The Role of Decision Support Systems in the Criminal Justice System

Olwyn Doyle

Department of Political Science

Spring 2020

Committee:

Advisor: Joshua Strayhorn, Political Science

Janet Donavan, Political Science

Robin Burke, Information Science

Introduction

The criminal justice system in the United States is increasingly the subject of more and more controversy, with the rate of accusations of racism in police departments consistently increasing as prison populations skyrocket. Similarly, variations in sentencing decisions have raised many questions of discrimination, and thus individual components of the system are under significant scrutiny and pressure to improve policy. Thus, it seems perfectly logical that police departments and judicial systems are turning to decision support systems to help in their decision-making processes. Decision support systems are computational algorithms that inform judgement; some are based on machine learning systems so that the algorithm can learn from previous data what makes an individual more likely to recommit a crime, while others learn from research in criminology and instead just analyze the factors in an individual case. The perceived objectivity of computer systems is very attractive, both to those who genuinely want to improve their practices and those who just want to skirt blame, as individual biases should theoretically be counteracted in a decision informed by a sophisticated algorithm. These algorithms are currently largely used for risk assessment of individual offenders or for predictive policing practices. However, because these algorithms are important intellectual property, the actual logic behind the decisions they are informing is often lost, leaving those who implement them and the those they affect at the mercy of a supposedly objective black box.

As the number of decision support systems implemented in the criminal justice system increases, it is vital to understand their strengths and limitations, especially as they are often not well understood by anyone involved. This technology is consistently improving, but implementing convoluted processes lacking transparency in a system that is already flawed has the unfortunate potential to ruin lives and reinforce existing discriminatory practices. These algorithms need to be tested for bias and fairness before their introduction into life or death scenarios.

The question which this paper seeks to answer is: what underlying assumptions are decision support systems making, how does this impact those who use them, and to what extent is the use of decision support systems in the criminal justice system successful in reducing crime rates? This question is important because up to this point most scholars have addressed this idea from either a political science lens or a computational perspective, but it is the intersection of those two things that provides the best opportunities for understanding. I will argue that the current usage of decision support systems in the criminal justice system is problematic as ensconces bias and discrimination present in the system and provides police departments and court systems with a scapegoat for any perceived discrimination, especially as these algorithms are not widely understood, and furthermore that these systems can be helpful but should be used to root out problematic data rather than iterate on existing problems.

Literature Review

The focal point of this investigation is the impact of decision support systems on the criminal justice system. Decision support systems are machine learning algorithms that compile large amounts of data and apply that learning in order to synthesize those trends into a decision. In the realm of criminal justice, this can be used to assist judges in determining the length of a sentence for a particular offender by analyzing the previous sentences given for similar crimes, along with any additional factors that may be relevant, including criminal history and perceived remorse (Bayamlıoğlu 2018). Data-driven methods are often claimed to be more forward thinking than previous methods as they are working under predictive rather than descriptive models (Conti-Cook 2019).These algorithms are becoming more prevalent in every aspect of society, yet most who interact with them do not understand how they work, or the extent of their deficiencies. The tendency in society is to view computers as objective actors, not swayed or limited by emotional understanding (Ferguson 2019). This can be problematic when the algorithm is synthesizing data that has been generated by people, as although the algorithm may not be intrinsically biased, learning from biased information results in biased systems (Brantingham 2018; Mayson 2018; Richardson, Shultz, Crawford 2019). Moreover, even in instances where individuals understand the subjective elements of decision support systems, they may feel pressure to default to them anyway (Conti-Cook 2019). If a person with an algorithmically low risk of recidivism commits a crime, those involved with their parole or sentencing can point to the computer in defense, whereas actors may be held politically accountable if they choose to override a high risk score and it yields negative results (Conti-Cook 2019). Even more concerningly, the question of bias has only recently come to the forefront of the academic communities that study and develop decision support systems and other AI systems. While these systems have been implemented to varying degrees across the country, many of them have not been tested for problematic results which could come from a flawed algorithm or from flawed data.

