the “true” DGP is unseen, it is nearly impossible to determine whether a proposed measure captures the phenomenon or outcome of interest for decision-makers.

Defining what is considered objective data is a particularly acute problem in criminal justice. Dating back to the turn of the 20th century, statisticians and criminologists have raised concerns over the operationalization and measurement of crime.<sup>8</sup> At its core, crime is a social phenomenon that has had multiple definitions and interpretations across time. Since the early 1930s, the United States Department of Justice uniform crime reporting data—considered the official assignment of national crime patterns—are based on crimes known to the police, from reporting by the public or witnessed by members of law enforcement.<sup>9</sup> While this operationalization of crime is reliable in the statistical sense (i.e. it will consistently measure the same concept over time), multiple scholars have pointed out this approach systematically undercounts crimes.<sup>10</sup> These “hidden” or reported instances are often referred to as the “dark figure of crime.”<sup>11</sup>

One of the most common reasons for the emergence of “dark figures” has been the policies and practices of individual police departments. One of the first studies to make a linkage between systemic bias in the criminal justice system and measurement found that arrest data for delinquency of juveniles in New York State (defined as truancy, theft, or malicious mischief) was not a function of a person’s race or socioeconomic status directly—as previously theorized—but rather the

---

*Make Us Less Racist: Computing Power Can Help Us Make Efficient Decisions About Who Is Friend or Foe*, ZÓCALO PUBLIC SQUARE (Apr. 28, 2016), <http://www.zocalopublicsquare.org/2016/04/28/how-big-data-can-make-us-less-racist/ideas/nexus/> [<https://perma.cc/RRX7-ZA8Z>].

<sup>7</sup> Lars-Erik Cederman & Nils B. Weidmann, *Predicting Armed Conflict: Time to Adjust Our Expectations?*, 355 SCI. 474 (2017).

<sup>8</sup> William Douglas Morrison, *The Interpretation of Criminal Statistics*, 60 J. ROYAL STAT. SOC’Y 1 (1897).

<sup>9</sup> FED. BUREAU OF INVESTIGATION, UNIFORM CRIME REPORTING, <https://ucr.fbi.gov/> [<https://perma.cc/VM2K-7GH5>] (last visited Apr. 3, 2018); CLAYTON J. MOSHER ET AL., *THE MISMEASURE OF CRIME* 43 (2d ed. 2011).

<sup>10</sup> MOSHER ET AL., *supra* note 9, at 101.

<sup>11</sup> *Id.* at 31.

differential treatment of these individuals by the criminal justice system.<sup>12</sup> Ronald Beattie in the 1940's noted that in addition to demographic factors, police statistics were likely manipulated based on the local political conditions, often with a tendency to "report those facts which show a good administrative record on the part of the department."<sup>13</sup>

More recently, Steven Levitt analyzed crime victimization and reporting data and found that the likelihood of a crime being reported to the police increases as the size of the city's police force increases.<sup>14</sup> Ziggy MacDonald assesses the likelihood of reporting crime to law enforcement in the United Kingdom, and finds that non-white (except Asian residents), unemployed, and low-income residents were less likely to report crimes.<sup>15</sup> A longitudinal study by Eric Baumer and Janet Lauritsen of crime reporting from the US National Crime Victimization Study (NCVS) between 1973 and 2005 found similar findings.<sup>16</sup> While their study found increasing rates of reporting over time, non-white victims and male victims were much less likely to report crimes to the police.<sup>17</sup> More surprising was that overall, "just 40 percent of the nonlethal violent incidents and 32 percent of the property crimes recorded in the national crime surveys during this period were reported to the police."<sup>18</sup> These findings would suggest that not only is the "dark figure" of crime very large, it is often unrepresentative of the overall picture of crime in a given area.

Simply put, crimes recorded by the police departments are not a complete census of all criminal offenses, nor do they constitute a representative random sample. Police records are actually a complex interaction between criminality, policing strategy, and community-police relations. And, while the debate about how to measure and operationalize crime was previously confined to esoteric debates among academic criminologists, it has become more relevant as the use of criminal justice data has moved into the big data era. The impact of poor input data on analysis and prediction is not a new concern—anyone who has taken a course on data analysis has heard the saying "garbage in, garbage out"—but in an era of an ever-expanding array of statistical models presented as panaceas to large and complex real-world data, we often forget to think carefully about the quality,

---

<sup>12</sup> SOPHIA MOSES ROBISON, CAN DELINQUENCY BE MEASURED? 64 (1936).

<sup>13</sup> Ronald H. Beattie, *The Sources of Criminal Statistics*, 217 ANNALS AM. ACAD. POL. & SOC. SCI. 19, 21–22 (1941).

<sup>14</sup> Steven D. Levitt, *The Relationship Between Crime Reporting and Police: Implications for the Use of Uniform Crime Reports*, 14 J. QUANTITATIVE CRIMINOLOGY 61, 64 (1998).

<sup>15</sup> Ziggy MacDonald, *Official Crime Statistics: Their Use and Interpretation*, 112 ECON. J. 85, 99–100 (2002).

<sup>16</sup> Eric P. Baumer & Janet L. Lauritsen, *Reporting Crime to the Police, 1973–2005: A Multivariate Analysis of Long-Term Trends in the National Crime Survey (NCS) and National Crime Victimization Survey (NCVS)*, 48 CRIMINOLOGY 131, 135 (2010).

