

IN DETAIL

To predict and serve?

Predictive policing systems are used increasingly by law enforcement to try to prevent crime before it occurs. But what happens when these systems are trained using biased data?

Kristian Lum and **William Isaac** consider the evidence – and the social consequences





Kristian Lum, PhD is the lead statistician at the Human Rights Data Analysis Group



William Isaac, MPP is a doctoral candidate in the Department of Political Science at Michigan State University

In late 2013, Robert McDaniel – a 22-year-old black man who lives on the South Side of Chicago – received an unannounced visit by a Chicago Police Department commander to warn him not to commit any further crimes. The visit took McDaniel by surprise. He had not committed a crime, did not have a violent criminal record, and had had no recent contact with law enforcement. So why did the police come knocking?

It turns out that McDaniel was one of approximately 400 people to have been placed on Chicago Police Department's "heat list". These individuals had all been forecast to be potentially involved in violent crime, based on an analysis of geographic location and arrest data. The heat list is one of a growing suite of predictive "Big Data" systems used in police departments across the USA and in Europe to attempt what was previously thought impossible: to stop crime before it occurs.¹

This seems like the sort of thing citizens would want their police to be doing. But predictive policing software – and the policing tactics based on it – has raised serious concerns among community activists, legal scholars, and sceptical police chiefs. These concerns include: the apparent conflict with protections against unlawful search and seizure and the concept of reasonable suspicion; the lack of transparency from both police departments and private firms regarding how predictive policing models are built; how departments utilise their data; and whether the programs unnecessarily target specific groups more than others.

But there is also the concern that police-recorded data sets are rife with systematic bias. Predictive policing software is designed to learn and reproduce patterns in data, but if biased data is used to train these predictive models, the models will reproduce and in some cases amplify those same biases. At best, this renders the predictive models ineffective. At worst, it results in discriminatory policing.

Bias in police-recorded data

Decades of criminological research, dating to at least the nineteenth century, have shown that police databases are not a complete census of all criminal offences, nor do they constitute a representative random sample.²⁻⁵ Empirical evidence suggests that police officers – either implicitly or explicitly – consider race and ethnicity in their determination of which persons to detain and search and which neighbourhoods to patrol.^{6,7}

If police focus attention on certain ethnic groups and certain neighbourhoods, it is likely that police records will systematically over-represent those groups and neighbourhoods. That is, crimes that occur in locations frequented by police are more likely to appear in the database simply because that is where the police are patrolling.

Bias in police records can also be attributed to levels of community trust in police, and the desired amount of local policing – both of which can be expected to vary according to geographic location and the demographic make-up of communities. These effects manifest as unequal crime reporting rates throughout a precinct. With many of the crimes in police databases being citizen-reported, a major source of

What is predictive policing?

According to the RAND Corporation, predictive policing is defined as “the application of analytical techniques – particularly quantitative techniques – to identify likely targets for police intervention and prevent crime or solve past crimes by making statistical predictions”.¹³ Much like how Amazon and Facebook use consumer data to serve up relevant ads or products to consumers, police departments across the United States and Europe increasingly utilise software from technology companies, such as PredPol, Palantir, HunchLabs, and IBM to identify future offenders, highlight trends in criminal activity, and even forecast the locations of future crimes.

What is a synthetic population?

A synthetic population is a demographically accurate individual-level representation of a real population – in this case, the residents of the city of Oakland. Here, individuals in the synthetic population are labelled with their sex, household income, age, race, and the geo-coordinates of their home. These characteristics are assigned so that the demographic characteristics in the synthetic population match data from the US Census at the highest geographic resolution possible.

How do we estimate the number of drug users?

In order to combine the NSDUH survey with our synthetic population, we first fit a model to the NSDUH data that predicts an individual’s probability of drug use within the past month based on their demographic characteristics (i.e. sex, household income, age, and race). Then, we apply this model to each individual in the synthetic population to obtain an estimated probability of drug use for every synthetic person in Oakland. These estimates are based on the assumption that the relationship between drug use and demographic characteristics is the same at the national level as it is in Oakland. While this is probably not completely true, contextual knowledge about the local culture in Oakland leads us to believe that, if anything, drug use is even more widely and evenly spread than indicated by national-level data. While some highly localised “hotspots” of drug use may be missed by this approach, we have no reason to believe the location of those should correlate with the locations indicated by police data.

- bias may actually be community-driven rather than police-driven. How these two factors balance each other is unknown and is likely to vary with the type of crime. Nevertheless, it is clear that police records do not measure crime. They measure some complex interaction between criminality, policing strategy, and community–police relations.