This topic is also important because of the specific implications it poses in the criminal justice system. Especially given the multitude of controversies facing police departments and court systems across the United States, it seems natural for these groups to want both to evade direct responsibility and legitimately reduce bias in their policing strategies and sentencing tactics. With an algorithm as a scapegoat or a justification, the criminal justice system can feel more comfortable with its ability to pass judgement (Ferguson 2019). This argument is convincing, but many point to its limitations. On the one hand, depending on how they are implemented, decision support systems can reduce some of the human factors that frustrate sentencing policy especially. Surely with an algorithm, the length of a prison sentence will not be dependent on how recently the judge has eaten or if their favorite sports team lost the game last night, right (Dazinger 2011)? Or, more seriously, can the algorithm reduce the impact of race and other identities that often face discrimination in the criminal justice system? This could be true, but it is easy to forget that the algorithm would still be trained on sentences with lengths that were impacted by human factors. Additionally, it has been repeatedly demonstrated that these algorithms look for the strongest predictors of a specific classification (for example high risk of recidivism), and can often uncover demographic information, even if not stated explicitly, if it is one of the principal components that explain the most variation in the data (Dastin 2018). This is to say that if one of the best predictors of recidivism in a city's crime statistics is race, which can often be the case as a result of biased policing practices, even if the city takes the time to remove the race of all the individuals in the dataset it is still likely that the algorithm will be able to detect race from other factors and use it in decision making thereby iterating on the bias already present in the system.

One of the other aspects that makes this topic important is the lack of transparency surrounding decision support systems. The algorithms themselves are not usually released to the public, as they are the backbone of an industry surrounding AI-based systems. If the algorithm was publicly available any number of other individuals could reproduce it and the companies involved would not have anything to sell (Diega 2018). This means back solving from data, also known as third party analysis, if it is available to try and determine how the algorithm is weighting different factors (which requires both the data and explicit knowledge of considered factors), or just a complete lack of transparency (Ferguson 2017). As a result of this, many scholars argue that these systems do not reduce bias at all, that they work as a new layer to obfuscate larger problems in the criminal justice system because the problems lie in the data itself and biased or flawed data can only produce biased or problematic results (Ferguson 2019). Others suggest that while these algorithms are largely flawed in their current implementation, they can be shifted to root out bias and serve as a check on the system and its components. The last large camp of researchers points to the incredible benefit that can result from decision support systems as it reduces the subjectivity of each decision, and so should result in a reduction in bias over time. This literature review will provide background as to the perspectives of the three predominant arguments on the subject to address the extent to which decision support systems are beneficial in the criminal justice system.

Those who see decision support algorithms as helpful point to the fact that there are trends in sentence lengths and that there are trends in which individuals are most likely to recommit criminal offenses (Bayamlıoğlu 2018). This is true, in fact Dr. J.C. Oleson from the University of Auckland has published a paper finding “seventeen discrete variables that appeared to be significantly associated with recidivism” (Oleson 2011). Additionally, court systems have previously used statistical analyses on individuals to help determine their relative risk, and thus the use of algorithms would mainly help to facilitate the speed and scope of these analyses (Kopf 2015). Given the large number of tests and measures present in criminal justice, it makes sense to synthesize these into a more tangible response system to predict risk levels (Kopf 2015). On top of this, proponents of the system point to systems such as the Federal Post-Conviction Risk Assessment (PCRA) which do not score for race, gender, and age, which is to say that these factors do not impact the risk assessment score, but are instead used to be able to determine systemic barriers that face these groups in order to account for them in correctional treatment, so these same considerations could be imbued in a machine learning system as well (Kopf 2015). While opponents of the usage of decision support systems in the criminal justice system often claim that using algorithmic risk assessment so heavily is punishing offenders for crimes they have not yet committed, scholars in this camp point to Congressional code which states that the court should also evaluate the need “to protect the public from further crimes of the defendant”, and thus decision support systems are acting in accordance with this initiative (Imposition of a Sentence 1984). Thus, these scholars see data-driven policing as realizing the potential of data collection and expanding on current practices in ways that can reduce discrimination and crime rates simultaneously.