<sup>17</sup> *Id.*

<sup>18</sup> *Id.* at 153.

instead of simply the quantity, of our data sources. As David Lazer and his colleagues point out, high quantity of data does not provide latitude to ignore foundational issues of measurement, construct validity, and dependencies among elements of our data.<sup>19</sup>

This is particularly true for machine learning in criminal justice, as these models are heavily reliant on the training dataset to estimate predictions. As I discuss below, machine learning algorithms are unaware, and in many cases, unable to adjust for institutional biases embedded within policing data. As a result, the presence of bias in the initial (training) dataset leads to predictions that are subject to the same biases that already exist within the dataset. Further, these biased forecasts can often become amplified if practitioners begin to concentrate resources on an increasingly smaller subset of these forecasted targets. Thus, a failure to understand the limitations of data used in these predictive tools—and create more transparent and accountable mechanisms to mitigate these potential harms—may simply perpetuate historical discrimination toward underrepresented groups and violate their civil and human rights.

## II. CASE STUDY: PREDICTIVE POLICING

One of the most popular and fastest growing ADSs in criminal justice are “predictive policing” tools, which are generally defined as applications designed to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions.<sup>20</sup> Much like Amazon or Facebook’s use of consumer data to serve up relevant ads or products to consumers, police departments across the United States and Europe increasingly utilize software from Silicon Valley based companies, such as Predpol and Palantir, to identify future offenders, highlight trends in criminal activity, and even forecast the locations of future crimes.<sup>21</sup> Police departments around the country are increasingly relying on a growing suite of predictive policing ADSs to more efficiently allocate policing resources amid shrinking public budgets and increasing pressure to be more responsive to the communities they serve.<sup>22</sup> A recent survey of police agencies found that 70% planned to implement or increase use of predictive policing technology in the next two to five years.<sup>23</sup> Outside the United States, European

---

<sup>19</sup> David Lazer et al., *The Parable of Google Flu: Traps in Big Data Analysis*, 343 SCI. 1203, 1203 (2014).

<sup>20</sup> WALTER L. PERRY ET AL., RAND CORP., PREDICTIVE POLICING: THE ROLE OF CRIME FORECASTING IN LAW ENFORCEMENT OPERATIONS xiii (2013).

<sup>21</sup> Kristian Lum & William Isaac, *To Predict and Serve?*, SIGNIFICANCE MAG., Oct. 2016, at 15.

<sup>22</sup> PERRY ET AL., *supra* note 20, at 127–28.

<sup>23</sup> CMTY. ORIENTED POLICING SERVS., U.S. DEP’T OF JUSTICE, ET AL., FUTURE TRENDS IN POLICING, 3 (2014), [http://www.policeforum.org/assets/docs/Free\\_Online\\_Documents/Leadership/future%20trends%20in%20policing%202014.pdf](http://www.policeforum.org/assets/docs/Free_Online_Documents/Leadership/future%20trends%20in%20policing%202014.pdf) [https://perma.cc/3VPG-STTA].

cities such as Kent, London, and Berlin are considering the use of predictive policing (or pre-crime) tools to predict potential violent gang members.<sup>24</sup>

While proponents of predictive policing have viewed this trend as a significant step towards transparency and pragmatic, data-driven policymaking, the use of predictive ADSs within police departments has also raised very serious concerns among activists and scholars regarding this new intersection between statistical learning and public policy.<sup>25</sup> Civil liberties advocates have argued the growth of predictive policing means that officers in the field are more likely to stop suspects who have yet to commit a crime under the guise of historical crime patterns that are not representative of all criminal behavior.<sup>26</sup> As Ezekiel Edwards of the ACLU noted, “It is well known that crime data is notoriously suspect, incomplete, easily manipulated and plagued by racial bias.”<sup>27</sup> In their excellent report on predictive policing, David Robinson and Logan Koepke point out that reported crime data are “greatly influenced by what crimes citizens choose to report, the places police are sent on patrol, and how police decide to respond to the situations they encounter.”<sup>28</sup> Legal scholars such as Joh note that, “Police are not simply end users of big data. They *generate the information* that big data programs rely upon. Crime and disorder are not natural phenomena. These events have to be observed, noticed, acted upon, collected, categorized, and recorded—while other events aren’t.”<sup>29</sup>

Is there any evidence of these potential harms? To date, there is only one empirical study to examine the impact of police-recorded training data on predictive policing models. In their analysis, Kristian Lum and William Isaac

<sup>24</sup> Chris Baraniuk, *Pre-Crime Software Recruited to Track Gang of Thieves*, NEW SCIENTIST (Mar. 11, 2015), <https://www.newscientist.com/article/mg22530123-600-pre-crime-software-recruited-to-track-gang-of-thieves/> [<https://perma.cc/5462-ZKGU>]; Sarah Griffiths, *Predicting Crimes BEFORE They Happen: Berlin Police Adopt Minority Report-Style Software That Seeks Out Criminal Behaviour*, DAILY MAIL (Dec. 2, 2014, 3:40 PM), <http://www.dailymail.co.uk/sciencetech/article-2857814/Predict-crimes-happen-Berlin-police-adopt-Minority-Report-style-software-seeks-criminal-behaviour.html> [<https://perma.cc/Z3DT-GRYY>].

<sup>25</sup> *Statement of Concern About Predictive Policing by ACLU and 16 Civil Rights Privacy, Racial Justice, and Technology Organizations*, ACLU (Aug. 31, 2016), <https://www.aclu.org/other/statement-concern-about-predictive-policing-aclu-and-16-civil-rights-privacy-racial-justice> [<https://perma.cc/D269-XQ77>].

