# Crime in America: An Application of Machine Learning in Criminology

Giang Nguyen[1], Sophie Lee, Allison Conners, Asaph Young Chun
International Strategy and Reconciliation Foundation Center for Interdisciplinary Research

## 1. Introduction

The United States has a higher percentage of its population incarcerated than any other country in the world. While there are many different angles from which this issue can be studied, the aim of this paper is to explore the societal factors which contribute to one's likelihood of committing a crime and the consequences that this criminal record will have on one's life. We intend to establish the correlates of recidivism – the tendency of a criminal to commit another crime – using prediction, and regression models.

We do this not in a vacuum, but with the knowledge that the answers to these questions are sought by actors in the criminal justice system. Currently, ten states use statistical models formally in the sentencing of convicted criminals. Many of these models are "black boxes". That is, the variables considered in sentencing, (which may include race, gender, creed) are not publicly known. The goal of this practice is to incapacitate individuals at the highest risk of committing crime in the future.

Because these models are used in the process of delivering criminals into the custody of the criminal justice system, they should be considered inside the context of the criminal justice system. The three purposes of any justice system are to prevent crime, to punish crime, and to rehabilitate criminals.

Indeed, these models are designed to prevent crime, so the first point is met. According to the American Psychological Association, the justice system has become more punitive over the last several decades. It is important to note that these models estimate the likelihood of a future crime being committed based on observables (race, gender, etc.). Therefore, when these models lead to an increase in a sentence, people are being punished not for actual crimes, but for these observable characteristics. Finally, these models emphasize prevention through "taking them off the streets", not prevention through rehabilitation. We turn the question on its head. If the criminal justice system does indeed rehabilitate, statistical models could be used to find the optimal sentence that leads to less recidivism rather than merely measuring the probability of recidivism.

To explore these issues, the paper first reviews the literature on the use of machine learning in the field of criminology, then the research questions are fully stated and elaborated on. Section 4 contains a description of the data, and section 5 contains a description of the methodology used. Finally, we present our findings in section 6.

## 2. Literature Review

### 2.1 Predicting recidivism and consequences of a criminal record for individuals

The purpose of the paper is to demonstrate how results from data visualizations and machine learning techniques may help the U.S. government, the United Nations and non-

[1] Giang Nguyen, Junior Fellow for Da Vinci Science Diplomacy and Senior Research Coordinator, International Strategy and Reconciliation Foundation Center for Science Diplomacy, www.isr2020.org You may contact giang.huong.nguyen92@gmail.com or ISRFoundation38@gmail.com if you have questions and inquiries about this paper.

governmental agencies develop data-driven decisions for humanitarian and development programs that aim to reduce crime. In order to do so, we review papers that provide essential information on crime and recidivism in the United States. The papers address our research questions regarding the correlates and predictor variables of recidivism and the consequences of criminal records for individuals.

Multiple studies support that the number of prior arrests is positively correlated to recidivism, meaning that the higher the number of prior arrests, the more likely it is for an individual to reoffend. In terms of demographics, we found that offenders who are male, of a minority race, and from a low socioeconomic background are more likely to reoffend than those who do not have these characteristics. Langan and Levin (2002) claims that there was no evidence that spending more time in prison increases recidivism rate and that those who served the longest time in fact had a significantly low rearrest rate. Cottle, Lee and Heilbrun (2001) state that longer incarcerations are positively correlated with recidivism. While the two pieces of information seem contradictory, we believe this may be due to the fact that the latter study only surveyed those aged 12 to 21, whereas the former tracked inmates of all ages.

Another topic the papers discuss in common is the role of dynamic risk factors in predicting criminal recidivism. Following Cottle, Lee and Heilbrun (2001), dynamic risk factors are factors that have the potential to change through planned intervention. Examples of dynamic factors include substance abuse and negative peer associations. Public Safety Canada (2010) defines static risk factors, on the other hand, are variables that predict recidivism but are not able to change under any intervention, such as prior offences. Gendreau, Little and Goggin (1996) emphasize the importance of including and reassessing these dynamic factors when measuring recidivism, stating that the dynamic predictor domains performed at least as well as static domains when predicting recidivism rates in their studies.

