**MACHINE LEARNING APPLICATIONS TO CRIMINAL JUSTICE POLICY**

Maura Lievano

# Introduction

Imagine that a researcher can tell you with over 95 percent accuracy what is the likelihood that a person might re-offend, carry a weapon, commit a crime or desist from committing crime, given a set of individual characteristics and environmental variables. As an ordinary person,  you can now have a probability -based on real world data- as a reference to understand and navigate the world. For a criminal justice official, say a police officer, prosecutor, public defender, or judge, this information is crucial and necessary to make informed decisions when interacting with individuals involved with the criminal justice system. Officials can use this information to allocate scarce public resources to competing  public safety strategies. Computational statistical learning  can provide such answers.

In fact, questions related to probabilistic assessment of crime, recidivism, gun violence, or an individual’s sentence compliance are typical classification problems for which supervised learning techniques have effective solutions.

**Footnote:**Generally speaking, big data is considered to have three attributes (the three V’s) Volume, Velocity, and Variety and has in the scope of hundred of millions of observations. Fat or wide data is a term used to refer to a dataset with several predictors. In social sciences big data is increasingly used in combination with machine learning to tackle a wide extent of issues, from medical to service consumption patterns.

In criminal justice and other areas of policy, supervised and unsupervised statistical learning algorithms have provided a way to tackle prediction questions and implement solutions where traditional statistical techniques, such as ordinary least squares regression have failed. Some popular examples from public health, include emergency room triage (Almeida et al. 2014), cancer diagnosis22Miles Wernick known for his work on prostate cancer detection is also the leading researcher of the predictive policing program of the Chicago Police Department and the National Institute of Justice.(Ozer et al. 2010; Kourou et al. 2015), and prevention of childhood lead poisoning (Potash et al. 2015). Additionally, targeted learning, the use of machine learning algorithms within a causal inference framework, offers alternatives to improve over the current methods of causality in observational studies.(Petersen, Sinisi, and van der Laan 2006; Van der Laan and Rose 2011)

In criminology, there is a considerable body of empirical research focusing on different criminal justice issues from a predictive perspective.  Core functions of police offices, prosecutorial institutions, public defenders’ offices, probation departments, courts, and prisons, have been analyzed through the lenses of statistical learning.

The work on policing strategies and racial discrimination of (Goel, Rao, and Shroff 2016) and (Berk and Bleich 2014) on courts and sentencing and forecasting criminal behavior, as well as in risk of criminal behavior of parolees and probationers  (Berk et al. 2009; Monahan and Skeem 2015) are cases in which statistical learning has been a crucial tool in providing robust solutions to improve the performance of criminal justice institutions. This approach to criminal justice, sometimes called actuarial, has had several critiques from practitioners and academics For example, former US Attorney General Eric Holder has opposed to this view arguing that statistical learning tools use immutable individual traits over which persons do not have control and cannot possible change in the short run to assess criminal behavior; in his view, those features such as education, socioeconomic status and neighborhood, when included in designing . algorithms will only deepen the existing disparities affecting the poor(Holder 2015).

Furthermore, the discussion has not explored applications of machine learning methods for causal inference. The limitations of parametric models can, in many instances, be offset by computational statistical learning. Targeted learning, a technique that allows to estimate a single parameter of interest with machine learning methods has been successfully applied to estimate the effect of different policies at the individual level.

In this paper, I focus on  the question of the extent to which machine learning is a tool to better understand data related to criminal justice policy . To this end, I organize this paper as follows.  The **first section** provides working definitions of machine and statistical learning, and presents common statistical algorithms used in the field of criminology.  Then, in the **second section**, I discuss a set of cases of criminology problems analyzed through the lenses of machine learning methods. These cases were selected to illustrate the practical application of machine learning to criminal justice and provide insight into the debate about the role of computational statistical techniques in  dealing with real world policy issues.

Then,  in the **third section**of this paper I review the case of assessment of risk of recidivism and violent behavior in order to illustrate the two poles of an undergoing debate about machine learning and criminal justice.  At the center of the such debate is the guiding question of this article:  are such statistical methods suitable to understand criminal behavior and design strategies to promote public safety?

The choice of risk assessment tools used in sentencing to analyze the  usefulness of using machine learning in the field of criminology serves two purposes. **One** is to present the discussion in a practical fashion and not only in mathematical terms.  Decision making processes such as sentencing are a practical  and relevant for society as a whole, as it affects the lives of individuals and impacts public safety. The **other objective**is to exhibit the common misunderstandings and pitfalls that occur when academics and public officials communicate. Risk assessment in sentencing meets both aims.  In the **fourth section** of this paper, I conclude by answering the guiding question and identifying ways to make the use of machine learning productive for criminal justice policy.

# What is Machine Learning?

Machine learning is a sub-field of computer science based on the study of pattern recognition using computational tools (software and hardware) in order to identify mathematical rules. Such mathematical rules  allow to predict future outcomes from existing data. Machine learning can be broken down into three actions, namely, defining an algorithm, training the algorithm on data, and collecting an expected outcome using using a computer  and specialized software.  The mediation of computers  to learn from data explains the use of the term computational statistical learning as a synonym of machine learning.

Thus, the characteristic feature of machine learning is the use of computational tools understood as  powerful computer hardware and statistical software, both needed to handle large datasets. Aside from this characteristic, the elements of machine learning are essentially the same than those of the statistical methods used in the last century in criminal justice.  Some examples of such statistical methods of estimation include expert’s forecast of criminal behavior in a neighborhood, a linear projection of crime relative to a set of sociodemographic variables,  an ordinary least squares regression procedure to establish crime incidence or the computation of a score of risk of violent behavior on the basis of a test designed by professionals with clinical expertise.

In social sciences, including criminology, research involving statistical analysis has shown a growing reliance on specialized software first developed in the 1960s. Below a table shows the most widely known specialized computer programs used in social sciences, including criminology, particularly in subfields such as quantitative criminology.

**TABLE HERE**

SPSS, Stata and R are arguably the most widely used programs to conduct statistical analysis in criminology studies using quantitative data. SPSS and Stata are paid programs, while R is an open source software accessible free of charge and crowd-sourced, as any individual can contribute with a library or package containing algorithms to perform computations and implement statistical model.  A relevant caveat to examining the trend in use of these three programs is the fact that only R supports the implementation of machine learning algorithms and unlike Stata and SPSS is increasingly improving packages to efficiently conduct complex and recursive computations using big data. It is only in the past few years that Stata researchers have placed growing attention to the development of implementations of popular machine learning techniques such as random forest and support vector machine.

Thus data availability paired with computational power allow for the use of elaborated mathematical formulas, as opposed to simple linear ones, and a more frequent use of wide and big data to answers policy question (billions of observations of multiple variables in different formats), as opposed to small data (imaginable number of observations of a few defined variables).

DISTINGUISHING STATISTICAL METHODS

Methods for

Machine Learning and other methods

Machine Learning of Actuarial, Clinical

Using as reference the working definition of machine learning provided above, we can distinguish these methods from those called actuarial or clinical. The term actuarial

With the previous definition of machine learning

Many of the critiques of statistical prediction methods are vague in terms of the specific methods that are employed by researchers.

The terms actuarial approaches, statistical learning and prediction are used liberally when scholars and practitioners blame the black box as the cause of bias, particularly of racial biases, when it comes to the potential dangers of using machine learning to support criminal justice decisions.

This type of argumentation that targets the messenger rather than the message fits the definition of a logical fallacy where the undesirable results of the implementation of an algorithm are mistaken for the algorithm itself. In this sense, I argue that statistical methods are sometimes the red herring for unexpected results of policy design and implementation. 33For example, Bernard Harcourt in his article The Shaping of Chance: Actuarial Models and Criminal Profiling at the Turn of the Twenty-First Century, critiques actuarial models in criminal justice and nowhere in the text defines the term. It is a nice touch to quote Jean-Luc Godard as a sign of the philosophical tradition supporting his critique; however, ”being slaves of probabilities” might not be a bad thing when it can outperform human judgment in guaranteeing societal outcomes.