Others see decision support systems as imperfect but potentially helpful. The overall argument from this faction suggests that the current implementations of decision support systems are largely problematic, but that they can be redirected in order to focus on finding bias, rather than perpetuating it, or they can be trained to account for the bias present in the data (Brantingham 2018; Ferguson 2019). These are two varying levels of trust in the use of algorithms, the first suggests that they should not be used to determine risk factors for criminal offenders but rather be appropriated to root out problems, while the second suggests tweaking and adjusting the existing algorithms to provide updated weights for factors, or to post-process recommendations to reduce some of the known systemic biases (Brantingham 2018; Ferguson 2019). The large quantity of data available could be used to develop different sorts of recommendations and mapping that focus on need rather than danger (Ferguson 2019). Rather than using these data-driven methods to root out criminals it could help to find systemic processes that result in crime and thereby address the cause of the issue instead of the symptoms.

This could take the form of police departments reaching out to families who live in underserved areas, and it could even help governments to adjust policy to fix problems and help communities at a more granular level (Ferguson 2019). Additionally, it could potentially help to direct policing based on factors such as mental health and addiction, which often get overlooked in the broader push to reduce crime (Ferguson 2019). Police departments under pressure from local and federal governments could use these systems to point to real solutions that areas need, and shift some of the pressure back on to policymakers. Given that a lot of issues in biased or incomplete crime statistics come from departments trying to under or over report in order to reach quotas, this could provide some ability for police departments to push back and reduce the bias in their systems. More generally, this tool set allows the sharing of data across different regulatory bodies in order to better help the population, if police officers log the locations of people who are arrested or ticketed in conjunction with mental health or addiction, the city can direct its resources to these areas in order to ameliorate the effects via public health programs, which will both help the society as a whole and reduce crime by addressing the source of the issue.

Adjusting the algorithms could be greatly facilitated by using a shifted algorithm that looks for bias in the data to change the weights produced by an initial synthesis of the data. Given the fact that simulations that artificially added biased downgrading of crimes (understating the severity of a crime) and biased upgrading of crimes (overstating the severity, usually to maximize liability) impacted the risk assessments of the predictive policing model ETAS (Epidemic-Type Aftershock Sequence), it would seem that in instances where overall or specific biases can be determined, the algorithm could be adjusted to account for these problems (Brantingham 2018).

As another example, researchers have developed newer models for predictive policing to mitigate the feedback loops present in the current models (Ensign 2018). Feedback loops are a major issue with predictive policing systems and decision support systems in general because as police officers increase their presence in areas suggested by the algorithm, they are more likely to find crime there (Ensign 2018). This new arrest data gets added to the system and reinforces the idea that the community in question needs more police presence and this process continues recursively (Ensign 2018). Thus reducing the effect of the feedback loop is likely to reduce discrimination in the algorithm, demonstrating that there are still many different facets of these systems than can be altered to improve their accuracy and reduce bias (Ensign 2018). Similarly, the algorithms could be adjusted using an unmonitored approach by dipping into the literature on biases in crime statistics and pre-processing the data to account for these. Overall, the mixed perspective focuses on the good that could be done with the imperfect tools available, while acknowledging their limitations.

Others see decision support algorithms as not addressing the problem in any way, as the problem comes from the data itself and not the way it is wielded. They point to the fact that crime is significantly underreported as police departments try to select out their crime statistics under pressure from the government, and so the available data is already only an unrepresentative sample (Brantingham 2018; Mayson 2018; Richardson, Shultz, Crawford 2019). This limits the effectiveness of artificial intelligence-based approaches to learning. More than this though, these algorithms have a racialized component, as the criminal justice system has often targeted black men, which means that “criminal history now correlates with race” (Mayson 2018; Richardson, Shultz, Crawford 2019). This is problematic when decision support systems calibrated for risk assessment utilize criminal history, as the results will be heavily correlated with race (Mayson 2018). Given that most police departments are at least trying to avoid the perception of race-based decision making, implementing a system with disparate impacts across racial lines is incredibly problematic. Despite this, recommendation technologies are largely viewed as race-neutral and objective, even though this claim has been falsified repeatedly (Ferguson 2019; Mayson 2018). While it is deeply problematic that these systems are biased and produce flawed results, they can be even more insidious because of the (somewhat inherent) lack of transparency associated with them, and because they can be used to justify unfair policing practices (Ferguson 2019).