Another key concept the papers bring up is the “collateral consequence” of criminal conviction, Kurlychek, Brame and Bushway (2006). An example of these consequences are the difficulties the offenders face when trying to obtain employment after their release. This is ethically and morally concerning, as it amplifies the offenders’ punishment beyond sanctions imposed on them by the criminal justice system. The study is aimed to find out the relevance of criminal history records for predicting employment behavior and reached the conclusion that while people with prior arrest never become completely indistinguishable from those without prior offense, the risk of a new criminal event among a population of non-offenders and a population of prior offenders do become similar over time.

### 2.2 Review of machine learning and its applications in the social sciences

Machine learning is a rapidly growing technical discipline that is applicable to multiple fields, including science, technology, health care and education. It stands at the core of artificial intelligence and data science, and scientists and engineers turn to machine learning for solutions to problems that need useful insights, predictions and decisions from specific data sets Jordan and Mitchell (2015). The two literature we reviewed focus on two different topics: the tasks that machine learning systems can and cannot perform and its impact on the workforce, and its effect on measuring economic development.

In the first study, Brynjolfsson and Mitchell (2017) discuss the key implications of machine learning for the workforce, and what the current machine learning systems can and cannot do. As for the benefits of machine learning, the authors state that the automated process of running machine learning can produce more accurate and reliable programs and dramatically lower costs for creating and maintain new software. They also claim that machine learning can sometimes produce computer programs that outperform the best humans at the task and improve on human decisions.

However, the authors also point out that while machine learning may prove to be beneficial in some tasks, it is impossible for machines to perform the full range of tasks that humans can do. They explain that machine learning systems are not equally suitable for all tasks, and that their competence is narrower and more fragile than human decision making. It is also mentioned that while machine learning systems are strong at learning empirical associations, they are less effective in tasks that require complex reasoning or those that are related to interactions with people. Regardless of these shortcomings, the authors conclude by stating that since it is clear that machine learning will grow at a faster rate and become capable to perform a much broader set of tasks, we should pay attention to its implications and understand the precise applicability of each type of machine learning.

The second study, Jean et al. (2016), we reviewed gives a specific example of how a machine learning approach called *daytime satellite imagery* can be used to accurately measure the economic development of developing nations. The study claims that daytime satellite imagery is an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth. It also states that it is a far more efficient method than the other available methods, such as household surveys or nightlights, especially for identifying clusters below the international poverty line. This method involves transfer learning, which uses the convolutional neural network (CNN) model. The authors once again emphasize the strength of machine learning in measuring the economic development of nations by stating that the CNN model outperforms other approaches such as common general-purpose image features such as color histograms.

These studies demonstrate how machine learning can be applicable in various fields, and remind us of its strengths and weaknesses. This is helpful to our study, as it allows us to pay more attention to the results that we retrieve from machine learning methods and carefully consider its limitations when forming a conclusion.

The literature on comparing the effectiveness of machine learning and classical models in estimating recidivism is inconclusive. Berk and Bleich (2013) compares a baseline model (logistic regression) to modern machine learning techniques favorably. They find that machine learning techniques often yield models with lower standard errors than classical models. BB also gives a cohesive discussion as to why the accuracy of such models is important (as they are being used in sentencing decisions). The authors emphasize that they are agnostic towards which predictors should be used to estimate the probability of recidivism and interested in fitting a model. Our paper differs in two ways. First, we propose a Cox regression as better than a logistic regression. Second, we are more interested in understanding what correlations feature prominently in these models (using LASSO). This then lets us pose the question, 'Are we comfortable using this set of observables as determinants in sentencing decisions?'.

Tollenaar (2012) runs a similar horse race to BB (2013)- that is, comparing classical models to machine learning approaches; but reaches the opposite conclusion. Again, our paper contributes to this line by examining whether the Cox model (which would still be considered classical by Tollenaar's standard) performs better than Logistic regression, the method identified as superior to machine learning models such as neural networks by Tollenaar.

Looking to the literature of estimating recidivism more broadly, Rhodes (2013), in a response to Berk and Bleich note that there are four generations or schools of thought in estimating recidivism. Newer schools of thought prefer dynamic approaches to risk assessment (an individual's risk of recidivism may change over time, and in response to various treatments). While Berk and Bleich does not explicitly incorporate dynamic concerns, this is more likely to be due to limitations in the data rather than shortcomings in the approach. However, our baseline Cox regression specification improves upon BB's baseline model in including some dynamic variables.