It is true that the higher complexity and limited to null causal inference properties of predictive algorithms are indeed problematic. But, as empirical studies have shown, the purpose, structure, validation, and implementation of statistical learning techniques to policy are in no way homogeneous or easily labeled in a single category. Additionally, the variety of machine learning methods is broad and some algorithms are relatively easier to understand than others, and, in some cases, the explanation of the basic two-dimensional underlying model can make the typology of methods manageable and understandable, in a way that it is possibly to associate different criminal justice policy questions with specific algorithmic settings. To illustrate what I mean by that, in the following paragraphs I provide a brief overview of the machine learning methods more broadly used in criminology research. 44Explain why it is sort of a justifiable sample of studies consulted

As mentioned, the main distinction between machine learning and traditional statistical learning is the mediation of computers to obtain information from data.

A defining characteristic of machine learning is, therefore, a technological interface to guide the process of learning. According to this definition, the line dividing statistical learning from machine learning is small, and although there is no single definition, in practice by ”mediation of computers”, a high computational power, enough to scale and study big data is assumed. Machine learning is broadly divided into supervised and unsupervised, where the former requires data of an outcome variable (numerical or categorical) and a set of features or input variables, associated with that outcome. The later, does not require information about the outcome to guide the process of learning (Friedman, Hastie, and Tibshirani 2001).

A distinctive trait of areas where machine learning is applied is when large databases are available and what is called wide data is at hand, wide data refers to a large number of variables or traits, in database terms it indicate the number of columns. In the presence of multidimensional data, techniques to identify association or select variables are valuable and statistical learning achieves this goal. A typical statistical learning problem is that of classification into binary categories, such as having a disease versus not having a disease, success or failure or, high or low risk.

A classification of statistical analysis techniques is provided by (Varian 2014), who identifies four categories - prediction, summarization, estimation and hypothesis testing- and considers that machine learning is mostly related to prediction. In this paper, I focus on prediction techniques and mention briefly machine learning methods for classification and visualization, as I consider them a practical tool for exploratory data analysis usually conducted prior to prediction.

Supervised and unsupervised learning, as defined previously, build learning models to either predict/classify (supervised) or represent associations among variables (unsupervised). Among the most important and used families of techniques in the fields of statistics, social research, engineering, finance and artificial intelligence, are the following: 1) clustering and principal components analysis, 2) linear methods for variable selection, 3) tree methods, 4) boosting, bagging, and bootstrapping.

The central objective of supervised learning algorithms for prediction is to make quality – i.e., precise and unbiased – out-of-sample forecasting of events. For this reason, predictive algorithms in machine learning usually are trained on a subset of the data (trainer or training set) and the remainder, (the test set) is used to evaluate the performance of the algorithm and its parameters in out of sample observations. Validation and cross-validation of algorithms are statistical techniques to increase the prediction power of a model so that it can perform well when used to forecast out-of-sample events. In addition to validation in the test set, there are two basic protocols to decrease out-of-data forecasting error, *k* -fold validation and holdout approaches, being the former most common. *k* -fold validation works by subsetting the data into *k* folds (Where K ¿ 2) and testing each recursively to calibrate the parameters. This way of validating the performance of an algorithm allows researchers to take advantage of big data sets and prevents overfitting of the algorithm to the data at hand. There is no general rule as the proportion of the size of the training set to the validation set, and of the number of folds to use in multiple fold validation. As for the holdout, this is typically a five to ten percent of the original dataset that is excluded from all analyses and used once the final model is selected (through one of the previous protocols) to fine tune the parameters. These validation protocols are ways of taking advantage of large amounts of data and computational capacity to expose the model to many different potential distributions of the observations, calibrate parameters and a good performance outside the original data.

## Clustering Methods

Visualizing Data: Clustering and principal component analysis are both canonical examples of techniques used in unsupervised learning to find association among variables, identify latent variables, and reduce data dimensions. Clustering is a family of methods that group variables according to a given definition of similarity - or dissimilarity, this approach is also used as a form of describing data as it provides information about the existence or not of groups or subgroups.

The definition of similarity can be thought of as the loss or cost function that is typical in prediction problems, such as the square distance, in the case of the more traditional ordinary least squares regression. Similarity is the central concept associated with clustering methods. This notion, although mathematically represented, can only be defined based on knowledge of the subject matter and the specific variables available as the most important to operationalize the definition in the algorithm.

For example, while minimizing distance across observations can be central to one definition of similarity, for others a sharing a particular attribute or a specific number of groupings may be key to the definition of similarity. Based on such definition, the objective of the computations is to group elements such that pairwise dissimilarities between elements of a group are smaller that those in different groups.

Clustering a large number of variables and observations becomes complex pretty fast. There are several popular algorithms to manage the grouping, namely, k-means, k-medoids, proximity matrices, nearest neighbors, self-organizing maps, spectral clustering, and so on. These algorithms can be classified into three types according to their assumptions about the way in which the elements to be grouped are distributed: combinatorial, mixture modeling, and mode seeking. Combinatorial models do not assume any particular probability distribution of the observations; mixture modeling assumes that the observations are independent and identically distributed and mode seeking is a non-parametric form of clustering that estimates modal probability density functions.

Clustering methods are suitable for the analysis of geographically based data. In criminal justice it is a popular technique to detect hotspots or to deploy resources based on selected characteristics of neighborhoods in order to prevent crime. (Eck and Weisburd 2015) conducted a review of 14 studies using different clustering techniques concluding that mapping crime patterns is helpful to prevention strategies, as the benefits of policies are diffuse in unprotected locations. This dispersion suggests that using clustering to adapt policies by geographic location is an efficient approach to prevention. One example of this particular approach to policy is illustrated by (Singh 2006), who used hierarchical clustering techniques to identify and map areas of high crime concentration and high risk of crime to design a prevention program targeting those spots and specific crimes.

Clustering techniques are complex and the definition of similarity is critical when applying such algorithms to policy, as illustrated by (Grubesic 2006). The author applied several clustering techniques to crime data in Cincinnati, Ohio. To identify areas of high crime concentration, the author concluded that fuzzy clustering is the appropriate approach to delineate hotspots in urban settings, Grubesic also showed that the spatial configuration of crime in his analysis changed with different algorithm specifications, showing how different concepts of similarity and grouping produce varied results.

## Principal Components Analysis

Principal components analysis or PCA is a technique that creates new variables based on existing variables, the new variables being the so-called components. Through a multiplication of the eigenvalues of the existing variables, the PCA produces the components which are orthogonal transformations of linearly dependent predictors and helpful to reduce the number of dimensions of a dataset. Applying the technique implies matrices operations that transform the original vector of parameters into their corresponding matrices of eigenvalues. Then, via multiplication, the new variables or so-called ”components” are created and represent combinations of the eigenvalues of the original predictors. PCA techniques are a common way of addressing multicollinearity.

It can be validly argued that the components do not have a practical meaning when talking about policy, as they represent a mixture of the original variables; however, these operations allow researchers to identify the original variables that have highest variance across their values and combine them into ”components” that represent those variables or features that combined yield a more precise measure of variation across observations in a dataset. Examples of application of PCA techniques in criminal justice are several. Among them, (Cooper et al. 2016) explore how race and ethnicity variables affect drug consumption and criminal activity in different groups, and (Ayoola et al. 2015) estimate crime rates in Nigeria using PCA techniques.

## Linear Methods

Linear models for regression or classification work under the assumption that whatever the underlying data generation process of the subject of study, it can be represented as linear. The simplicity of this type of model makes it powerful, under certain conditions it can outperform complex non-linear algorithms in the task of regression and classification (Friedman et al., 2001). Linear regression, as a method of the pre-computer era is the most popular one among social researchers. The parsimony of its structure and its power to identify causal effects once the assumptions are met make of this approach the preferred among social scientists, including criminology. However, satisfying the assumptions of a causal model is a non-trivial endeavor, as in criminology data is rarely normally distributed and variables uncorrelated. As Richard Berk puts it, in practice, the practical scope of regression analysis in criminology might be overestimated (Berk 2010).

(Varian 2014) considers that issues with big datasets require different tools. The author specifically mentions data manipulation tools such as computational software, mathematical algorithms for variable selection, given that there is more availability of potential predictors for estimation and techniques to model non-linear relationships, factors that machine learning offers.

### Ordinary Least Squares

Among linear methods for regression three categories stand out: least squares, subset selection, and shrinkage estimation, these three categories have in common an optimization function that typically consist in minimizing a given loss function. In the case of least squares, the optimization function is the minimum of the squared errors across observations. This model is widely used, but it has two main limitations. First is the low prediction accuracy due to the low bias but high variance of the coefficient estimates. The second limitation concerns the problem of interpreting coefficients in the presence of multiple predictors (Friedman et al., 2001, p. 56).