<sup>26</sup> William Isaac & Kristian Lum, Opinion, *Predictive Policing Violates More Than It Protects*, USA TODAY (Dec. 2, 2016, 5:18 PM), <https://www.usatoday.com/story/opinion/policing/spotlight/2016/12/02/predictive-policing-violates-more-than-protects-column/94569912/> [<https://perma.cc/9873-2VTC>].

<sup>27</sup> Jamiles Lartey, *Predictive Policing Practices Labeled as ‘Flawed’ by Civil Rights Coalition*, GUARDIAN (Aug. 31, 2016, 2:55 PM), <https://www.theguardian.com/us-news/2016/aug/31/predictive-policing-civil-rights-coalition-aclu> [<https://perma.cc/J88V-L8ZX>].

<sup>28</sup> David Robinson & Logan Koepke, *Stuck in a Pattern: Early Evidence on “Predictive Policing” and Civil Rights*, UPTURN (Aug. 2016), <https://www.teamupturn.com/reports/2016/stuck-in-a-pattern> [<https://perma.cc/8TW8-BRUT>].

<sup>29</sup> Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data & Algorithms*, 26 WM. & MARY BILL RTS J. 287, 289 (2017).

replicate the Epidemic-Type Aftershock Sequence or ETAS crime forecasting model outlined in a 2015 study by George Mohler and colleagues used commercially by the predictive policing vendor Predpol to generate predictions on model to publicly available data on drug crimes in the City of Oakland from 2009 to 2011.<sup>30</sup> In the study, the authors argue that the unrepresentative nature of recorded crime data could bias predictive policing forecasts in two ways.<sup>31</sup>

First, the presence of bias in the initial training data leads to predictions that are subject to the same biases that already exist within the records.<sup>32</sup> Because these predictions are likely to over-represent areas that were already known to police, police become increasingly likely to patrol these same areas and observe new criminal acts that confirm their prior beliefs regarding the distributions of criminal activity.<sup>33</sup> Figure 1 below,<sup>34</sup> reprinted from the original study, plots the estimated number of drug users in Oakland, California divided by 150 x 150 meter bins, with the greater intensity indicating a higher number of estimated drug users in a particular location. As we see from the figure, the spatial distribution of drug users appears spread across the city, with elevated levels along International Boulevard, which is a primary thoroughfare in the City of Oakland. This is compared to figure 2,<sup>35</sup> reprinted from the same study, which plotted the number of reported crimes across the city divided by bins and demonstrates a much more concentrated view as the recorded crimes seem to be exclusively isolated to neighborhoods around West Oakland and Fruitvale, two neighborhoods with largely non-white and low-income populations. Variations in the latter are driven primarily by differences in population density, as the estimated rate of drug use is relatively uniform across the city.<sup>36</sup>

---

<sup>30</sup> Lum & Isaac, *supra* note 21, at 17; G.O. Mohler et al., *Self-Exciting Point Process Modeling of Crime*, 106 J. AM. STAT. ASS'N 100, 100 (2011).

<sup>31</sup> Lum & Isaac, *supra* note 21, at 16–19.

<sup>32</sup> *Id.* at 16.

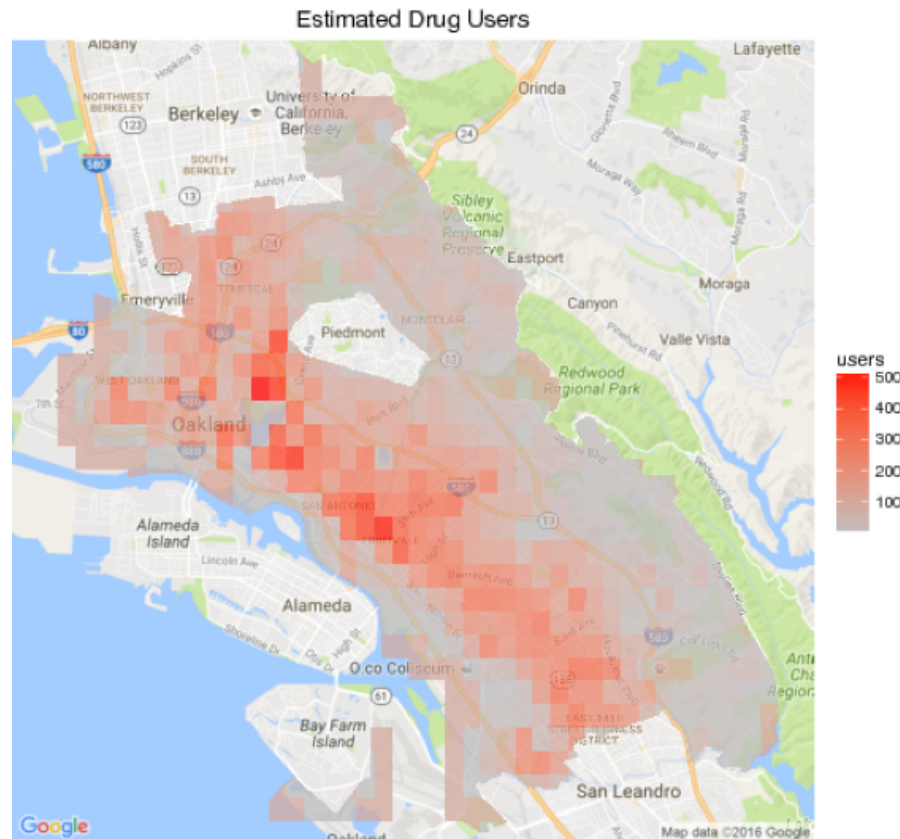
<sup>33</sup> *Id.*

<sup>34</sup> *Id.* at 17.