## 3. Research Questions

This paper seeks to address three key questions. First, what influences recidivism? It is important to identify these variables to understand what lies behind modern machine learning models used by the justice system.

Second, we seek to consider various specifications for models of recidivism to address both statistician's and criminologist's theoretical concerns.

Finally, statistical models of recidivism are used to influence sentencing with the notion that someone more likely to recidivate should be taken off the streets. We argue that there is a second effect in play that the justice system should account for. Not only does imprisonment incapacitate criminals during their time served, but it also affects their behavior after release (possibly making crime more likely). Therefore, the sentence should be handed down not only with the short-term goal of incapacitating the criminal, but the long term goal of rehabilitating the criminal.

In pursuit of these goals, first, we use NLSY97 to study criminal behavior in young people. The nature of the data (many variables, few criminals), require parsimony in variable selection. We use model selection techniques to identify pertinent variables in recidivism. The silver lining to this need for parsimony is we identify a manageable list of relevant variables and can pose the normative question 'should sentencing be based on these observable variables'.

Second, our paper responds to the growing field of quantitative criminology. Berk and Bleich (2013), and Tollenaar (2012), run a horse race between the classical Logistic Regression, and more modern models such as neural networks to determine which models provide the best fit. By proposing the Cox Regression as a better classical model than logistic, we address theoretical concerns brought up in Rhodes (2013) (that previous work lacks dynamic or life-cycle elements which criminologists believe to be important).

## 4. Methodology and Data Analysis Plan

We begin by performing basic analysis on the Recidivism in the NSLY97 Standalone Data. We compare the sample probabilities of recidivism conditional on race, gender, income, and other observables. We create data visualizations to highlight notable correlations or differences across populations.

We define recidivism as occurring if an individual is arrested at least N>1 times conditional on the individual being arrested once. Because we have NSLY data on sampled individuals for twelve years of their life, censoring can occur. That is, observation of an individual may cease before the individual recidivates again.

In keeping with the structural criminology literature, we start our formal analysis by performing logistic regressions to characterize the odds of recidivism across individuals. Because our dataset is "wide" rather than "long", that is, there are many observable characteristics and comparatively few individuals with criminal histories in the sample, we turn to variable selection techniques provided by machine learning to build a model.

LASSO is used for variable selection, and important covariates are discussed. This methodology is very versatile, as it consists of applying an upper bound to the sum of the absolute value of a regression model's coefficients. Therefore, a wide number of regression models can be used.

Using LASSO provides two advantages. First, because the NLSY contains a fairly large number of variables compared to the number of people who have been incarcerated at least once, parsimony is necessary when selecting variables for a model. LASSO allows for variable selection or elimination in a sound and agnostic way. Second, as noted above in the introduction, one goal of this paper is to distill the model to relevant variables and frame the question 'should these variables be used to influence sentencing decisions'. LASSO provides a means of doing just that.

Proceeding forward, we use a Cox Proportional Hazards Model with LASSO, asserting that this is a more fitting "classical model" than the logistic model used by Berk and Bleich (2013). The Cox model corrects for censoring in the data, and incorporates a dynamic aspect to the model which the logistic model fails to display.

The Cox Proportional Hazards Model estimates the effect covariates have on the rate at which an event of interest (sometimes called a "death" or "failure") occurs. Critically, this model is best suited for dealing with problems of censoring. Censoring occurs when the event in question is not observed within the timeframe of the study, but the event could happen in the future as the subject lives on.

While this model has been used in criminology for at least 40 years, some recent papers have run horse races, comparing modern machine learning methods to the Logistic regression, which suffers problems with censoring.

Finally, clustering techniques are used to identify commonalities among individuals who recidivate and those who do not. Data visualizations are again created to illustrate the results of the regression analysis.

## 5. Description of the Data

We performed analysis on part 1 of the Recidivism in the National Longitudinal Survey of Youth 1997 Standalone Data. The NLSY97 is a nationally representative survey of nearly 12,000 individuals who were between the ages of 12 and 16 on December 31, 1996. The interviews were conducted from 1997 to 2009. The Recidivism Standalone dataset contains information on arrests and incarcerations, as well as self-reported crime activity. Part 1 of the dataset contains data on self-reported criminal activities for 8,984 individuals, whereas part 2 of the dataset contains data on self-reported illegal activities for 2,977 individuals.