However, ordinary least squares regression is still the most widely used method in practice, as illustrated by the proportion of academic articles published in journals. Given the limitations of this approach, as mentioned above, it is important to consider what are the expectations and the scope of such analysis based on OLS regression. (Berk 2010) characterizes three levels of possible regression analysis as follows, Level I, descriptive regression analysis, is an exploratory exercise with no assumptions about the data generation process that simply describes patterns and relations observed across variables. Level II refers to inferential statistical analysis, which requires a well-defined population and a sample obtained through a probability sampling technique, where the probability of each observation being selected is known. This type of analysis uses hypothesis testing, confidence intervals and estimation of key parameters to add statistical inference to the description, but does not convey any causal finding. Finally, Level III, or causal regression analysis, requires a model specification with very small room for error and compliance with strong assumptions about the data generation process in order to make causal claims about the regression parameters. Model specification and assumptions are necessary to make this type of analysis work, but in criminology this type of analysis is not common by the nature of the data.

From a similar perspective, (Varian 2014) considers linear regression to be a conventional statistical technique that has been widely used in social sciences, but has substantially changed in its relevance as the availability of new data and computational power has benefited the development of new algorithms to analyze data. The author reviews the main machine learning algorithms and shows some examples of their comparative advantages to identify and modeling non-linear relationships among variables.

### Logistic Regression

Logistic regression is a powerful tool to model the probability of K possible outcomes using a linear function and ensuring that the probabilities add up to 1, and in the range of [0,1]. When K =2, typically a binary classification problem, the model is simple with a unique linear function. Variations of logistic regression, such as sparse logistic regression in its binomial or multinomial version have proven to be useful in complex classification problems and text analysis (Sculley et al., 2011). (Sculley and others 2011).

### Subset Selection

One potential solution to tackle the problem of multiple predictors is to select a subset of variables, in which the least squares technique is applied to only a subset of selected variables. To do so, there are several techniques, such as best-subset selection, forward-and-backward step-wise selection and, forward-stage-wise regression.

A *subset selection* method finds a subset of size *k* that yields the smallest residual sum of squares; this procedure can be efficiently applied for a number of predictor *p* of around 40 (Friedman, Hastie, and Tibshirani 2001). The decision to choose *k* is usually a trade-off between bias and variance, as well as a researcher’s preference for parsimony, but typically a k is chosen that minimizes the expected prediction error. To find this *k*, the algorithm searches through all possible combinations of predictors, ordering the possible sets.

When the number of predictors is substantially over 40 a forward-step- wise selection of variables can be applied, this method builds a model adding one variable at a time starting with the intercept and following with the variable that is a best fit. Inversely, the backward-step-wise technique starts with a full model and excludes one by one the variable that most affect the fit. When there are a large number of variables these methods are heavily computational but offer several potential models to set on the trade-off between explanatory power and parsimony. Because these methods add or subtract a variable at a time, in a discrete process, they might lead to high variance, affecting the overall prediction accuracy of the final model. This is one of the reasons why shrinkage methods aimed at reducing the number of variables but in a continuous fashion might be preferred.

### Shrinkage: Ridge and Lasso

Selecting a few variables has the advantage of providing a more parsimonious model that can be easily interpreted and has more prediction accuracy than a full or bigger model. Instead of selecting one variable at a time to achieve a more parsimonious model, shrinkage methods introduce a penalty associated with the number of regression coefficients. The penalty parameter, usually denoted by , determines the size of shrinkage. There are two common shrinkage techniques, ridge and lasso.

Ridge shrinkage penalizes the sum of square residuals and pushes the estimators towards zero and each other, and do not penalizes the intercept (Hoerl and Kennard 1970). The penalization parameter is the norm L- 2 or median. A similar approach is Lasso regression, a method that also penalizes coefficients with the important difference that it does so by applying a penalization of a different form, namely L- 1 or the mean making the solution a quadratic computation. An alternative approach to deal with a large number of parameters and multicollinearity is using principal components as regressors.

## Tree Methods

Decision trees are a tool for classification problems, when the objective is to predict a 0-1 outcome or, in other words, place an observation in one of two mutually exclusive categories. Examples of classification problems would be to place individuals into high-risk or low-risk segments, predict if an individual will or will not develop a disease, label an email as spam or not-spam, or other discrete categories. Although trees are mostly used for classification problem, they can also be used in a regression setting by using the leafs or branches as variables.

The classification task is based on a set of predictors and might be carried out via a logit or probit model.

However, an alternative to these methods is to grow a tree classifier that models a sequence of partitions. While a partition can only handle two variables, a tree manages an unlimited number of predictors and, what is more important, there are efficient computational ways to carry out this process.

This method is particularly helpful for settings where there are relevant non-linear relationships and interactions among variables. It also happens to handle missing data very well.

A good example of the application of this technique is the Titanic survival, described by (Varian 2014), in which a classification tree shows extreme age, namely, being very young or very old was decisive in survival rates, but for passengers in the middle of the distribution of age, variables different than age played a most important role. These types of insights from data are easily extracted by a tree algorithm and cannot otherwise be revealed. An example that I particularly like for its policy implications is the data analysis conducted by (Varian 2014) using the same data set than (Munnell et al. 1996) use to analyze the effects of the Home Mortgage Disclosure Act enacted in 1975 in the access of low-income individuals to the housing market.

In their paper, using a logistic regression approach, (Munnell et al. 1996) concluded that minorities are more than twice as likely to be denied a mortgage as whites; therefore, they maintain, ”race continued to play an important […] role in the decision to grant a mortgage”. Yet, several years later, Varian used a random forest method, a variation of the tree method described in previous paragraphs, to analyze the data on which the original article was based, Varian unveiled a slightly different conclusion: race was not the first but the second most important variable in explaining the difference in mortgage credits granted between the two groups. It turned out that ”dmi” or denial of mortgage insurance - a previous step in the process of applying for a mortgage credit- was the first predictor, which accounted for improved prediction accuracy of the random forest model by 10 percent, classifying adequately 223/2380 cases.

The narration of the authors regarding this variable suggests that previous banking history as well as other predictors of economic stability were generally more positive among whites relative to minorities, which can be attributed to a general economic trend and underlying dynamic, rather than to the specific event of insurance authorization. It might be argued that race, again, was the main predictor in denial of an insurance that later in the process will influence the decision of approval or denial of a mortgage; however, in terms of tracking the process by which minority individuals were denied mortgage credit, it is salient that the influence of race is partly exogenous, as insurance is necessary to access a housing credit.

While a single decision tree is relatively easy to interpret, techniques such as random forest, where hundreds of random trees are created and de-correlated is more difficult to make sense of, partly because of the non-linearities and interactions between variables that end up in different partitions. Another potential issue with tree methods is that they tend to overfit the data; for this reason, pruning the tree is important and that is a decision of the researcher as to where to stop partitioning so that the resulting classification algorithm can be useful to analyze new data efficiently (Friedman, Hastie, and Tibshirani 2001).

## Boosting, Bagging and Bootstrapping

Boosting is relatively new method, introduced in the last couple of decades, mostly used to transform a weak learning algorithm into a strong one by adapting to weighted versions of the training dataset and adopting a majority rule to produce a final prediction. One of the best known algorithms of this type is Adaptive Boosting or AdaBoost (Freund, Schapire, and Abe 1999).

Bagging stands for ”bootstrap averaging” or ”bootstrap aggregating,” which is a method that uses bootstrap samples to train an algorithm and average the results obtained for each sample to produce a classifier with reduced noise. This approach can be applied to virtually any algorithm to produce a model with parameters calibrated from as many bootstrapping samples as the computational capacity makes possible. This approach is particularly useful for unstable procedures such as tree methods (Breiman 1996).

## Targeted Learning

# Machine Learning and Criminal Justice

ADD somewhere here: Similarly, (Harcourt 2003) has argued that actuarial approaches exacerbate the racial imbalance in the prison populations and do not solve the root problem of having too many incarcerated individuals. Harcourt maintains the strong position on the topic and proposes to stop using computational statistical approaches to justice and focus on the early stages of the criminal process, that is at the point of intake. The author argues that algorithms in the prosecutorial or sentencing stage (2003, 2010) will reproduce the biases generated earlier in time, which he considers the origin of problems and racial disparity in the criminal justice system. It is not clear if he believes that statistical approaches can help officials dealing with the first stages in the criminal process to handle their functions better, for example in providing heuristics for detention, pre-trial release or propensity for success in non-custodial treatment.