Machine learning algorithms of the kind predictive policing software relies upon are designed to learn and reproduce patterns in the data they are given, regardless of whether the data represents what the model’s creators believe or intend. One recent example of intentional machine learning bias is Tay, Microsoft’s automated chatbot launched earlier this year. A coordinated effort by the users of 4chan – an online message board with a reputation for crass digital pranks – flooded Tay with misogynistic and otherwise offensive tweets, which then became part of the data corpus used to train Tay’s algorithms. Tay’s training data quickly became unrepresentative of the type of speech its creators had intended. Within a day, Tay’s Twitter account was put on hold because it was generating similarly unsavoury tweets.

A prominent case of unintentionally unrepresentative data can be seen in Google Flu Trends – a near real-time service that purported to infer the intensity and location of

influenza outbreaks by applying machine learning models to search volume data. Despite some initial success, the models completely missed the 2009 influenza A–H1N1 pandemic and consistently over-predicted flu cases from 2011 to 2014. Many attribute the failure of Google Flu Trends to internal changes to Google’s recommendation systems, which began suggesting flu-related queries to people who did not have flu.⁸ In this case, the cause of the biased data was self-induced rather than internet hooliganism. Google’s own system had seeded the data with excess flu-related queries, and as a result Google Flu Trends began inferring flu cases where there were none.

In both examples the problem resides with the data, not the algorithm. The algorithms were behaving exactly as expected – they reproduced the patterns in the data used to train them. Much in the same way, even the best machine learning algorithms trained on police data will reproduce the patterns and unknown biases in police data. Because this data is collected as a by-product of police activity, predictions made on the basis of patterns learned from this data do not pertain to future instances of crime on the whole. They pertain to future instances of *crime that becomes known to police*. In this sense, predictive policing (see “What is predictive policing?”) is aptly named: it is predicting future policing, not future crime.

To make matters worse, the presence of bias in the initial training data can be further compounded as police departments use biased predictions to make tactical policing decisions. Because these predictions are likely to over-represent areas that were already known to police, officers become increasingly likely to patrol these same areas and observe new criminal acts that confirm their prior beliefs regarding the distributions of criminal activity. 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. This creates a feedback loop where the model becomes increasingly confident that the locations most likely to experience further criminal activity are exactly the locations they had previously believed to be high in crime: selection bias meets confirmation bias.

Predictive policing case study

How biased are police data sets? To answer this, we would need to compare the crimes recorded by police to a complete record of all crimes that occur, whether reported or not. Efforts such as the National Crime Victimization Survey provide national estimates of crimes of various sorts, including unreported crime. But while these surveys offer some insight into how much crime goes unrecorded nationally, it is still difficult to gauge any bias in police data at the local level because there is no “ground truth” data set containing a representative sample of local crimes to which we can compare the police databases.

We needed to overcome this particular hurdle to assess whether our claims about the effects of data bias and feedback in predictive policing were grounded in reality. Our solution was to combine a demographically representative *synthetic population* of Oakland, California (see “What is a synthetic

population?") with survey data from the 2011 National Survey on Drug Use and Health (NSDUH). This approach allowed us to obtain high-resolution *estimates* of illicit drug use from a non-criminal justice, population-based data source (see "How do we estimate the number of drug users?") which we could then compare with police records. In doing so, we find that drug crimes known to police are not a representative sample of all drug crimes.

While it is likely that estimates derived from national-level data do not *perfectly* represent drug use at the local level, we still believe these estimates paint a more accurate picture of drug use in Oakland than the arrest data for several reasons. First, the US Bureau of Justice Statistics – the government body responsible for compiling and analysing criminal justice data – has used data from the NSDUH as a more representative measure of drug use than police reports.² Second, while arrest data is collected as a by-product of police activity, the NSDUH is a well-funded survey designed using best practices for obtaining a statistically representative sample. And finally, although there is evidence that some drug users do conceal illegal drug use from public health surveys, we believe that any incentives for such concealment apply much more strongly to police records of drug use than to public health surveys, as public health officials are not empowered (nor inclined) to arrest those who admit to illicit drug use. For these reasons, our analysis continues under the assumption that our public health-derived estimates of drug crimes represent a ground truth for the purpose of comparison.