Of the 8,984 individuals surveyed in part 1 of the dataset, 584 have been incarcerated for at least one month and 318 have been incarcerated for at least one month a year in more than one year during the surveyed timeframe (1997 to 2009). In other words, 6.5% of respondents were incarcerated at least one month, and 3.5% of respondents were incarcerated at least one month per year in more than one year. 32.83% of individuals sampled have been arrested.

In the table below, we provide statistics on race and gender, conditional on being incarcerated. In the table to follow, these same statistics are shown for the full sample.

| | **Black** | **Hispanic** | **Mixed race non-Hispanic** | **Non-black / non-Hispanic** | **Total** |
|---|---|---|---|---|---|
| **Male** | 32.71 | 17.81 | 0.17 | 31.85 | 82.54 |
| **Female** | 5.30 | 3.25 | 0.34 | 8.56 | 17.45 |
| **Total** | 38.01 | 21.06 | 0.51 | 40.41 | 100% |

Table 1: By race and gender, conditional on incarceration. (NLSY97)

Comparing to the demographics in the NLSY for the full sample:

| | **Black** | **Hispanic** | **Mixed race non-Hispanic** | **Non-black / non-Hispanic** | **Total** |
|---|---|---|---|---|---|
| **Male** | 13.01 | 10.87 | .45 | 26.86 | 51.19 |
| **Female** | 12.98 | 10.28 | .48 | 25.07 | 48.81 |
| **Total** | 25.99 | 21.15 | .93 | 51.96 | 100% |

Table 2: By race and gender, full sample (NLSY97)

We observe a disproportionate share of males, blacks, and Hispanic men are incarcerated. Black females are incarcerated only slightly proportionally more than their female counterparts, and Hispanic females are incarcerated proportionally less. This suggests an interaction between race and gender which should be explored, that is, black and Hispanic males rather than females drive the difference in incarceration rates between the races. Mixed race people represent less than one percent of the full sample, and Blacks and Hispanics are over sampled compared to the US population. A clear depiction of the effects of gender and race is plotted in the data visualizations below.

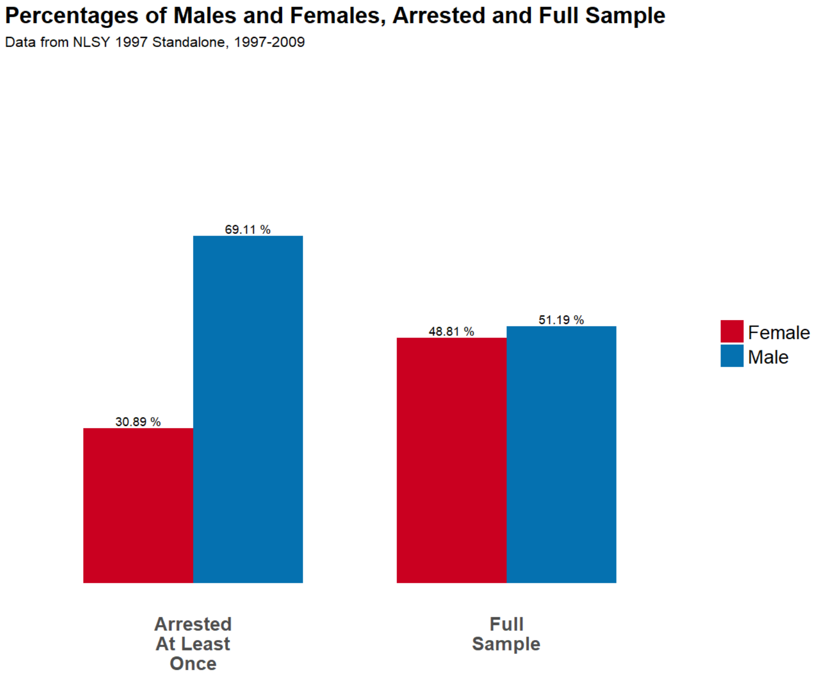

Figure 1: Full Sample by Arrest Occurrence (NLSY97)

In the above, it is clear that males are incarcerated proportionally more often.