Proponents and detractors of the use of machine learning algorithms to support decision making in criminal justice have several points of dissent that can be organized in two general types, one is mathematical and the other is in regards to policy-making. The mathematical concern is the algorithmic complexity and opacity of machine learning techniques, often called ”black box” methods.  The main concern, from a policy perspective is implementation and its consequences. Thus,  the idea that black box approaches might generate or exacerbate biases, especially racial discrimination is one of the main focus in the public debate. A lack or clarity and rigor in the debate has lead to mistake the policy critiques  with those challenging  implementation.

The application of clustering methods, principal component analysis, regression, tree methods, bagging and bootstrapping as techniques to prediction and classification problems using large data sets is relatively recent, as the computational power needed to handle millions of data points and numerous predictors has only recently become widely available. Prediction itself however is not. In criminology, prediction has been a common approach and it is documented in the academic literature, as well as in government documents and practices since the early 1920s (Ohlin and Dudley, 1949).

While machine learning algorithms are quite good at providing answers to classification and prediction problems, this approach is considered a ”black box” when it comes to interpretability and causal inference. There is concern among criminal justice policy researchers and practitioners as to the effects of the opacity of algorithms in producing and exacerbating undesirable outcomes such as racial discrimination (Starr 2016; Goel, Rao, and Shroff 2016; Harcourt 2008; Holder 2015). Several examples of this peril illustrate the claim, as a frequent critique of the risk assessment tools used in 27 American States currently utilizes actuarial approaches to support the decisions on parole, bail, and sentencing.

The black box is an issue of concern, especially because sometimes algorithms used in criminal justice institutions to guide decision process are developed by the private sector and not by public institutions, adding layers of complexity to the matter of public transparency. Such is the case of the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS), developed by Northpointe Inc. and challenged at length by the investigative journalism of the non-profit ProPublica11Since May, a discussion has unfolded regarding ProPublica’s claim that COMPAS might have a bias against African American individuals, for more on this claim see —(Angwin et al. 2016)—.

In criminal justice and other policy areas there is a growing trend to advocate for transparency in design and implementation of machine learning algorithms in order to prevent unintended consequences of data-driven policy decision making.

It would be fair to say that this trend pointing out the need for transparency in algorithms is analogous to the demand for transparency and accountability in the criminal justice system policy making and implementation. However, there is one particular aspect of development of machine learning tools to criminal justice policy that make this situation specially complex, and that is the private institutions behind the development of such data-driven tools. As private providers generate algorithms and do not disclose the process of design and validation, public institutions have less control over the process, limited tools to adjust the algorithm and fewer elements to be accountable to the public.

As (Petersilia 2014) points out, in addition to the mathematical make up of the algorithm, two factors are relevant to ensure that actuarial tools are effective in achieving their goals: time and location.

In other words, the author suggests that algorithms must be time and location sensitive and provides the example of the Missouri Sentencing Advisory Commission that uses an actuarial tool called Automated Sentencing Application (ASA), available online 22http://www.mosac.mo.gov/ to support the decisions of judges. Petersilia further explains that such a tool, based on a set of factors both fixed and variable, such as age, gender, education, ties to the community, type of crime, and employment status, generates a report with suggested sentence length. The author points out that the ASA was developed in the 1980’s when unemployment rates were below five percent, so being unemployed was very unlikely and was a stronger predictor of criminal behavior; however, over 30 years later, this factor is not as strongly predictive.

**rewrite** Although actuarial methods have been used since the 1940’s when actuarial methods for risk assessment were a promise of science to control crime and reduce recidivism (Petersilia 2014), it is not until recent years that the growing incarcerated population in the United States, currently the country with the highest rate incarceration rate in the world33In 2016 the rate per 100,000 inhabitants is 737 and the United States has lead the list in the last decade, according to the data published by Prison Policy Initiative brought attention to different alternatives to curb the trend (Raphael and Stoll 2013). Evidence-based policy and the application of machine learning algorithms arise as practical tools to be used in criminal justice institutions to control crime and provide public safety.

\pretitle\posttitle\preauthor\postauthor\predate\postdate

**end rewrite**

The capacity of machine learning algorithms to predict certain outcomes with a high level of accuracy, and reveal non-linear relationships among variables are considered the main advantages of these methods over conventional assessments made by experts and inferential studies based on linear regression (Berk and Bleich 2014; Athey and Imbens 2015; Varian 2014).

Furthermore, some researchers argue that these predictive methods provide a solid and superior alternative to both clinical or expert assessments in the case of practitioners and linear regression models in criminology research, where data available typically compromise the basic assumptions that produce desirable statistical properties for causal inference.

**rewrite** No matter the predictive merits of statistical learning argued by researchers and practitioners, there are as many proponents as detractors in both the academic and the practitioner side of their use in policy. They are a pink elephant in the room of the many criminal justice institutions at the state and federal levels currently using predictive tools.

\pretitle\posttitle\preauthor\postauthor\predate\postdate

**end rewrite**

The discussion has advanced beyond the mathematical black box, which in itself is a matter of concern, to the implementation realm, as evaluations of the effect of evidence based policies have suggested that bringing algorithms to the field does not have the expected results and can even bias the behavior of criminal justice officials in an unpredictable way. This was seen in the broadly known Chicago policing experiment, which included the release of the Strategic Subject List (SSL) containing the names of 426 individuals identified by a predictive algorithm as being at high risk of being involved in gun violence events.

In a recently published evaluation of the implementation of this program that has been running for three years, (Saunders, Hunt, and Hollywood 2016) found no conclusive evidence of its effectiveness but indicators suggesting that police officers behavior might have been used the list to disproportionately target those individuals, regardless of their actual behavior. It can be argued that the discriminatory effect of the SSL cannot be attributed to the algorithm, but rather was not an algorithm- the one being challenged, as it classified individuals by their risk of being involved -as victims or perpetrators- in gun violence episodes, according to certain classification rules and set of predictors - which, by definition do not imply causation- but failed to anticipate the heuristics of public officials in using the tool in the field.

Given the complexity of the problem of policy implementation and critiques of detractors that point to policy outcomes rather than policy design, it would be reasonable to state that critiques of actuarial methods in criminal justice are aiming at the wrong target, as it seems that the focus of the challenges mix the part of algorithmic transparency and policy design and implementation.

This framework puts in context the assertion of statistician Richard Berk (2013) who states that quantitative criminology tools are a black box but ”no apologies are made” for such a character because they are effective in predicting criminal behavior. Similarly, \citeauthor{kleinberg2015prediction} \citeyear{kleinberg2015prediction} maintain that no apologies are necessary if such methods succeed in producing key information for policy decisions.

Beyond the politics of implementation, there is broad consensus in the powerful contribution of statistical learning to approach critical criminology questions, such as how to allocate resources for policing operations (Perry 2013), frame the decisions of judges (Ridgeway 2003), analyze and rethink the effects of stop-and-frisk policies (Goel, Rao, and Shroff 2016) or determine the risk and need level of individuals involved with the criminal justice system to provide treatment (Monahan and Skeem 2015). Still, some criminal justice areas are more problematic than others in the degree to which machine learning is seen as a tool to improve human-made decisions by providing data. One of such areas regards assessment of individual risk based on population level data.

Perhaps, the most critical application of risk assessment tools is the one that determines the decision of placing an individual in incarceration or leaving him free to comply with non-incarceration measures, as well as the determination of sentence length. This decision is arguably the highest form of power of the state over the individual; the responsibility of taking such a decision is made by a person, the judge. Under this scenario, the availability of tools to make a ”fair” decision, pondering individual and collective benefits for society is key to improve the ability of the state to provide public safety.

For this reason, in the next section I will further analyze the use of risk assessment tools in the sentencing stage of the criminal process, as I consider this case a good example to illustrate in a practical way many of the valid and invalid critiques of machine learning and actuarial methods as a black box to be discarded and replaced by prevention policies or, more broadly a more humane and less automated approach. I expect this example to shed light into the black box and distinguish the algorithmic transparency from criminal justice policy design and implementation in a way that is beneficial to the field.