<sup>35</sup> *Id.* at 18.

<sup>36</sup> *Id.*

Figure 1



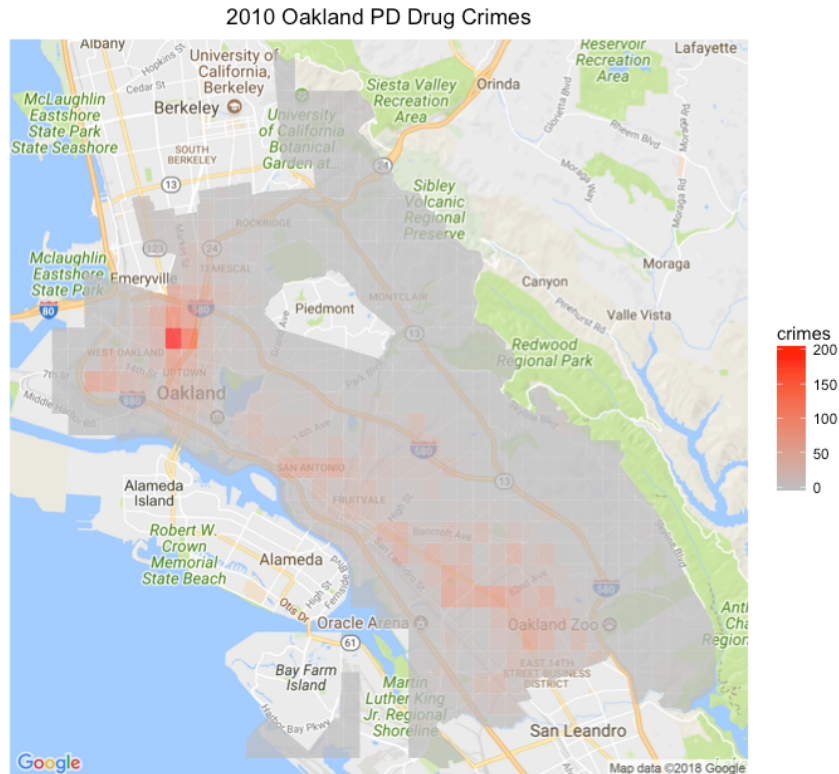
From these figures it is clear that police databases and public health-derived estimates tell dramatically different stories about the pattern of drug use in Oakland. The important question is whether the predictions generated from the ETAS model are more closely aligned with the crime data or estimates from public health data. Figure 3 plots the number of days each bin would have been flagged by Predpol for targeted policing, with greater color intensity indicating higher number of days targeted.<sup>37</sup> Given the few number of bins targeted by the ETAS algorithm, it is clear the model failed to capture the larger spatial variation found in the public health data and more closely resembles the reported crime data. Further, this narrower targeting led to a higher concentration of targeted policing among minority and low-income neighborhoods.<sup>38</sup> As the authors note in applying the ETAS algorithm to crime data in Oakland, “Black people would be targeted by predictive policing at roughly twice the rate of Whites. Individuals classified as a

<sup>37</sup> *Id.* at 18–19.

<sup>38</sup> *Id.*

race other than white or black would receive targeted policing at a rate 1.5 times that of Whites.”<sup>39</sup>

Figure 2



Second, the newly observed criminal acts that police document as a result of these targeted patrols then feed into the predictive policing algorithm on subsequent days, generating increasingly biased predictions.<sup>40</sup> This feedback loop or “ratchet effect” can lead to model over-fitting in that the locations most likely to experience further criminal activity are exactly the locations they had previously believed to be high in crime.<sup>41</sup> In their study, Lum and Isaac attempt to address this issue by simulating a scenario of the application of the ETAS model where in addition to the observed Oakland crime data there is an additional 20% chance that additional crimes are found in the targeted bin for a given day.<sup>42</sup> In this scenario, the study finds a dramatic increase in the predicted odds of targeting previous bins

<sup>39</sup> *Id.* at 18.

<sup>40</sup> *Id.* at 19.

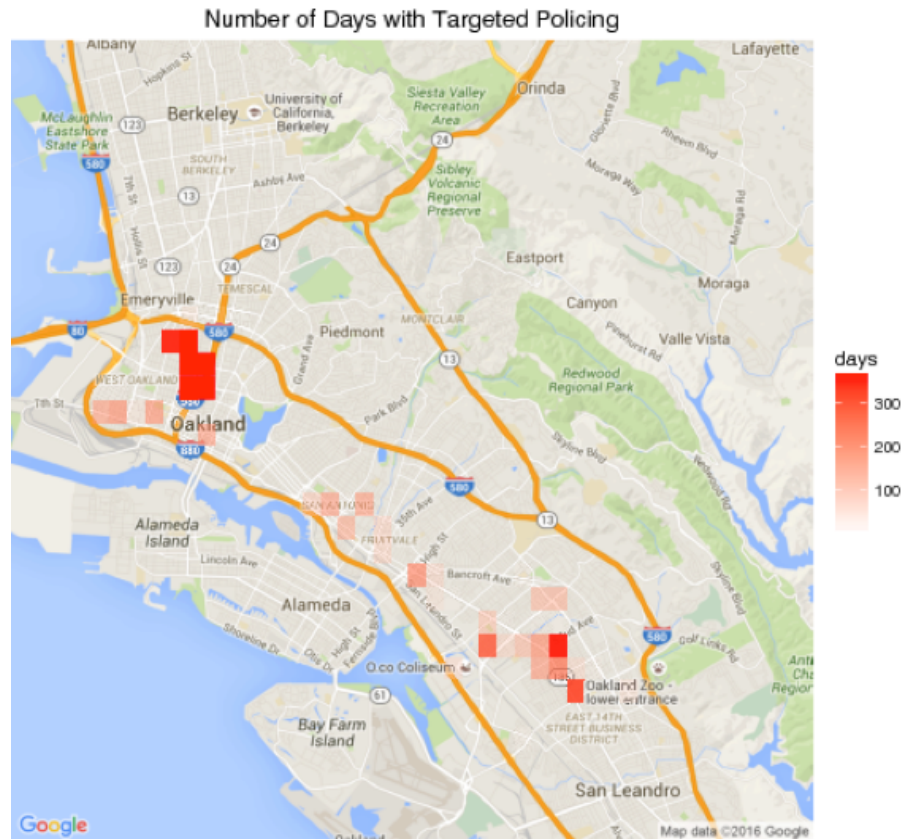
<sup>41</sup> Robinson & Koepke, *supra* note 28.