Depending on how the algorithm is calculated, risk scores may be reinforcing larger theories about crime in a less transparent and obvious way. Thus not only can algorithms justify unfair policing practices, but they can also be used to reinforce a more general framework about who commits crimes. As discussed earlier, these decision support systems can create feedback loops which serve as an algorithmic confirmation bias; if the algorithm previously determined someone to be high risk and then that person is monitored more closely by the police they may be arrested for more crimes than they would have otherwise, and those who were assumed to be low risk may skirt criminal charges because they are not the focal point (Ensign 2018). Moreover, if certain groups of people are overrepresented in the data used to train the algorithm, then those who use it may begin to develop stronger biases against that group of people because they are constantly associated with higher scores (Ferguson 2019). This is especially true as police officers may or may not know what factors are used in calculating these scores, and even when they do, they may not know how heavily they are weighted.

When factors are highly subjective their impacts can be even more problematic. For example, some decision support systems have implemented levels of remorse as a variable which is clearly not an objective fact and can be manipulated by the exact human elements these algorithms are supposed to minimize (Bayamlıoğlu 2018). In the example of the Correctional

Offender Management Profiling for Alternative Sanctions (COMPAS), which is the recidivism algorithm used in many states including New York to determine parole eligibility, one of the questions utilized by the algorithm is “Does this person appear to have notable disciplinary issues?” (Conti-Cook 2019). The question is highly subjective, those answering it do not receive any training about what constitutes disciplinary issues and thus there is high variation in the responses, and it has been experimentally deduced to be a highly weighted question that can completely shift an individual’s likelihood of being granted parole (Conti-Cook 2019). Furthermore, the parole board almost never ignores the recommendation of COMPAS because if an offender was granted parole over the judgement of COMPAS and then committed another crime it would be incredibly controversial (Conti-Cook 2019). Without the understanding of the underlying mechanisms of a decision support system, those who implement them may just see the classification of high risk and treat it as fact. Algorithms are often instituted without the full understanding of the public, which can create problems when a population is told that their risk level is being assessed without any understanding as to how or how this information will be used (Ferguson 2017). Additionally, given these metrics can be used to assess appropriate levels of force and suspicion for the police to use, it is important that they are fully understood both by the citizenry and by the academic community so that they have been properly tested and assessed for bias and problems in logic (Ferguson 2017).

Furthermore, nuances present in data-driven approaches become “legal truth” when a judge or police officer sees a recommendation without any supporting information. Even if the system provides a breakdown of the factors that went into the decision, it does not necessarily mean that those who use them will look into the factors and how they are weighted for the individual, but rather see the overall recommendation provided by the algorithm and take it as fact (Bayamlıoğlu 2018). This means that while a system may be intended to serve as a guide or suggestion, it is given more weight because it does not provide more than a classification as higher or lower risk in terms of recidivism, or a general sentence (Bayamlıoğlu 2018). Along with this, given algorithms normally rate offenders along a spectrum from low risk to high risk (though it may even use three discrete categories which is even more problematic as it flattens nuance to a greater extent), which frames the individual in a negative light regardless of what they have actually done (Bayamlıoğlu 2018). These systems always suggest some level of risk of recidivism, and although this may make sense logically as we can never be completely certain that an individual will not commit another crime, it paints each person as a risk to society. Also, from a criminal justice context, this works against the idea of innocent until proven guilty by predicting guilt in future crimes and framing everyone as a potential repeat offender (Bayamlıoğlu 2018).

The focal point of these set of arguments is the flaws that have been identified with these systems and with the data sets they are trained on. There exists a gradient amongst the arguments; some opponents suggest that their implementation is the problematic aspect, while others believe that the idea of using prediction as a metric at all is flawed, that there is no fashion in which the algorithms can be changed or updated to account for this, and that using these metrics even without an algorithm is still problematic as it perpetuates the cycle of incarceration for certain communities (Mayson 2018). However, the latter is an extreme view, and even most of those individuals who disagree with the use of decision support systems do not dismiss them outright.