## types or risk assessment tools

fairness and bias in clinical, actuarial and adaptive-algorithmic assessment of criminal behavior and treatment needs is like a family issue: there might be disagreement but we all have the same purpose in mind: reducing crime rates and favor public safety in an efficient manner.11There is no current consensus around the defining elements of a behavioral assessment tool to be considered clinical, actuarial or black-box. We considered that the main difference originates from the source of data and the calculation method. Based on these two factor, we identify three broad categories of tools for behavioral assessment in the context of the criminal justice system: clinical, actuarial and adaptive-algorithmic. The risk assessment tool under analysis in this article can be consider as in the canonical approach that aims at predicting future behavior based on present individual traits via specialized judgment. Tools in this category, collect data primarily through human interaction and compute a overall measure following a pre-determined and finite algorithm previously agreed upon by professionals in the field. Actuarial tools of assessment obtained data for a non-primary source, that is not necessarily through human interaction and measurement, but by all available means of data collection and report and estimates an overall measure of risk following a probabilistic model. Finally, the more contentious category, to which we can attribute the renewed and more intense debate on the matter of fairness are adaptive-algorithmic tools, which rely on non-primary data and flexible statistical models that calculate an end measure of risk through a computer, tweaking the algorithm parameter to maximize performance in-sample and out of sample.

\pretitle\posttitle\preauthor\postauthor\predate\postdate

# Promise of Peril? Prediction of Risk

Among the tools for assessment in the criminal justice system, those related to sentencing or so called front-end decisions are the most contentious, as there is a qualitative dimension that distinguishes them from others and influences the placing of a person in or out of incarceration. Individual assessments of different types, such as risk, needs, success or failure relative to an outcome are useful tools in different stages of the criminal justice process, from pretrial detention (Milgram et al. 2015), early release, prosecution, sentencing, treatment, and post-release supervision for adults and for juveniles.

Judges make those decisions under human constraints, and as(Starr 2016) argues they balance competing sentencing criteria and make difficult decisions all the time. Starr adds that in many cases, pre-sentence reports based on risk assessment considerations make the decision more complex rather than helping the process; although she does not explain her reasoning, other than emphasizing that such reports are an additional variable for the judges to ponder in an already complex decision making framework. However, it can be argued that it is also the case that the individual judgment of criminal justice officials such as those in charge of sentencing, can obey rules that are not visible as opposed to the potentially more transparent information derived from an assessment tool.

To summarize the debate around the role of risk assessment tools in sentencing, a brief conceptual framework is needed. A general definition of risk assessment is the process of using risk factors to estimate and manage 11—(Webster et al. 2006)— add ”manage” to the original definition by —(Kraemer et al. 1997)— as a relevant element to indicate the purpose of a tool of this nature.the probability of an outcome occurring in a population. The concepts of risk and risk factor might vary across instruments, but their explicit definitions are necessary to evaluate the degree to which they are useful. For example, risk factors can be categorized into fixed or variable, and the idea of causality if introduced, must have corresponding empirical evidence, as opposed to speculations or beliefs (Kraemer et al. 1997).

According to this definition, at any step of the criminal process or function of a justice sector institution an assessment of expected outcomes can be conducted based on a set of predictors, by starting a process of inputting data based on a previously articulated rule of prediction. Leaving aside for a moment the tool (mathematical or clinical) to carry out the process, we have a general structure that requires an input to generate an expected outcome. This generic definition of a risk assessment tool acknowledges the relevance of using front-end decisions as an example to illustrate the point that the black box has different shadowy angles to uncover.

Beyond this generally accepted definition of risk assessment, the discussion on actuarial techniques to predict criminal behavior has been characterized by an array of terms and labels that rather than facilitate communication make it more difficult. Risk has been confounded with need, and no meaningful distinctions between the different usages have been made. So the debate presents a variety of adjectives attached to the term risk, such as static and dynamic, preventive, actuarial, predictive, and others. (Monahan and Skeem 2015) contribute to the discussion of assessment of risk in sentencing by providing a framework to select an actuarial assessment instrument that satisfies the claims of both sides of the debate, namely those who care about the input and those who care about the output. According to the authors, risk assessment tools can be ordered along three dimensions: purpose, degree of structure and quality of validation.

In building this framework, the authors provide basic definitions and a typology of relevant terms, a statement of two main goals of assessment tools, an overview of the role of such tools in the sentencing process, and a discussion of the varying extents of validation of the actuarial tools. Finally, and perhaps most importantly, they conclude by providing two guiding criteria to choose an assessment tool in sentencing: purpose of the evaluation and a principle of fairness, defined as minimal bias and lowest mean score across groups. Mathematically, this fairness principle can be expressed as minimizing the mean squared error of prediction associated with the highest prediction accuracy. In doing so, (Monahan and Skeem 2015) put into words the mathematical trade-off that exist between bias and variance which in terms of risk scores would result in having similar predictive accuracy across different groups of the population and minimal variance. It is Worth noting that even if this statement poses a neat optimization problem with a single solution, that is achievable through computation, the complex part of partitioning the population into groups and operationalizing risk factors is a matter of policy design.

\citeauthor{skeem2015risk} define a risk factor as a variable that precedes and is associated with increased likelihood of criminal behavior. Risk factors can be categorized as one of the following: i) fixed marker, ii) variable marker, iii) variable risk factor or iv) causal risk factor, which vary in their propensity to change over time and as a result of intervention. By further explaining that a risk factor can be appropriately named a proxy factor for criminal behavior, they maintain that considering risk a proxy -such as criminal history- for intrinsic features of individuals such as race, gender, or socioeconomic status, implies a causal link that is not stated, explained and operationalized. According to this view, critiques targeting risk factors as a veiled instrument for discrimination miss the important point of analyzing the concept and purpose of the term ”risk factor”. Therefore, in challenging the outcome, critics overlook the necessary step of analyzing the purpose, operationalization, functioning, and implementation of an algorithm.

In addition to the imprecise terminology used by scholars discussing the issue, there is a lack of clarity regarding the ways in which actuarial methods differ in purpose. Arguably the most important one is the distinction between predicting and managing risk. The former seeks to exclusively describe potential future outcomes, while the latter is intended to inform supervision and treatment of individuals so that such predicted risk is managed and reduced (Gerlinger and Turner 2015; Turner 2015; Lin, Grattet, and Petersilia 2010).

The policy implications of this lack of rigor in the debate are enormous when one considers the extended use of risk assessment tools across the country. State and federal institutions have in place major investments in ambitious initiatives to build up a data-driven approach in criminal justice as well as in other areas of government. The Data Driven Justice Initiative launched by the White House in 2015 aimed at curbing incarceration trends by making the criminal justice system smarter. It brings together state and city-level justice institutions, with private and non-profit organizations and has support from both major political parties to advance the use of data to design effective criminal justice policies. Additionally, during the last few decades, at least 27 subnational units have introduced evidence-based policies (Lawrence 2013) including risk assessment tools for sentencing.

The most salient critiques of the trend of using statistical learning tools can be summarized as focused on the outcome rather than in the algorithm itself, or as former US General Attorney (Holder 2015) has expressed, the concerns concentrate on the effects of such actuarial mechanisms of prediction, which might have a disparate effect on racial minorities and underprivileged communities.

Although recent events widely covered by media, specifically the recent evaluation of the Chicago experiment described before and the ProPublica assessment of the COMPAS algorithm, have brought fresh air to the academic debate on discriminatory algorithms, those are only two out of many cases where the claim of algorithmic impartiality and utility has been challenged. As (Ridgeway 2013) points out prediction has played a major role in criminology over decades, before computational capabilities made possible the use of complex algorithms to forecast criminal justice outcomes and justice institutions were gathering huge amounts of data. In his opinion, the task of prediction, with or without statistical learning, has had recurrent pitfalls that need to be considered in order to get the most out of the latest scientific developments in prediction. Ridgeway specifies seven common pitfalls of prediction made by criminal justice institutions and officials and consider them as a case to use computational algorithms to overcome such challenges as they have proven to outperform human judgment.

Ridgeway’s seven practical problems of applying machine learning methods to criminal justice are i) trusting expert opinion too much, ii) clinging to basic statistic concepts but failing to understand prediction models characteristics (such as performance criteria, accuracy, computational efficiency and the trade-off between parsimony and interpretability), iii) assuming that one method works best for all problems, iv) trying to interpret too much when the algorithm is not transparent (example, tree method or components in PCA), v) forsaking model simplicity for predictive strength (or viceversa), vi) expecting perfect predictions instead of efficiency improvements, and vii) failing to consider the unintended consequences of prediction (Berk et al. 2009).