<sup>42</sup> Lum & Isaac, *supra* note 21, at 18.



versus to non-targeted bins compared to the baseline example.<sup>43</sup> This evidence led the authors to conclude the feedback scenario “causes the Predpol algorithm to become increasingly confident that most of the crime is contained in the targeted bins.”<sup>44</sup> In short, the feedback mechanism is selection bias meets confirmation bias.

Figure 3



As the findings from Lum and Isaac make clear, crime data is not inherently objective and is ultimately reflective of institutional behavior and norms, and any predictive policing algorithm—not just Predpol’s ETAS algorithm—that fails to consider this limitation of recorded crime data will simply generate predictions that reproduce the embedded biases.<sup>45</sup> That said, a few intentional decisions made by Predpol’s ETAS algorithm could lead to biased targeting that is unique to other

<sup>43</sup> *Id.*

<sup>44</sup> *Id.* at 19.

<sup>45</sup> *Id.*

methods. Specifically, the algorithm's use of historical baselines means that neighborhoods which have been chronically over-policed will have higher baseline values in the initial estimates. Biased targeting is further compounded when you consider that Predpol ranks the bins based on their conditional intensity scores and reports only small subset for targeted policing. This means the top ranked bins are unlikely to change unless there is a dramatic surge in reported crime for an extended period of time. The "stickiness" of the baseline values combined with the ratcheting effect from newly reported crimes would suggest that neighborhoods that have received disproportionate targeting from the police historically will continue to receive targeted policing under the ETAS model. Further, the sticky baselines could also prevent the police from being alerted to small but noticeable upticks in crime in other neighborhoods, potentially making them less responsive to changing conditions in the community.

Some critics of Lum and Isaac have questioned the appropriateness of using drug crime data for generating forecasts using Predpol's ETAS algorithm, as Predpol has claimed their model has not been used for forecasting drug crimes,<sup>46</sup> but this claim fails on many fronts. First, as the authors stated clearly in the study, drug crimes were used because it allowed for the use of public health data on illicit drug use to serve as a comparison against observations recorded by police departments, not because Predpol or other major predictive policing vendors were known to be widely forecasting drug crimes.<sup>47</sup> Second, most major police departments use some form of "Hot Spot" policing tactics<sup>48</sup>—which serve as the theoretical foundation for predictive policing—to isolate areas where they believe drug activity to be occurring. Lastly, despite predictive policing vendors' claims, governments are interested in using predictive ADSs to target drug crimes. For example, the National Institute of Justice included drug crimes as part of their recent \$1.2 million crime forecasting challenge.<sup>49</sup> And, local police departments have explicitly proposed using predictive policing for predicting the location of drug crimes and gang activity.<sup>50</sup> With the increased concern within the Trump

---

<sup>46</sup> Jack Smith IV, (*Exclusive*) *Crime-Prediction Tool PredPol Amplifies Racially Biased Policing, Study Shows*, MIC (Oct. 9, 2016), <https://mic.com/articles/156286/crime-prediction-tool-pred-pol-only-amplifies-racially-biased-policing-study-shows> [<https://perma.cc/GW3L-P4GX>].

<sup>47</sup> Lum & Isaac, *supra* note 21, at 16.

<sup>48</sup> Anthony A. Braga, *Hot Spots Policing and Crime Prevention: A Systematic Review of Randomized Controlled Trials*, 1 J. EXPERIMENTAL CRIMINOLOGY 317, 317 (2005); Tanvi Misra, *What This Salt Lake City Heatmap Tells Us About Drug Crime*, CITYLAB (Aug. 11, 2017), <https://www.citylab.com/equity/2017/08/what-this-map-of-salt-lake-citys-drug-hotspot-really-show/536214/> [<https://perma.cc/G4VD-JUBH>].

<sup>49</sup> *Real-Time Crime Forecasting Challenge Posting*, NAT'L INST. OF JUSTICE (July 31, 2017), <https://nij.gov/funding/Pages/fy16-crime-forecasting-challenge-document.aspx> [<https://perma.cc/8N52-W5VH>].