In this paper I will argue that although decision support systems are problematic in their implementation in the criminal justice system thus far, they can be used to root out the problems that undermine their current usage or tested in order to assess the bias and fairness of the algorithm. Using big data analyses could allow departments to try and find bias in their data in order to address them moving forward. These techniques should be shifted to discover flawed data and practices rather than iterating on bias. This will take some calibration as the current algorithms have to be tweaked, but research should be pursued to develop our understanding of the extent to which they can help. Decision support systems could be used to compare sentencing decisions from various judges, complete with a detailed readout that expresses the factors of analysis and the raw score as well as the decision itself.

The argument that decision support systems are just speeding up the tools already used by court systems such as using a model to plot the AUC of a specific individual is missing the fact that these models allow for more nuance, the person creating the model can see the components that are impacting the final result and choose to ignore the results. This is usually not the case for decision support algorithms as the algorithms themselves are not explained to those who utilize them, and they often only output a classification (Brantingham). Additionally, many of the factors that scholars have identified in larger analyses of recidivism are subjective. Using “whether the person has antisocial attitudes and values” as a marker requires an officer to determine the extent to which this is true, further undermining the conceptualization that data-driven methods can be objective (Kopf 2015). Furthermore, this subjective indicator will get magnified when fed into any form of machine learning system, and those who collect the data inputted into the decision support systems often do not understand the extent to which their answer impacts the outcome of the system as a whole (Kopf 2015; Conti-Cook 2019; Oleson 2011). Without the knowledge that a question is highly rated, people may be flippant in their response without realizing the severe impact that action carries (Conti-Cook 2019). Also, the extent to which these data-driven methods are protecting “the public from further crimes of the defendant” can be highly limited by the problems present in a given algorithm (Imposition of a Sentence 1984). An algorithm that is systematically biased against certain groups is not analyzing the specific case of a defendant but rather the demographic data that comprises them which is not the desired outcome from these systems (Richardson 2019). The current implementation of decision-making systems provides a litany of problems and limitations, but this kind of technology can be recalibrated to create helpful checks on systems and support populations rather than perpetuate cycles of incarceration.

Research Design

This paper seeks to determine the underlying assumptions made by decision support systems and the extent to which they are successful in reducing crime rates. It is important to analyze a real instance of these algorithms in order to increase the transparency of that system in particular, as well as demonstrate the limitations in a specific instance, so I will be assessing the Strategic Subject List from Chicago, Illinois. I will outline my research design and data and then provide relevant analysis before addressing broader issues in relation to the topic.

For this project it is most important to fully define decision support systems. There has been some shift in the definition of this term over time. In this paper, they will be defined as “a computer-based information system that combines models and data in an attempt to solve semi-structured and some unstructured problems with extensive user involvement” (Turban, Rainer, Potter 2005). This definition is used as opposed to the others as it is the only definition to highlight the user involvement, and it addresses that the problems do not necessarily have to have structure. These aspects are important because they emphasize that the algorithm is not meant to make decisions on its own, it is meant to inform a user to support their decision-making process. This is different from a recommendation algorithm as it is not personalized to the individual traits or tastes of a user; risk assessment measures do not vary based on the judge as law should not vary in different courts.

In this analysis I focus mainly on the Strategic Subject List (SSL) from Chicago. This algorithm was selected as the Chicago Police Department has an especially bad reputation resulting from accusations of corruption and racism and operates in the third largest city by population in the United States (Andonova 2019; United States Census Bureau, 2018). This means that the impacts of the implementation of a decision support system in this context has a significant impact on the citizens of Chicago and the city had a high incentive to switch to algorithmic methods to reduce backlash. As a result of this, their decision support system was fully implemented early on, resulting in a larger temporal dataset than from other cities. Also, the SSL suffered lots of backlash as a result of its initial lack of transparency; the public had no conception or how these scores were calculated or what the police were using them for (Posadas 2019). This makes it interesting to study as a major concern of many critics of decision support systems is that they are largely hidden from the public. The SSL is also significant in that the full data set is now published online in an attempt to improve transparency, complete with the specific factors the police department used to generate risk assessment scores, which allows for the reverse engineering of the algorithm being used ("Strategic" 2017). Given that the algorithms are almost never released to the public because they are the intellectual property of the groups who develop them, the ability to reconstruct an algorithm is rare and provides a deeper view as to how these kinds of algorithms function (Diega 2018). Along with this, because the data set includes demographic and location data it lends itself to more in-depth analyses of bias in the data.