Figure 1(a) shows the number of drug arrests in 2010 based on data obtained from the Oakland Police Department; Figure 1(b) shows the estimated number of drug users by grid square. From comparing 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. In Figure 1(a), we see that drug arrests in the police database appear concentrated in neighbourhoods around West Oakland (1) and International Boulevard (2), two areas with largely non-white and low-income populations. These neighbourhoods experience about 200 times more drug-related arrests than areas outside of these clusters. In contrast, our estimates (in Figure 1(b)) suggest that drug crimes are much more evenly distributed across the city. Variations in our estimated number of drug users are driven primarily by differences in population density, as the estimated rate of drug use is relatively uniform across the city. This suggests that while drug crimes exist everywhere, drug arrests tend to only occur in very specific locations – the police data appear to disproportionately represent crimes committed in areas with higher populations of non-white and low-income residents.

To investigate the effect of police-recorded data on predictive policing models, we apply a recently published predictive policing algorithm to the drug crime records in Oakland.⁹ This algorithm was developed by PredPol, one of the largest vendors of predictive policing systems in the USA and one of the few companies to publicly release its algorithm in a peer-reviewed journal. It has been described by its founders

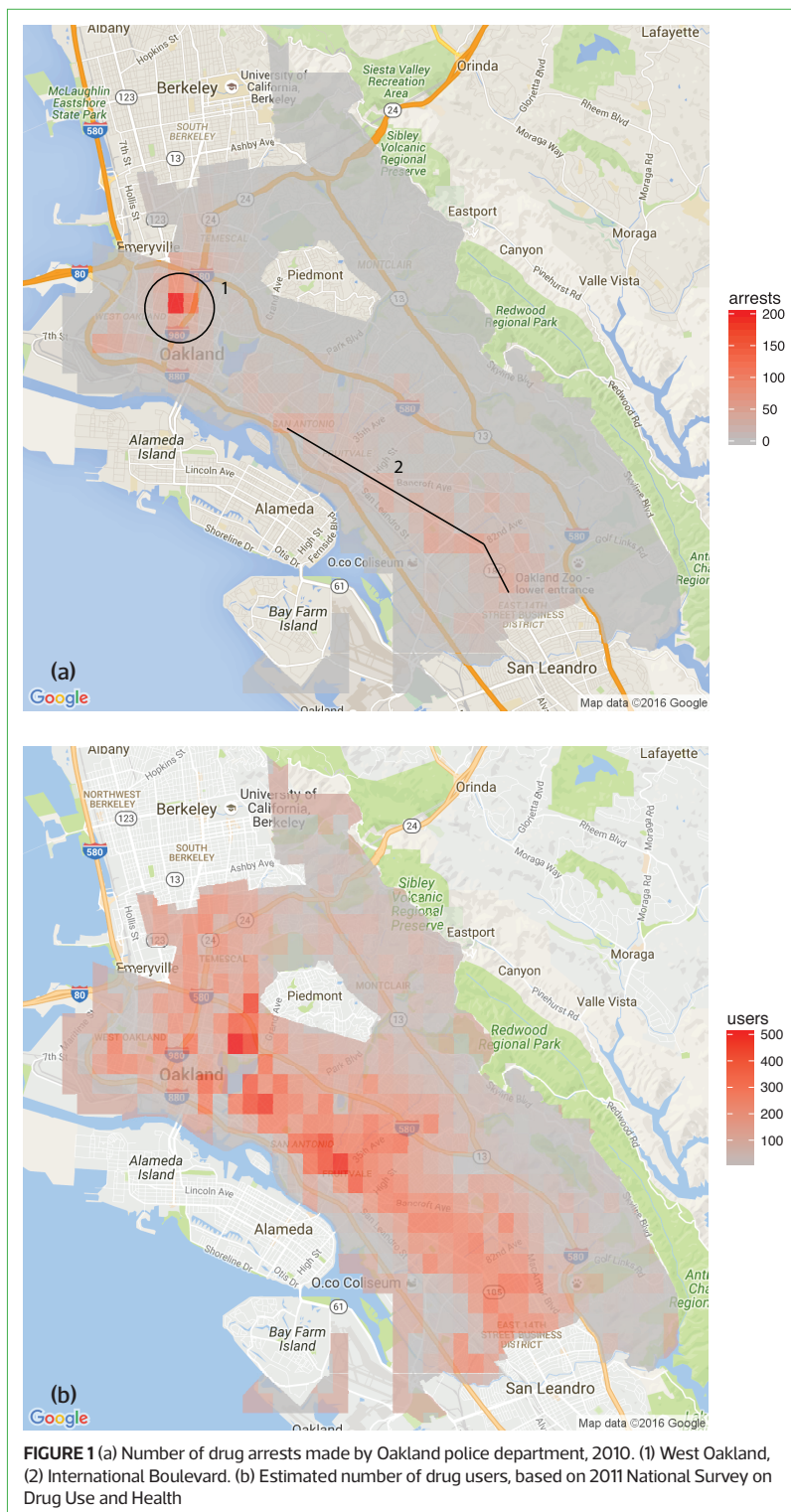
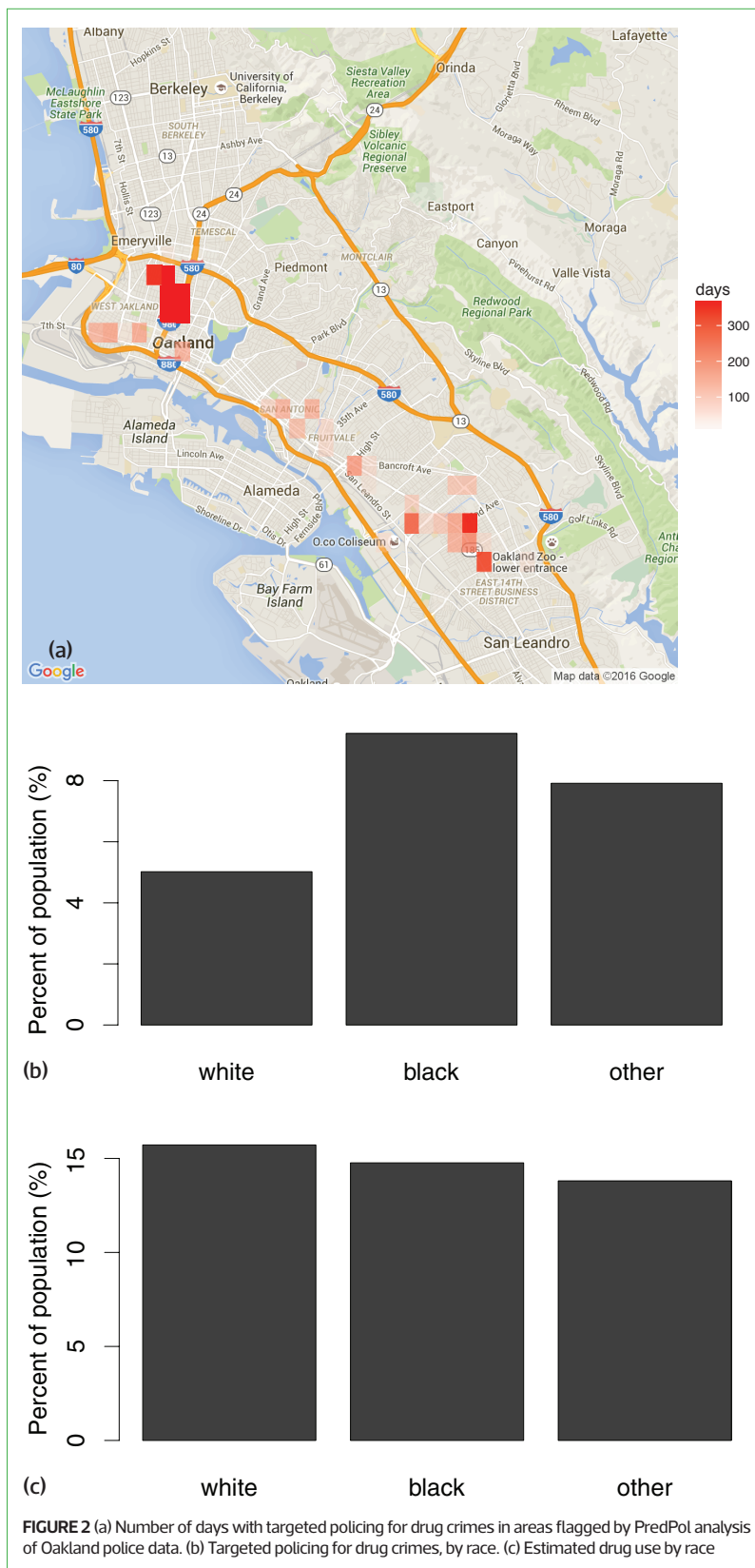


FIGURE 1 (a) Number of drug arrests made by Oakland police department, 2010. (1) West Oakland, (2) International Boulevard. (b) Estimated number of drug users, based on 2011 National Survey on Drug Use and Health



as a parsimonious race-neutral system that uses “only three data points in making predictions: past type of crime, place of crime and time of crime. It uses no personal information about individuals or groups of individuals, eliminating any personal liberties and profiling concerns.” While we use the PredPol algorithm in the following demonstration, the broad conclusions we draw are applicable to any predictive policing algorithm that uses unadjusted police records to predict future crime.