<sup>50</sup> See Kate Ramunni, *Hamden Police Using App, Eyeing Software to Help Fight Crime*, NEW HAVEN REG. (Nov. 15, 2015, 8:18 PM), <https://www.nhregister.com/connecticut/article/Hamden-police-using-app-eyeing-software-to-help-11345181.php> [<https://perma.cc/W6TF-GBT5>]; Press Release, Hamden Police Department, Edward Byrne Memorial Justice Assistance Grant (JAG)

administration about the growing opioid crisis,<sup>51</sup> it is certainly reasonable to assume that more departments will look to predictive policing to help combat this issue.

### III. DISCUSSION

Using predictive analytics in the real world is challenging, especially in high-stakes policy areas such as policing. However, this does not mean police departments should abandon the use of analytics or intelligence-led approaches to improve public safety. Rather, it is important for police departments and other law enforcement agencies to think more broadly about the potential impacts of implementing algorithmic decision-support tools and ensure they create internal and external systems to promote public safety while minimizing disparate impacts. Specifically, police departments or any other agencies that attempt to implement algorithmic decision-support tools should take steps to develop internal and external accountability, ensure operational transparency, and be aware of the long-run impact to the community.

The process to accountable and transparent use of algorithmic decision support systems must start with community stakeholders and police departments discussing policing priorities and measures of police performance. Currently, nearly all predictive policing systems are tasked with identifying neighborhoods or individuals determined to be high public safety risks, which departments then use to intervene with additional surveillance or negative enforcement actions (i.e. arrest or citation).<sup>52</sup> However, these negative enforcement actions have been found to be “contagious” in communities targeted by police,<sup>53</sup> leading to a deterioration of police-community relations and perpetuating the mass incarceration crisis in the United States.<sup>54</sup> Moreover, police departments often use metrics such as arrests or citations for internal promotion,<sup>55</sup> creating a perverse incentive for police officers to increase negative enforcement actions for institutional support. Predictive policing and algorithmic decision-support systems not properly implemented will

---

Program: FY 2015 Local Solicitation (July 28, 2014), (<http://www.hamden.com/content/219/228/9315/default.aspx> [<https://perma.cc/8UN7-GPZR>]).

<sup>51</sup> Brian Naylor & Tamara Keith, *Trump Says He Intends to Declare Opioid Crisis National Emergency*, NPR (Aug. 10, 2017, 4:41 PM), <http://www.npr.org/2017/08/10/542669730/trump-says-he-intends-to-declare-opioid-crisis-national-emergency> [<https://perma.cc/2MVX-77CN>].

<sup>52</sup> Robinson & Koepke, *supra* note 28.

<sup>53</sup> Kristian Lum et al., *The Contagious Nature of Imprisonment: An Agent-Based Model to Explain Racial Disparities in Incarceration Rates*, 11 J. ROYAL SOC’Y INTERFACE 1, 1 (2014).

<sup>54</sup> Anthony W. Flores et al., *False Positives, False Negatives, and False Analyses: A Rejoinder to “Machine Bias: There’s Software Used Across the Country to Predict Future Criminals. And It’s Biased Against Blacks.”*, 80 FED. PROB. 1, 3 (2016).

<sup>55</sup> Saki Knafo, *A Black Police Officer’s Fight Against the N.Y.P.D.*, N.Y. TIMES MAG. (Feb. 18, 2016), <https://www.nytimes.com/2016/02/21/magazine/a-black-police-officers-fight-against-the-nypd.html> [<https://perma.cc/V4GX-3VT9>].

only exacerbate the problem. As Jessica Saunders, Priscilla Hunt, and John Hollywood discovered in their assessment of the implementation of the Strategic Subjects List (SSL) for the Chicago Police Department (CPD), the individual-level ADS tool was rarely used to prevent victimization from violent crimes, rather the authors found that CPD officers often used the SSL as a way to generate leads for unsolved shooting cases.<sup>56</sup>

Yet, the promise of leveraging data to improve public safety also provides an opportunity to reform how we think about policing, and a growing number of innovative departments have started to pursue these alternative approaches. For example, Toronto and other cities in Canada have moved toward the “HUB and COR” model of predictive policing which allow police departments to serve as a conduit to access other social services rather than demanding departments to respond to a myriad of social problems with a narrow range of enforcement tools, and reflects recent experimental studies that suggest targeted use of social services can be an effective strategy for reducing crime, particularly on violent crime among juveniles.<sup>57</sup>

Under the “HUB and COR” model, a predictive model uses data collected from multiple governmental agencies to flag at-risk individuals (often minors) in need of urgent intervention by law enforcement.<sup>58</sup> The individuals identified by the risk models are then vetted by representatives from various agencies and community groups, and the persons identified as high-risk will have “rapid response” plans tailored to each individual or family’s needs prior to more serious problems occurring.<sup>59</sup> And, while civil society groups and journalists have rightly raised concerns about potential abuse of the system due to weak privacy protections such as personally identifiable information of the targeted persons,<sup>60</sup> new advances in blockchain technology—the algorithm that serves as the foundation to the cryptocurrency bitcoin<sup>61</sup>—could potentially allow HUBs to

---

<sup>56</sup> Jessica Saunders et al., *Predictions Put into Practice: A Quasi-Experimental Evaluation of Chicago’s Predictive Policing Pilot*, 12 J. EXPERIMENTAL CRIMINOLOGY 347, 366 (2016).

<sup>57</sup> Brad J. Bushman et al., *Youth Violence: What We Know and What We Need to Know*, 71 AM. PSYCHOL. 17, 29 (2016); Sara B. Heller, *Summer Jobs Reduce Violence Among Disadvantaged Youth*, 346 SCI. MAG. 1219, 1219 (2014); Andrew Barr & Chloe R. Gibbs, *Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood* (UC Davis Ctr. for Poverty Research, Working Paper, 2017), [http://www.chloegibbs.com/s/Barr-Gibbs\\_Intergenerational\\_2017.pdf](http://www.chloegibbs.com/s/Barr-Gibbs_Intergenerational_2017.pdf) [<https://perma.cc/3GVD-ZJDB>].