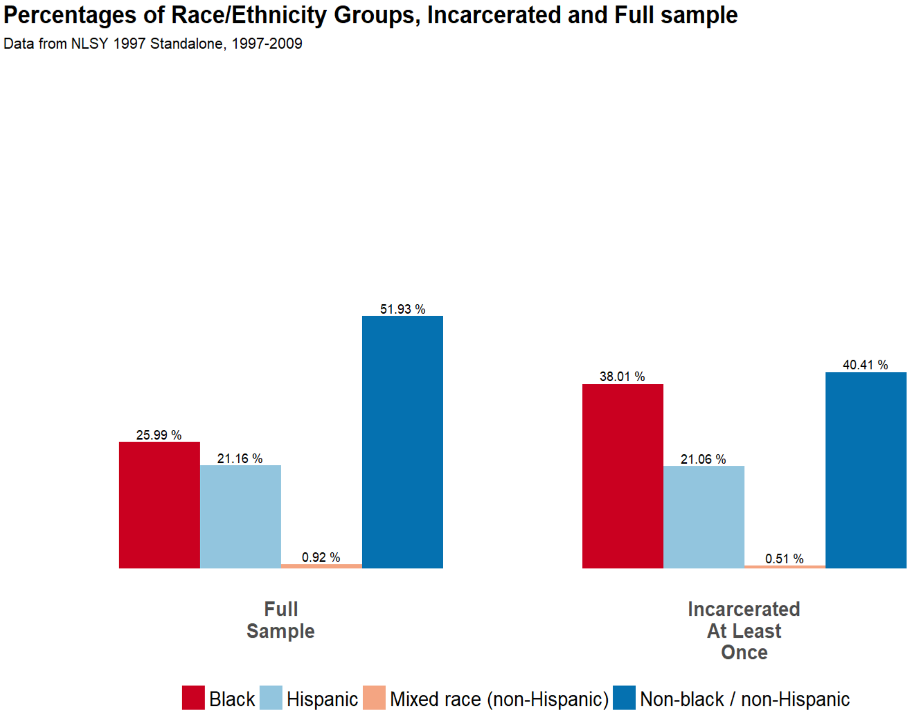

Figure 2: Sample type by race/ethnicity (NLSY97)

When viewing by race, we see that blacks are incarcerated proportionally more, and non-black / non-Hispanic proportionally less. The proportion of Hispanics incarcerated does not change significantly.

Of a total of 20878 self-reported crimes categorized as assault, selling drugs, or property crimes during the sampled years, 24% were assault, 60% were theft or property related, and 16% involved selling drugs.

| | **Self-Reported Violent** | **Self-Reported Property** | **Self-Reported Drug** |
|---|---|---|---|
| **Percent Incarcerated in year** | 24.00 | 59.80 | 16.20 |

Table 3: Self-Reported crimes by type (NLSY97)

Note that an individual could commit more than one crime in a given year, and be incarcerated for a crime different from the self-reported crime. These statistics therefore are noisy and do not represent a one to one relationship between crimes and judicial punishment.

## 6. Analysis and Visualizations

We begin our analysis by supplementing a logistic regression with a LASSO (Least Absolute Shrinkage and Selection Operator) to inform variable selection. This operator penalizes all non-zero coefficients in estimating a model, therefore encouraging parsimony in the choice of non-zero covariates. This method is useful in cases such as ours in which the number of possible covariates is large compared to the number of observations in the model. (Here in our study of recidivism, we compare the population who has recidivated to the population who have committed crimes only once. The restriction that all individuals have committed a crime once limits the number of viable observations in the dataset to 584.) These selection model employed allows us to agnostic about the choice of covariates, while also avoiding issues with overfitting.