The PredPol algorithm, originally based on models of seismographic activity, uses a sliding window approach to produce a one-day-ahead prediction of the crime rate across locations in a city, using only the previously recorded crimes. The areas with the highest predicted crime rates are flagged as “hotspots” and receive additional police attention on the following day. We apply this algorithm to Oakland’s police database to obtain a predicted rate of drug crime for every grid square in the city for every day in 2011. We record how many times each grid square would have been flagged by PredPol for targeted policing. This is shown in Figure 2(a).

We find that rather than correcting for the apparent biases in the police data, the model reinforces these biases. The locations that are flagged for targeted policing are those that were, by our estimates, already over-represented in the historical police data. Figure 2(b) shows the percentage of the population experiencing targeted policing for drug crimes broken down by race. Using PredPol in Oakland, black people would be targeted by predictive policing at roughly twice the rate of whites. Individuals classified as a race other than white or black would receive targeted policing at a rate 1.5 times that of whites. This is in contrast to the estimated pattern of drug use by race, shown in Figure 2(c), where drug use is roughly equivalent across racial classifications. We find similar results when analysing the rate of targeted policing by income group, with low-income households experiencing targeted policing at disproportionately high rates. Thus, allowing a predictive policing algorithm to allocate police resources would result in the disproportionate policing of low-income communities and communities of colour.

The results so far rely on one implicit assumption: that the presence of additional policing in a location does not change the number of crimes that are discovered in that location. But what if police officers have incentives to increase their productivity as a result of either internal or external demands? If true, they might seek additional opportunities to make arrests during patrols. It is then plausible that the more time police spend in a location, the more crime they will find in that location.

We can investigate the consequences of this scenario through simulation. For each day of 2011, we assign targeted policing according to the PredPol algorithm. In each location where targeted policing is sent, we increase the number of crimes observed by 20%. These additional simulated crimes then become part of the data set that is fed into PredPol on subsequent days and are factored into future forecasts. We study this phenomenon by considering the ratio of the predicted daily crime rate for targeted locations to that for non-targeted locations. This is shown in Figure 3, where large values indicate that many more crimes are predicted in the targeted locations

relative to the non-targeted locations. This is shown separately for the original data (baseline) and the described simulation. If the additional crimes that were found as a result of targeted policing did not affect future predictions, the lines for both scenarios would follow the same trajectory. Instead, we find that this process causes the PredPol algorithm to become increasingly confident that most of the crime is contained in the targeted bins. This illustrates the feedback loop we described previously.

Discussion

We have demonstrated that predictive policing of drug crimes results in increasingly disproportionate policing of historically over-policed communities. Over-policing imposes real costs on these communities. Increased police scrutiny and surveillance have been linked to worsening mental and physical health;^{10,11} and, in the extreme, additional police contact will create additional opportunities for police violence in over-policed areas.¹² When the costs of policing are disproportionate to the level of crime, this amounts to discriminatory policy.

In the past, police have relied on human analysts to allocate police resources, often using the same data that would be used to train predictive policing models. In many cases, this has also resulted in unequal or discriminatory policing. Whereas before, a police chief could reasonably be expected to justify policing decisions, using a computer to allocate police attention shifts accountability from departmental decision-makers to black-box machinery that purports to be scientific, evidence-based and race-neutral. Although predictive policing is simply reproducing and magnifying the same biases the police have historically held, filtering this decision-making process through sophisticated software that few people understand lends unwarranted legitimacy to biased policing strategies.