In the case of statistical assessment of risk, (Monahan and Skeem 2015) conclude that four steps are necessary to select a risk assessment instrument: first, a definition of its purpose, understood as it policy objective or ultimate expected outcome of its implementation. Second, the degree of structure in terms of the institution and public official who will use and apply the instrument. Third, make sure that the instrument is validated and, finally, that it has an operational definition of fairness.

It seems that the example of front-end assessment tools based on prediction sheds light into the several angles of the black box: it is not only a matter of algorithmic transparency but an issue of design and implementation of the instrument that is out of the scope of a purely mathematical model. Failing to acknowledge the different components of the problem does not help a fruitful discussion, as the parties are talking different languages. In a sense, it is a logical trap, as the critiques are auto-referential: the root of racial discrimination is racial discrimination. Evidence suggests that as with any scientific tool, statistical learning algorithms of all sorts are to be used responsibly and in a deliberate and transparent fashion, pointing out at each stage - from design, validation and implementation- the objectives and mathematical limitations of any given algorithm. But, in my opinion, it would be a huge pitfall not to incorporate statistical learning tools into the craft of criminal justice policy - or more generally into the design of public policy-, as they offer a good shot at operationalizing a concept of fairness as the perfect balance between bias and variance, the core trade-off of machine learning tools. Furthermore, the discipline of machine learning is a viable way of taking a stab at optimization problems with multiple solutions, such as those modeled by neural networks, deep learning and virtual reality algorithms(Woodie 2015; Fox, Arena, and Bailenson 2009; Byrne and Marx 2011; Wiggins 2006; Bailenson et al. 2008; Bailenson et al. 2008).

\pretitle\posttitle\preauthor\postauthor\predate\postdate

# Algorithm versus Judge

I have reviewed several empirical examples of successful and unsuccessful applications of statistical learning algorithms to criminal justice policy. The results of the models considered are associated to gains, such as a potential cut of 2,000 post-arraignment arrests for domestic violence(Berk, Sorenson, and Barnes 2016) and losses, such as the use of the Strategic Subject List by police officers to disproportionally target individuals classified as of high risk (Saunders, Hunt, and Hollywood 2016). This spectrum sets the horizon for future applications of machine learning to criminal justice policy questions.

The case of statistical assessments of risk illustrates the complexity of designing a policy instrument to support judges’ decision making process by showing the necessity of algorithmic and policy transparency as it will have an effect over the physical freedom of a fellow individual. The contentious case of modeling an imperfect world at the risk of reproducing its disparities is a hazard of the application of machine learning to criminal justice. However, this hazard is not different from those posed by other methods already used by justice officials to decide, from individual heuristics to clinical assessments, judges already balance competing criteria when deciding over individual sentences and make decision under conditions that vary from case to case and from judge to judge (Starr 2016; Laqueur and Copus 2015). A question to consider is if machine learning or other mathematical approaches to policy might outperform humans and increase the overall efficiency of the system providing a basic framework to support judges -and other public officials- work.

There are several examples suggesting that analyzing the costs and benefits of different policy alternatives, including that of incorporating machine learning algorithms, is safer way to craft policies that yield the highest overall benefits to society, as it is suggested in the economics of crime literature(Domínguez and Raphael 2015). Such a framework would allow policy makers to ponder the effect that a particular policy might have in the overall system and, perhaps take action. One example is predictive policing, which has been shown to be effective in forecasting crime hotspots, leading police departments to patrol with higher intensity those areas, which in the process has alerted potential offenders and moved them to other locations, leading researchers to reevaluate the predictive model and make it dynamic, so it can account for potential changes in the overall distribution of crime patterns and police resources deployment (Ferguson 2016; Kitchin 2015)

\pretitle\posttitle\preauthor\postauthor\predate\postdate

# Conclusion and Next Steps

Criminal justice institutions are in constant search of policy alternatives to increase efficiency in maximizing public safety while curbing incarceration rates. Data-driven policy is the latest trend in achieving this goal by shaping decision making processes in different stages of the criminal justice process. Two technological developments from the last couple of decades have placed machine learning methods as the cutting-edge tool to explore critical problems of criminology and shape policy. I have provided an overview of machine learning methods used to visualize and analyze criminal justice data. Clustering and principal component analysis; linear methods for regression,including penalized regression, variable and model selection; tree methods and, bootstrapping, boosting, bagging are approaches to unveil relationships among variables and predict events of interest for the institutions in the criminal justice system.

I have provided a summary of the critiques and empirical examples supporting and challenging the scope and reach of such computational statistical methods. My review has shown that there are two main focus of critique to machine learning methods. First, the claims of mathematical complexity and opacity which has gained them the name of black box methods. Second, complexity and opacity in policy design and implementation. I have proposed that contrary to a trending belief reproduced by scholars and media, black box methods, as applied to criminal justice, have a policy dimension that needs to be addressed in order to tackle the issue of mathematical opacity. I claim this dimension of policy making is more challenging and dangerous that the issue presented by mathematical complexity alone.

Based on the analysis of front-end statistical risk assessment, fours steps are suggested to ensure that the selection of an computational statistical tool for policy has gone through ethical and methodological considerations: objective, design, implementation plan and validation. Objective refers to a definition of policy purpose. Design relates to identifying, defining and justifying relevant variables associated with the goal of the instrument. Implementation plan refers to identifying the institution, public official and stage in the process at which the instrument will be applied, as well as the role of the tool in the decision process. Finally, empirical validation of the instrument will make it time and location specific to serve its purpose. These steps will expand the extent to which machine learning can help to maximize the objective function of the criminal justice system and provide public safety at the same time that curbs the incarceration trend.

As for the existing policies that rely on machine learning methods, an approach to assess their efficacy is by conducting empirical analyses comparing geographical units -such states or counties- where a given policy has been implemented with one where it has not. One potential case of study is at the early stages in the criminal justice process, pre-trial detention, when a judge assess if the defendant has a flight risk or more likely to appear in court during the process. An analysis of rates of failure to appear by group in both scenarios, one where the judge decision was supported by a statistical tool and one where the judge decided with clinical assessments would provide additional information about the effect of introducing a machine learning methods to approach criminal justice policy.

# References

Almeida, Hayda, Marie-Jean Meurs, Leila Kosseim, Greg Butler, and Adrian Tsang. 2014. “Machine Learning for Biomedical Literature Triage”. *PloS One* 9 (12). Public Library of Science: e115892.

Ozer, Sedat, Deanna L Langer, Xin Liu, Masoom A Haider, Theodorus H van der Kwast, Andrew J Evans, Yongyi Yang, Miles N Wernick, and Imam S Yetik. 2010. “Supervised and Unsupervised Methods for Prostate Cancer Segmentation with Multispectral MRI”. *Medical Physics* 37 (4). American Association of Physicists in Medicine: 1873–83.

Kourou, Konstantina, Themis P Exarchos, Konstantinos P Exarchos, Michalis V Karamouzis, and Dimitrios I Fotiadis. 2015. “Machine Learning Applications in Cancer Prognosis and Prediction”. *Computational and Structural Biotechnology Journal* 13. Elsevier: 8–17.

Potash, Eric, Joe Brew, Alexander Loewi, Subhabrata Majumdar, Andrew Reece, Joe Walsh, Eric Rozier, Emile Jorgenson, Raed Mansour, and Rayid Ghani. 2015. “Predictive Modeling for Public Health: Preventing Childhood Lead Poisoning”. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2039–47. ACM.

Petersen, Maya L, Sandra E Sinisi, and Mark J van der Laan. 2006. “Estimation of Direct Causal Effects”. *Epidemiology* 17 (3). LWW: 276–84.

Van der Laan, Mark J, and Sherri Rose. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data*. Springer Science & Business Media.

Goel, Sharad, Justin M Rao, and Ravi Shroff. 2016. “Precinct or Prejudice? Understanding Racial Disparities in New York City’s Stop-and-Frisk Policy”. *Annals of Applied Statistics* 10 (1): 365–94.

Berk, Richard, and Justin Bleich. 2014. “Forecasts of Violence to Inform Sentencing Decisions”. *Journal of Quantitative Criminology* 30 (1). Springer: 79–96.

Berk, Richard, Lawrence Sherman, Geoffrey Barnes, Ellen Kurtz, and Lindsay Ahlman. 2009. “Forecasting Murder within a Population of Probationers and Parolees: a High Stakes Application of Statistical Learning”. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (1). Wiley Online Library: 191–211.

Monahan, John, and Jennifer L Skeem. 2015. “Risk Assessment in Criminal Sentencing”. *Annual Review of Clinical Psychology, Forthcoming*.