<sup>58</sup> Dale R. McFee & Norman E. Taylor, *The Prince Albert Hub and the Emergence of Collaborative Risk-Driven Community Safety*, CANADIAN POLICE COLLEGE DISCUSSION PAPER SERIES (2014), <http://www.cpc-ccp.gc.ca/sites/default/files/pdf/prince-albert-hub-eng.pdf> [<https://perma.cc/6EML-C7JH>]; Nathan Munn, *Canada’s ‘Pre-Crime’ Model of Policing is Sparking Privacy Concerns*, MOTHERBOARD (Jan. 19, 2017, 9:00 AM), [http://motherboard.vice.com/en\\_us/article/mg7w4x/canada-hub-and-cor-policing-privacy-police](http://motherboard.vice.com/en_us/article/mg7w4x/canada-hub-and-cor-policing-privacy-police) [<https://perma.cc/U8CF-S7K5>].

<sup>59</sup> See sources cited *supra* note 58.

<sup>60</sup> Munn, *supra* note 58.

<sup>61</sup> Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*, BITCOIN (Oct. 31, 2008), <https://bitcoin.org/bitcoin.pdf> [<https://perma.cc/8UQR-JBDW>].

generate predictive risk scores and identify at-risk persons in a manner which respects the privacy of persons in the datasets.<sup>62</sup>

In addition to rethinking what officers do when deployed into communities, the big data era of policing also forces us to question how we measure successful police strategies. The New York Police Department (NYPD) is taking an innovative approach to address the issue by moving away from measuring departmental performance based on enforcement actions—such as their pioneering CompStat model—and will instead measure performance based on large-scale community surveys sent out daily by the department.<sup>63</sup> The surveys, which ask residents about their level of trust in the department and safety in their neighborhood, will be used to generate a precinct level “sentiment meter” score between 100 and 900.<sup>64</sup> It is unclear if the NYPD has incorporated their sentiment scores into predictive policing models, but this approach does create the potential to allow police departments to tailor strategies around a measure that is independent of policing reporting practices and more closely aligns officer incentives to community needs. The potential use of alternative measures could address some of the concerns about reporting biases, but the scale of such a data collection effort (50,000 respondents per day), raises concerns from civil liberties groups on the potential privacy issues, in addition to challenges that most survey researchers face to generate representative samples of residents will severely limit how many cities could adopt similar measures.<sup>65</sup>

Another important question that must be addressed is whether ADS tools and their subsequent implementation actually lead to greater public safety and benefit to underserved communities. To date, only three empirical studies of predictive policing has been published. Saunders, Hunt, and Hollywood assess the Chicago Police Department’s SSL, a person-based predictive policing system created internally in 2013.<sup>66</sup> The authors use ARIMA time series models to estimate impacts of the deployment of the SSL in 2013 on city-level homicide trends.<sup>67</sup> While the authors find a decline in city-level homicides overall, the introduction of the SSL failed to have a measurable impact.<sup>68</sup> As the authors note, “the statistically significant reduction in monthly homicides predated the introduction of the SSL, and that the SSL did not cause further reduction in the average number

---

<sup>62</sup> Guy Zyskind et al., *Decentralizing Privacy: Using Blockchain to Protect Personal Data*, 2015 IEEE CS SECURITY & PRIVACY WORKSHOPS 180, 182–83.

<sup>63</sup> Al Baker, *Updated N.Y.P.D Anti-Crime System to Ask: ‘How We Doing?’*, N.Y. TIMES (May 8, 2017), <https://www.nytimes.com/2017/05/08/nyregion/nypd-compstat-crime-mapping.html> [<https://perma.cc/TC25-C9FN>].

<sup>64</sup> *Id.*

<sup>65</sup> *Id.*

<sup>66</sup> Saunders et al., *supra* note 56, at 347.

<sup>67</sup> *Id.*

<sup>68</sup> *Id.*

of monthly homicides above and beyond the pre-existing trend.”<sup>69</sup> Hunt, Saunders, and Hollywood conducted a randomized control trial on the deployment of a predictive policing system in Shreveport, Louisiana, and found no statistically significant change in property crime in the experimental districts that applied the predictive models compared with the control districts.<sup>70</sup>

The only study to find a statistically significant decline in reported crime is Mohler et al., which conducted a randomized control trial of Predpol’s ETAS model with the Los Angeles (United States) and Kent Police Department (United Kingdom).<sup>71</sup> The authors used a novel randomization approach by randomizing between crime maps created by the ETAS algorithm and one generated by human crime analysts.<sup>72</sup> Overall, the police patrols using ETAS forecasts led to an average 7.4% reduction in crime volume, while patrols based upon analyst predictions showed no significant effect on crime volume.<sup>73</sup> While this reduction in crime volume is notable, Thomas suggests that the reduction in crime may have been spurious, as LAPD’s crime statistics show other divisions that were not using Predpol also saw crime reduction as high as 16% during the same period.<sup>74</sup> Given these inconclusive peer-reviewed findings, some vendors point to internal testing done by departments themselves as evidence of the efficacy of predictive policing.<sup>75</sup> However, Robinson and Koepke note that although “system vendors often cite internally performed validation studies to demonstrate the value of their solutions, our research surfaced few rigorous analyses of predictive policing systems’ claims of efficacy, accuracy, or crime reduction.”<sup>76</sup>

The concerns about implementation and efficacy raise critical questions about the appropriate degree of transparency and regulation that algorithmic decision support tools should have. Much of the concerns originally expressed by civil society groups and other skeptics of ADS tools were centered around opacity of the underlying algorithms which generate the predictions provided to law enforcement agencies.<sup>77</sup> In response, many agencies have sought to move away

---

<sup>69</sup> *Id.* at 361.