Data

To analyze this data I will implement a host of tools. I will reconstruct the algorithm using regression analysis from the dataset and analyze the coefficients to see which variables have the strongest impacts on the raw SSL score. It is important to note that the Strategic Subject List risk scores try to predict the risk of being involved in a crime as either a victim or a perpetrator rather than just the risk of recidivism. I will also plot the average SSL score for each tract side by side in order to see the distribution of SSL scores across tracts. Then I will compare the location data available in the Strategic Subject List to the census records for Chicago and the community area dataset to determine if any demographics are over or underrepresented in the data set (“Chicago Community”; “Strategic” 2017). Additionally, I shall be assessing the introduction of the SSL as an intervention. The city of Chicago suggested that the SSL would be used to reduce gun violence in the area, and so I will use a lagged time series to compare gun violence in the city before and after the full implementation of this algorithm, in order to evaluate its success. I will also be comparing the crime rates in Chicago to those in Cincinnati to determine if any perceived reduction in crime is as a result of the intervention or if it is part of a larger trend of decreasing crime rates.

Analysis

I. Third Party Analysis

In attempting to reconstruct the algorithm that was used to develop the SSL scores, I ran a linear regression on the terms to determine how the different factors were combined. The eight factors that are used to calculate the SSL score are the individual's age at the time of their most recent arrest, the number of times the person has been the victim of a shooting, the number of times the person has been the victim of assault or aggravated battery, the number of times the individual has been arrested for a violent offense, a binary value with a value of one if the person is a member of a gang and 0 otherwise, the number of times the person has been arrested for narcotics, the trend in an individual's criminal activity, and the number of times the individual has been arrested for Unlawful use of Weapons. In the dataset, the age variable is divided into seven ranges which are shown separately in my analysis. From the regression table in Table 1 we can see the R^2 value is 0.953 which means that approximately 95.3% of the variation in the data can be explained by the included variables. This is a surprising result, getting an R^2 that high on a first cut linear regression signifies that the true algorithm behind the SSL is quite simple, not much more than a linear combination of terms. Additionally, we can see that all the variables are statistically significant at a p-value of 0.001. The largest single coefficient in the

Table 1: Full Regression on the raw SSL score

	Dependent variable:
	RAW.SSL.SCORE
GANG MEMBERSHIP	0.001*** (0.00001)
AGE AT LATEST ARREST 20-30	0.049*** (0.0002)
AGE AT LATEST ARREST 30-40	0.041*** (0.0002)
AGE AT LATEST ARREST 40-50	0.033*** (0.0002)
AGE AT LATEST ARREST 50-60	0.025*** (0.0002)
AGE AT LATEST ARREST 60-70	0.017*** (0.0002)
AGE AT LATEST ARREST 70-80	0.009*** (0.0002)
AGE AT LATEST ARREST LESS THAN 20	0.055*** (0.0002)
NUMBER OF TIMES VICTIM OF A SHOOTING	0.003*** (0.00003)
NUMBER OF TIMES VICTIM OF ASSAULT	0.006*** (0.00002)
NUMBER OF TIMES ARRESTED FOR VIOLENT OFFENSE	0.003*** (0.00001)
NUMBER OF TIMES ARRESTED FOR NARCOTICS OFFENSE	0.001*** (0.00000)
TREND IN RECENT CRIMINAL ACTIVITY	0.003*** (0.00001)
NUMBER OF TIMES ARRESTED FOR UNLAWFUL USE OF WEAPONS	0.002*** (0.00002)
Constant	0.042*** (0.0002)
Observations	398,684
R^2	0.953
Adjusted R^2	0.953
Residual Std. Error	0.002 (df = 398669)
F Statistic	576,462.000*** (df = 14; 398669)
Note:	*p<0.1; **p<0.05; ***p<0.01