Holder, Eric. 2015. “Speech Presented at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference, Philadelphia, PA”. *Federal Sentencing Reporter* 27 (4). University of California Press Journals: 252–55.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics Springer, Berlin.

Varian, Hal R. 2014. “Big Data: New Tricks for Econometrics”. *The Journal of Economic Perspectives* 28 (2). American Economic Association: 3–27.

Eck, John E, and David L Weisburd. 2015. “Crime Places in Crime Theory”. *Crime and Place: Crime Prevention Studies* 4.

Singh, Ajay K. 2006. “Hierarchical Clustering Techniques in Crime GIS”.

Grubesic, Tony H. 2006. “On the Application of Fuzzy Clustering for Crime Hot Spot Detection”. *Journal of Quantitative Criminology* 22 (1). Springer: 77–105.

Cooper, Hannah LF, Sabriya Linton, Mary E Kelley, Zev Ross, Mary E Wolfe, Yen-Tyng Chen, Maria Zlotorzynska, et al. 2016. “Risk Environments, Race/Ethnicity, and HIV Status in a Large Sample of People Who Inject Drugs in the United States”. *PloS One* 11 (3). Public Library of Science: e0150410.

Ayoola, J, MA Adeyemi, SO Jabaru, and others. 2015. “On the Estimation of Crime Rate in the Southwest of Nigeria: Principal Component Analysis Approach”. *Global Journal of Science Frontier Research* 15 (2).

Berk, Richard. 2010. “What You Can and Can’t Properly Do with Regression”. *Journal of Quantitative Criminology* 26 (4). Springer: 481–87.

Sculley, D, and others. 2011. “Results from a Semi-Supervised Feature Learning Competition”. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*.

Hoerl, Arthur E, and Robert W Kennard. 1970. “Ridge Regression: Biased Estimation for Non-Orthogonal Problems”. *Technometrics* 12 (1). Taylor & Francis Group: 55–67.

Munnell, Alicia H, Geoffrey MB Tootell, Lynn E Browne, and James McEneaney. 1996. “Mortgage Lending in Boston: Interpreting HMDA Data”. *The American Economic Review*. JSTOR, 25–53.

Freund, Yoav, Robert Schapire, and N Abe. 1999. “A Short Introduction to Boosting”. *Journal-Japanese Society For Artificial Intelligence* 14 (771-780). JAPANESE SOC ARTIFICIAL INTELL: 1612.

Breiman, Leo. 1996. “Bagging Predictors”. *Machine Learning* 24 (2). Springer: 123–40.

Harcourt, Bernard E. 2003. “The Shaping of Chance: Actuarial Models and Criminal Profiling at the Turn of the Twenty-First Century”. *The University of Chicago Law Review* 70 (1). JSTOR: 105–28.

Starr, Sonja. 2016. “The Odds of Justice: Actuarial Risk Prediction and the Criminal Justice System: Mary W. Gray, Column Editor”. *CHANCE* 29 (1). Taylor & Francis: 49–51.

Harcourt, Bernard E. 2008. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. University of Chicago Press.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. “Machine Bias: There’s Software Used across the Country to Predict Future Criminals. And It’s Biased against Blacks”. *ProPublica, May* 23.

Petersilia, Joan. 2014. “Joan Petersilia Presents Sentencing, Risk Assessment, and Re-Offending. Stanford, March, 2013”. <\url{https://www.youtube.com/watch?v=YjPm-gedWo8}.>

Raphael, Steven, and Michael A Stoll. 2013. *Why Are so Many Americans in Prison?*. Russell Sage Foundation.

Athey, Susan, and Guido W Imbens. 2015. “Machine Learning Methods for Estimating Heterogeneous Causal Effects”. *Stat* 1050: 5.

Saunders, Jessica, Priscillia Hunt, and John S. Hollywood. 2016. “Predictions Put into Practice: a Quasi-Experimental Evaluation of Chicago’s Predictive Policing Pilot”. *Journal of Experimental Criminology*, 1–25. doi:10.1007/s11292-016-9272-0.

Perry, Walt L. 2013. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Rand Corporation.

Ridgeway, Greg. 2003. “Strategies and Methods for Prediction”. *NIJ Journal*. Citeseer.

Milgram, Anne, Alexander M Holsinger, Marie Vannostrand, and Matthew W Alsdorf. 2015. “Pretrial Risk Assessment”. *Federal Sentencing Reporter* 27 (4). University of California Press Journals: 216–21.

Webster, Christopher D, Tonia L Nicholls, Mary-Lou Martin, Sarah L Desmarais, and Johann Brink. 2006. “Short-Term Assessment of Risk and Treatability (START): the Case for a New Structured Professional Judgment Scheme”. *Behavioral Sciences & the Law* 24 (6). Wiley Online Library: 747–66.

Kraemer, Helena Chmura, Alan E Kazdin, David R Offord, Ronald C Kessler, Peter S Jensen, and David J Kupfer. 1997. “Coming to Terms with the Terms of Risk”. *Archives of General Psychiatry* 54 (4). American Medical Association: 337–43.

Gerlinger, Julie, and Susan F Turner. 2015. “California’s Public Safety Realignment Correctional Policy Based on Stakes Rather than Risk”. *Criminal Justice Policy Review* 26 (8). SAGE Publications: 805–27.

Turner, Susan. 2015. “Predicting Risk: Who Knew It Was Such a Risky Business?”. In *Envisioning Criminology*, 205–10. Springer.

Lin, Jeffrey, Ryken Grattet, and Joan Petersilia. 2010. “‘BACK-END SENTENCING’ AND REIMPRISONMENT: INDIVIDUAL, ORGANIZATIONAL, AND COMMUNITY PREDICTORS OF PAROLE SANCTIONING DECISIONS”. *Criminology* 48 (3). Wiley Online Library: 759–95.

Lawrence, Alison. 2013. “Trends in Sentencing and Corrections: State Legislation”. In *National Conference of State Legislatures*.

Ridgeway, Greg. 2013. “The Pitfalls of Prediction”. *NIJ Journal* 271: 34–40.

Woodie, Alex. 2015. “What Police Can Learn from Deep Learning”. *Datanami*.

Fox, Jesse, Dylan Arena, and Jeremy N Bailenson. 2009. “Virtual Reality: A Survival Guide for the Social Scientist”. *Journal of Media Psychology* 21 (3). Hogrefe Publishing: 95–113.

Byrne, James, and Gary Marx. 2011. “Technological Innovations in Crime Prevention and Policing. A Review of the Research on Implementation and Impact”. *Journal of Police Studies* 20 (3): 17–40.

Wiggins, Elizabeth C. 2006. “The Courtroom of the Future Is Here: Introduction to Emerging Technologies in the Legal System”. *Law & Policy* 28 (2). Wiley Online Library: 182–91.

Bailenson, Jeremy N, Alexandra Davies, Jim Blascovich, Andrew C Beall, Cade McCall, and Rosanna E Guadagno. 2008. “The Effects of Witness Viewpoint Distance, Angle, and Choice on Eyewitness Accuracy in Police Lineups Conducted in Immersive Virtual Environments”. *Presence: Teleoperators and Virtual Environments* 17 (3). MIT Press: 242–55.

———. 2008. “The Effects of Witness Viewpoint Distance, Angle, and Choice on Eyewitness Accuracy in Police Lineups Conducted in Immersive Virtual Environments”. *Presence: Teleoperators and Virtual Environments* 17 (3). MIT Press: 242–55.

Berk, Richard A, Susan B Sorenson, and Geoffrey Barnes. 2016. “Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions”. *Journal of Empirical Legal Studies* 13 (1). Wiley Online Library: 94–115.

Laqueur, Hannah, and Ryan Copus. 2015. “Machines Learning Justice: The Case for Judgmental Bootstrapping of Legal Decisions”. *Available at SSRN*.

Domínguez, Patricio, and Steven Raphael. 2015. “The Role of the Cost-of-Crime Literature in Bridging the Gap Between Social Science Research and Policy Making”. *Criminology & Public Policy* 14 (4). Wiley Online Library: 589–632.

Ferguson, Andrew Guthrie. 2016. “Policing Predictive Policing”. *Washington University Law Review* 94.

Kitchin, Rob. 2015. “The Promise and Peril of Smart Cities”. *Computers and Law: the Journal of the Society for Computers and Law* 26 (2). Society for Computers and Law.