<sup>70</sup> PRISCILLIA HUNT ET AL., RAND CORP., EVALUATION OF THE SHREVEPORT PREDICTIVE POLICING EXPERIMENT (2014).

<sup>71</sup> G.O. Mohler et al., *Randomized Controlled Field Trials of Predictive Policing*, 110 J. AM. STAT. ASS’N 1399, 1399 (2015).

<sup>72</sup> *Id.*

<sup>73</sup> *Id.*

<sup>74</sup> Emily Thomas, *Why Oakland Police Turned Down Predictive Policing*, MOTHERBOARD (Dec. 28, 2016, 9:00 AM), [https://motherboard.vice.com/en\\_us/article/ezp8zp/minority-retort-why-oakland-police-turned-down-predictive-policing](https://motherboard.vice.com/en_us/article/ezp8zp/minority-retort-why-oakland-police-turned-down-predictive-policing) [<https://perma.cc/94VC-ZG2N>].

<sup>75</sup> Robinson & Koepke, *supra* note 28.

<sup>76</sup> *Id.*

<sup>77</sup> Laurel Eckhouse, Opinion, *Big Data May Be Reinforcing Racial Bias in the Criminal Justice System*, WASH. POST (Feb. 10, 2017), [https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d\\_story.html?utm\\_term=.57752c78d70e](https://www.washingtonpost.com/opinions/big-data-may-be-reinforcing-racial-bias-in-the-criminal-justice-system/2017/02/10/d63de518-ee3a-11e6-9973-c5efb7ccfb0d_story.html?utm_term=.57752c78d70e) [<https://perma.cc/JLB8-UC4K>]; Andrew Selbst

from third-party commercial vendors and opt for tools built in-house or in collaboration with universities.<sup>78</sup> Newer predictive policing companies such as Civicscape have committed to algorithmic transparency by publishing a version of their source code on the online code repository Github and pledged to not use their tools to predict drug crimes because of concerns that the bias present in crime data are too difficult to model out of their predictions.<sup>79</sup>

These efforts are certainly a laudable move toward transparency and accountability, although there are still important issues that need to be resolved. For example, a question that arises from the Civicscape transparency efforts is whether vendors should be responsible for defining what constitutes transparency, fairness, and oversight before policymakers set firm guidelines. Under ideal circumstances, vendors or police departments would disclose their code for public scrutiny after each major release, but there is little incentive to continue their attempts at transparency as future iterations of the software are released, perhaps allowing biases to creep back in as more features or different data are included.

The long-term success of ADS tools in criminal justice will depend on regulatory guidelines that can hold officials accountable and remove complex ethical decisions out of the hands of software developers, whose interests may not always be in alignment with the needs of the communities affected. What would a regulatory system for ADS tools look like? Shneiderman has outlined a three prong approach that could serve as a potential blueprint.<sup>80</sup> In the article, the author outlines three kinds of AI oversight mechanisms, 1) a review board model where vendors or agencies should submit their tool or algorithm before any real world implementation, 2) continuous monitoring or auditing oversight reminiscent of what companies and non-profit foundations are required to do for financial due diligence, and 3) retrospective analysis of “disaster” scenarios much like the NTSB does after a plane crash by reviewing the black box data and internal governance.<sup>81</sup> One shortcoming of this approach is that it would be very resource intensive to carry out, and most police departments or city governments do not have the capacity to rigorously assess these algorithms in the manner outlined in the paper. However, an alternative approach would be for civil society groups to provide a set of best practices at the procurement stage (or proposal stage for in-house tools) and have a volunteer panel of civil society groups and university experts to prepare

---

& Solon Barocas, *Regulating Inscrutable Systems* (Mar. 20, 2017) (unpublished article) (<http://www.werobot2017.com/wp-content/uploads/2017/03/Selbst-and-Barocas-Regulating-Inscrutable-Systems-1.pdf> [<https://perma.cc/25GK-25Y2>]).

<sup>78</sup> Mara Hvistendahl, *Crime Forecasters*, 353 SCI. MAG. 1484, 1485 (2016); Aaron Shapiro, Comment, *Reform Predictive Policing*, 541 NATURE 458, 459 (2017).

<sup>79</sup> Dave Gershgorn, *Software Used to Predict Crime Can Now Be Scoured for Bias*, QUARTZ (Mar. 22, 2017), <https://qz.com/938635/a-predictive-policing-startup-released-all-its-code-so-it-can-be-scoured-for-bias/> [<https://perma.cc/5GQL-S7BE>].

<sup>80</sup> Ben Shneiderman, Opinion, *The Dangers of Faulty, Biased, or Malicious Algorithms Requires Independent Oversight*, 113 PROC. NAT'L ACAD. SCI. 13538, 13539 (2016).

<sup>81</sup> *Id.*

oversight reports. If we are successful in developing better guidelines for ADS tools, cities will both improve the quality of the data they collect and implement more transparent and inclusive processes to build safer communities for all of their residents.