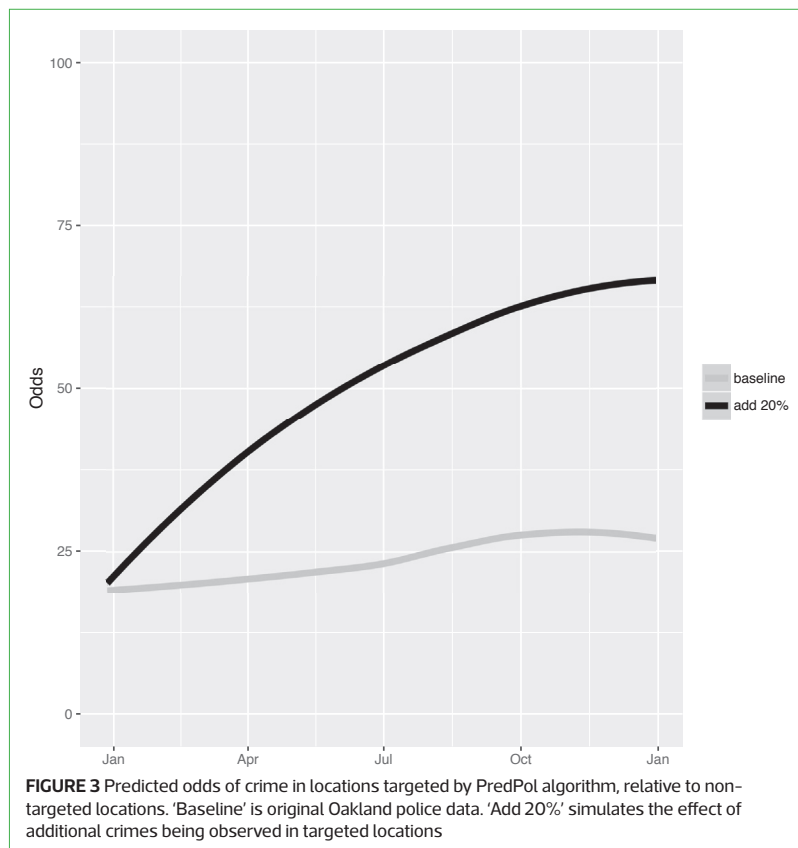
The impact of poor data on analysis and prediction is not a new concern. Every student who has taken a course on statistics or data analysis has heard the old adage “garbage in, garbage out”. In an era when an ever-expanding array of statistical and machine learning algorithms are presented as panaceas to large and complex real-world problems, we must not forget this fundamental lesson, especially when doing so can result in significant negative consequences for society. ■

Note

The authors would like to thank Bobbi Isaac, Corwin Smidt, Eric Juenke, James Johndrow, Jim Hawdon, Matt Grossman, Michael Colaresi, Patrick Ball and the members of the HRDAG policing team for insightful conversations on this topic and comments on this article.

References

1. Gerner, J. (2013) Chicago police use “heat list” as strategy to prevent violence. *Chicago Tribune*, 21 August.
2. Langan, P. A. (1995) The racial disparity in U.S. drug arrests. bit.ly/29B2pQu
3. Levitt, S. D. (1998) The relationship between crime reporting and police: Implications for the use of Uniform Crime Reports. *Journal of Quantitative Criminology*, **14**(1), 61–81.
4. Morrison, W. D. (1897) The interpretation of criminal statistics. *Journal*



of the Royal Statistical Society, **60**(1), 1–32.

5. Mosher, C. J., Miethe, T. D. and Hart, T. C. (2010) *The Mismeasure of Crime*. Thousand Oaks, CA: Sage Publications.
6. Gelman, A., Fagan, J., & Kiss, A. (2007) An analysis of the New York City Police Department’s “stop-and-frisk” policy in the context of claims of racial bias. *Journal of the American Statistical Association*, **102**(479), 813–823.
7. Lange, J. E., Johnson, M. B. and Voas, R. B. (2005) Testing the racial profiling hypothesis for seemingly disparate traffic stops on the New Jersey Turnpike. *Justice Quarterly*, **22**(2), 193–223.
8. Lazer, D., Kennedy, R., King, G. and Vespignani, A. (2014) The parable of Google flu: traps in big data analysis. *Science*, **343**(6176), 1203–1205.
9. Mohler, G. O., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L. and Brantingham, P. J. (2015) Randomized controlled field trials of predictive policing. *Journal of the American Statistical Association*, **110**(512), 1399–1411.
10. Sewell, A. A. and Jefferson, K. A. (2016) Collateral Damage: The Health Effects of Invasive Police Encounters in New York City. *Journal of Urban Health*, **93**(1), 42–67.
11. Sewell, A. A., Jefferson, K. A. and Lee, H. (2016) Living under surveillance: gender, psychological distress, and stop-question-and-frisk policing in New York City. *Social Science & Medicine*, **159**, 1–13.
12. Lerman, A. E. and Weaver, V. (2014) Staying out of sight? Concentrated policing and local political action. *Annals of the American Academy of Political and Social Science*, **651**(1), 202–219.
13. Perry, W. L. (2013) *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica, CA: Rand Corporation.