Neugebauer, Romain, Julie A Schmittdiel, and Mark J Laan. 2014. “Targeted Learning in Real-World Comparative Effectiveness Research with Time-Varying Interventions”. *Statistics in Medicine* 33 (14). Wiley Online Library: 2480–2520.

O’Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books.

Tett, Gillian. 2014. “Mapping Crime – or Stirring Hate?”. Edited by Financial Times.

Athey, Susan, and Guido Imbens. 2016. “The State of Applied Econometrics-Causality and Policy Evaluation”. *ArXiv Preprint ArXiv:1607.00699*.

Wu, Xindong, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, et al. 2008. “Top 10 Algorithms in Data Mining”. *Knowledge and Information Systems* 14 (1). Springer: 1–37.

Breiman, Leo. 2003. “Statistical Modeling: The Two Cultures”. *Quality Control and Applied Statistics* 48 (1). Executive Sciences Institute: 81–82.

Raphael, Steven. 2014. “How Do We Reduce Incarceration Rates While Maintaining Public Safety?”. *Criminology & Public Policy* 13 (4): 579–97.

Simoiu, Camelia, Sam Corbett-Davies, and Sharad Goel. n.d. “Testing for Racial Discrimination in Police Searches of Motor Vehicles⇤”.

Goel, Sharad, Justin M Rao, and Ravi Shroff. 2016. “Personalized Risk Assessments in the Criminal Justice System”. *The American Economic Review* 106 (5). American Economic Association: 119–23.

Yee, Nick, and Jeremy N Bailenson. 2006. “Walk a Mile in Digital Shoes: The Impact of Embodied Perspective-Taking on the Reduction of Negative Stereotyping in Immersive Virtual Environments”. *Proceedings of PRESENCE* 246: 147–56.

Groom, Victoria, Jeremy N Bailenson, and Clifford Nass. 2009. “The Influence of Racial Embodiment on Racial Bias in Immersive Virtual Environments”. *Social Influence* 4 (3). Taylor & Francis: 231–48.

Young, Caitlin O. 2014. “Employing Virtual Reality Technology at Trial: New Issues Posed by Rapid Technological Advances and Their Effects on Jurors’ Search for the Truth”. *Tex. L. Rev.* 93. HeinOnline: 257.

Segovia, Kathryn Y, Jeremy N Bailenson, and Carrie Leonetti. 2012. “Virtual Human Identification Line-Ups”. *Craniofacial Identification*. Cambridge University Press, 101.

Leonetti, Carrie, and Jeremy Bailenson. 2009. “High-Tech View: the Use of Immersive Virtual Environments in Jury Trials”. *Marq. L. Rev.* 93. HeinOnline: 1073.

Bailenson, Jeremy N, Jim Blascovich, Andrew C Beall, and Beth Noveck. 2006. “Courtroom Applications of Virtual Environments, Immersive Virtual Environments, and Collaborative Virtual Environments”. *Law & Policy* 28 (2). Wiley Online Library: 249–70.

Frase, Richard S. 2014. “Recurring Policy Issues of Guidelines (and Non-Guidelines) Sentencing: Risk Assessments, Criminal History Enhancements, and the Enforcement of Release Conditions”. *Federal Sentencing Reporter* 26 (3). University of California Press: 145–57.

Monahan, John, and Jennifer L Skeem. 2013. “Risk Redux: The Resurgence of Risk Assessment in Criminal Sanctioning”. *Virginia Public Law and Legal Theory Research Paper*, no. 2013-36.

Petersilia, Joan. 2003. *When Prisoners Come Home: Parole and Prisoner Reentry*. Oxford University Press.

Byrne, James, and Gary Marx. 2011. “Technological Innovations in Crime Prevention and Policing. A Review of the Research on Implementation and Impact”. *Journal of Police Studies* 20 (3): 17–40.

Turner, Susan, and Joan Petersilia. 1992. “Focusing on High-Risk Parolees: An Experiment to Reduce Commitments to the Texas Department of Corrections”. *Journal of Research in Crime and Delinquency* 29 (1). Sage Publications: 34–61.

Berendt, Bettina, and Sören Preibusch. 2014. “Better Decision Support through Exploratory Discrimination-Aware Data Mining: Foundations and Empirical Evidence”. *Artificial Intelligence and Law* 22 (2). Springer: 175–209.

Fix, Evelyn, and Joseph L Hodges Jr. 1951. “Discriminatory Analysis-Nonparametric Discrimination: Consistency Properties”. DTIC Document.

Berk, Richard A. 2008. *Statistical Learning from a Regression Perspective*. Springer Science & Business Media.

Harcourt, Bernard E. 2004. “Rethinking Racial Profiling: A Critique of the Economics, Civil Liberties, and Constitutional Literature, and of Criminal Profiling More Generally”. *The University of Chicago Law Review* 71 (4). JSTOR: 1275–1381.

———. 2010. “Risk as a Proxy for Race”. *Criminology and Public Policy, Forthcoming*.

Simon, J, and M Feeley. 1994. “Actuarial Justice: The Emerging New Criminal Law”. *The Futures of Criminology. Thousand Oaks, CA: Sage*.

Legewie, Joscha. 2016. “Racial Profiling and Use of Force in Police Stops: How Local Events Trigger Periods of Increased Discrimination”. *NYU: Working Chapter*.

Goel, Sharad, Maya Perelman, Ravi Shroff, and David Alan Sklansky. 2016. “Combatting Police Discrimination in the Age of Big Data”. *Forthcoming, New Criminal Law Review*.

Persico, Nicola. 2009. “Racial Profiling? Detecting Bias Using Statistical Evidence”. *Annu. Rev. Econ.* 1 (1). Annual Reviews: 229–54.

Neuilly, Melanie-Angela, Kristen M Zgoba, George E Tita, and Stephen S Lee. 2011. “Predicting Recidivism in Homicide Offenders Using Classification Tree Analysis”. *Homicide Studies* 15 (2). Sage Publications: 154–76.

Ridgeway, Greg. 2013. “Linking Prediction and Prevention”. *Criminology & Public Policy* 12 (3). Wiley Online Library: 545–50.

Tita, George, and Greg Ridgeway. 2007. “The Impact of Gang Formation on Local Patterns of Crime”. *Journal of Research in Crime and Delinquency* 44 (2). Sage Publications: 208–37.

Papachristos, Andrew V, Anthony A Braga, and David M Hureau. 2012. “Social Networks and the Risk of Gunshot Injury”. *Journal of Urban Health* 89 (6). Springer: 992–1003.

Papachristos, AV, AA Braga, and DM Hureau. 2011. “Six-Degrees of Violent Victimization: Social Networks and the Risk of Gunshot Injury”. *Retrieved February* 26: 2011.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer. 2015. “Prediction Policy Problems”. *The American Economic Review* 105 (5). American Economic Association: 491–95.

Ohlin, Lloyd E, and Otis Dudley Duncan. 1949. “The Efficiency of Prediction in Criminology”. *American Journal of Sociology*. JSTOR, 441–52.

Farrington, David P, and Roger Tarling. 1985. *Prediction in Criminology*. SUNY Press.

Casey, Pamela M, Roger K Warren, and Jennifer K Elek. 2011. *Using Offender Risk and Needs Assessment Information at Sentencing: Guidance for Courts from a National Working Group*. National Center for State Courts.

Artan, Yusuf, Masoom A Haider, Deanna L Langer, Theodorus H van der Kwast, Andrew J Evans, Yongyi Yang, Miles N Wernick, John Trachtenberg, and Imam Samil Yetik. 2010. “Prostate Cancer Localization with Multispectral MRI Using Cost-Sensitive Support Vector Machines and Conditional Random Fields”. *IEEE Transactions on Image Processing* 19 (9). IEEE: 2444–55.

Ohlin, Lloyd E, and Richard A Lawrence. 1952. “A Comparison of Alternative Methods of Parole Prediction”. *American Sociological Review* 17 (3). JSTOR: 268–74.

Wilkins, Leslie T, and Peter MacNaughton-Smith. 1964. “New Prediction and Classification Methods in Criminology”. *Journal of Research in Crime and Delinquency* 1 (1). Sage Publications: 19–32.

Houghton, John Theodore, YDJG Ding, David J Griggs, Maria Noguer, Paul J van der Linden, Xiaosu Dai, Kathy Maskell, and Cathy A Johnson. 2001. *Climate Change 2001: the Scientific Basis*. The Press Syndicate of the University of Cambridge.