The results of the LASSO Model are as follows, with coefficients set to zero omitted.

| Variable | Coefficient |
|---|---|
| **Intercept** | -298 |
| **Sex Dummy (Female)** | -.55 |
| **Race (Non-Black, Non-Hispanic)** | -.72 |
| **Race (Hispanic)** | -.22 |
| **Health Condition After Age 16** | -.07 |
| **Months Incarcerated Age 15** | .003 |
| **Months Incarcerated Age 17** | .08 |
| **Months Incarcerated Age 18** | .06 |
| **Died at age 21** | .13 |
| **Number of Siblings** | .03 |
| **Year of Birth** | .15 |
| **Total Arrests** | .00 |
| **Age at First Incarceration** | -.13 |
| **Died at Age 14** | .00 |

Table 4: Results of LASSO Logit Model

Sex and race have the predicted effect on the odds of recidivating. Whites are least likely to recidivate, and Hispanics are less likely than blacks. Health conditions appear to make recidivism less likely. This could be explained as an increase in the costs of committing crime (more difficult to perpetrate, more limited access to healthcare if incarcerated), or a decrease in the benefits enjoyed by successfully committing a crime.

Notably, as an individual ages, the number of months incarcerated has a growing effect on the odds of recidivism. This could stem from several causes. First, prisons could provide a chance for youth to network and share skills with other criminals. It is also possible that older criminals are more hardened, more likely to recidivate, and more likely to receive more stringent punishment. This effect will be explored by exploring self-reported crimes, and categorizing crimes by their severity.

It is worth noting here that If write a naive model such as a simple logit regression, as is sometimes used in the literature, we find that children are more likely to recidivate. This is not due to any characteristic of the individual himself, but it is an anomaly of the model specification. A younger first-time criminal is observed for longer in the data, and is more likely to recidivate. Therefore, using naive models in sentencing could lead to punishing individuals improperly

Perhaps just as important to discuss when using LASSO, is what variables do not make the cut. LASSO places very little weight on family and income related covariates (excluding the number of siblings variable). Recidivism is largely driven by sex, race, and prior experience in the criminal justice system.

| **Variable** | **Coefficient** |
|---|---|
| **Sex Dummy (Female)** | -.31 |
| **Race (Non-Black, Non-Hispanic)** | -.00 |
| **Health Condition Before Age 16** | .03 |
| **Birth Year** | .02 |
| **Ever Incarcerated calendar year of 14th birthday** | .10 |
| **Total Incarcerations** | .34 |
| **Incarcerate in calendar year of 15th birthday** | .16 |
| **Incarcerate in calendar year of 21st birthday** | .58 |
| **Incarcerate in calendar year of 22nd birthday** | .30 |
| **Incarcerate in calendar year of 25th birthday** | .13 |

Table 5: Results of LASSO with Cox Regression

Now turning to the Cox regression, we can glean the following insights. Crime has gone down in the 1990s. If born earlier, one might be a part of a more violent cohort. The effect of criminal activity on recidivism is more scattered across age groups. Indeed, controlling for the censoring problems of the logistic has alleviate the concern of bias towards high recidivism in young offenders. Again, we see a positive correlation between incarceration and recidivism. Of course, there is endogeneity here, but the correlation challenges the conventional wisdom that we should estimate recidivism to lock people away therefore preventing future crimes. While locking people away more frequently or for longer duration does prevent them from committing crime during this timeframe, this evidence suggests it may make criminal activity more likely when they are finally released.

## 7. Conclusion

This paper explores crime data in the United States using machine learning and data visualization techniques. This analysis was intended to demonstrate the merits of applying machine learning methods to inform evidence-based policy-making. While our discussion has focused on crime and recidivism in the U.S., the applications of machine learning can be used throughout the social sciences wherever a large volume of data exists. It is important to also recognize the limitations of machine learning – namely in understanding social, political and cultural contexts. With this said, we hope that our analysis has produced data-driven research that can inform policy-making in government and the nonprofit sector.

## References

Berk, Richard A., and Justin Bleich. 2013. “Statistical Procedures for Forecasting Criminal Behavior.” *Criminology & Public Policy* 12 (3): 513–44. https://doi.org/10.1111/1745-9133.12047.

Brennan, Tim, and William L. Oliver. 2013. “The Emergence of Machine Learning Techniques in Criminology.” *Criminology & Public Policy* 12 (3): 551–62. https://doi.org/10.1111/1745-9133.12055.

Brynjolfsson, Erik, and Tom Mitchell. 2017. “What Can Machine Learning Do? Workforce Implications.” *Science* 358 (6370): 1530–34. https://doi.org/10.1126/science.aap8062.

Chen, M. Keith, and Jesse M. Shapiro. 2007. “Do Harsher Prison Conditions Reduce Recidivism? A Discontinuity-Based Approach.” *American Law and Economics Review* 9 (1): 1–29. https://doi.org/10.1093/aler/ahm006.

COTTLE, CINDY C., RIA J. LEE, and KIRK HEILBRUN. 2001. “The Prediction of Criminal Recidivism in Juveniles: A Meta-Analysis.” *Criminal Justice and Behavior* 28 (3): 367–94. https://doi.org/10.1177/0093854801028003005.

Gendreau, Paul, Tracy Little, and Claire Goggin. 1996. “A Meta-Analysis of The Predictors of Adult Offender Recidivism: What Works!” *Criminology* 34 (4): 575–608. https://doi.org/10.1111/j.1745-9125.1996.tb01220.x.

“Giving Meaning to Risk Factors.” 2016. January 28, 2016. https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/mnng-fctrs/index-en.aspx.

Holtfreter, Kristy, Michael D. Reisig, and Merry Morash. 2004. “Poverty, State Capital, And Recidivism Among Women Offenders.” *Criminology & Public Policy* 3 (2): 185–208. https://doi.org/10.1111/j.1745-9133.2004.tb00035.x.

Jean, Neal, Marshall Burke, Michael Xie, W. Matthew Davis, David B. Lobell, and Stefano Ermon. 2016. “Combining Satellite Imagery and Machine Learning To Predict Poverty.” *Science* 353 (6301): 790–94. https://doi.org/10.1126/science.aaf7894.

Jordan, M. I., and T. M. Mitchell. 2015. “Machine Learning: Trends, Perspectives, and Prospects.” *Science* 349 (6245): 255–60. https://doi.org/10.1126/science.aaa8415.

Kurlychek, Megan C., Robert Brame, and Shawn D. Bushway. 2006. “Scarlet Letters and Recidivism: Does an Old Criminal Record Predict Future Offending?” *Criminology & Public Policy* 5 (3): 483–504. https://doi.org/10.1111/j.1745-9133.2006.00397.x.

Langan, Patrick A., and David J. Levin. 2002. “Recidivism of Prisoners Released in 1994.” *Federal Sentencing Reporter* 15 (1): 58–65. https://doi.org/10.1525/fsr.2002.15.1.58.

Olver, Mark E., Keira C. Stockdale, and J. Stephen Wormith. 2011. “A Meta-Analysis of Predictors of Offender Treatment Attrition and Its Relationship to Recidivism.” *Journal of Consulting and Clinical Psychology* 79 (1): 6–21. https://doi.org/10.1037/a0022200.

Pearson, Frank S., Douglas S. Lipton, Charles M. Cleland, and Dorline S. Yee. 2002. “The Effects of Behavioral/Cognitive-Behavioral Programs on Recidivism.” *Crime & Delinquency* 48 (3): 476–96. https://doi.org/10.1177/001112870204800306.

Rhodes, William. 2013. “Machine Learning Approaches as a Tool for Effective Offender Risk Prediction.” *Criminology & Public Policy* 12 (3): 507–10. https://doi.org/10.1111/1745-9133.12060.

Stevens, Dennis J., and Charles S. Ward. 1997. “College Education and Recidivism: Educating Criminals Is Meritorious.” *Journal of Correctional Education* 48 (3): 106–11.

Tollenaar, N., and P. G. M. van der Heijden. 2013. “Which Method Predicts Recidivism Best?: A Comparison of Statistical, Machine Learning and Data Mining Predictive Models.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176 (2): 565–84. https://doi.org/10.1111/j.1467-985X.2012.01056.x.

Uggen, Christopher. 2000. “Work as a Turning Point in the Life Course of Criminals: A Duration Model of Age, Employment, and Recidivism.” *American Sociological Review* 65 (4): 529–46. https://doi.org/10.2307/2657381.

United States Department Of Justice. Office Of Justice Programs. Bureau Of Justice Statistics. 2014. “Recidivism in the National Longitudinal Survey of Youth 1997 - Standalone Data (Rounds 1 to 13).” ICPSR - Interuniversity Consortium for Political and Social Research. http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/34562/version/1.