table is for age at latest arrest less than 20, with a value of 0.055. As many of the other factors are also binary, we can compare their values directly, with the exceptions being the number of narcotic arrests and the trend in criminal activity. For the former we can see that with every additional arrest related to narcotics results in an increase of 0.001 on the raw SSL score, and for the later for every one point increase in the trend of criminal activity there is an increase of 0.003 in the overall score. We can see that the coefficients for the age variables have the greatest impact on the raw SSL score. This is not automatically unreasonable; in criminology it is largely accepted that age is an excellent predictor of crime with people tending to be involved with fewer crimes as they get older (Hirschi 2017). To further analyze the extent to which the data is influenced by age, I also ran a regression on just the age variables. From this table (Table 2) we see that our R^2 value is still 0.886, meaning that age accounts for 88.6% of the variation in the data. Thus it would seem that the vast majority of the SSL score is actually determined by age. This would not necessarily be problematic if there was greater transparency for those interacting with these scores, but because this measure is supposed to be providing risk assessment, it could create serious issues.

Table 2: Regression for raw SSL score with only age variables

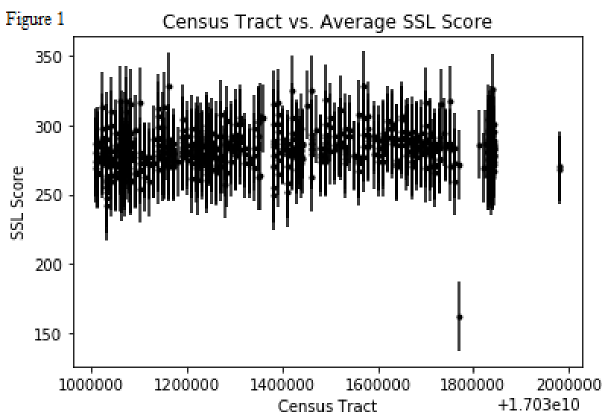
	Dependent variable:
	RAW.SSL.SCORE
AGE AT LATEST ARREST 20-30	0.049^{***} (0.0004)
AGE AT LATEST ARREST 30-40	0.041^{***} (0.0004)
AGE AT LATEST ARREST 40-50	0.033^{***} (0.0004)
AGE AT LATEST ARREST 50-60	0.025^{***} (0.0004)
AGE AT LATEST ARREST 60-70	0.017^{***} (0.0004)
AGE AT LATEST ARREST 70-80	0.009^{***} (0.0004)
AGE AT LATEST ARREST LESS THAN 20	0.056^{***} (0.0004)
Constant	0.043^{***} (0.0004)
Observations	398,684
R^2	0.886
Adjusted R^2	0.886
Residual Std. Error	0.004 (df = 398676)
F Statistic	$443{,}811.500^{***}$ (df = 7; 398676)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

The other pertinent step in understanding the algorithm and the logic behind it is to determine how the coefficients were developed. There are two main options for methodology, either the researchers at the Illinois Institute of Technology determined them artificially based on theories of criminology or they developed them from existing data. The Strategic Subject List dataset includes two values for SSL score, the raw SSL score and the SSL score after it is processed to fit it to the desired range ("Strategic" 2017). This regression uses the raw score rather than the standardized values in order to see what the true coefficients are. In the case of a set of theory based coefficients we would anticipate that the values would be round numbers distributed evenly as they would have been chosen by humans while in the event that they were developed empirically we would expect more irregular values. From the coefficients in the model (Table 1) it appears more likely that the researchers were working from a data-driven approach. This is problematic because it means that the Chicago Police Department is likely iterating on any problems that were present in their existing data, and those biases and issues present in the initial data are now compounded and reinforced by the algorithm and the use of the decision making system. This system has likely created a feedback loop; if the police think that certain people are high risk, they will pay more attention to them and as a result be more likely to have problems with them (Ensign 2018). Additionally, if a police officer notices that every person under 30 who they interact with has a high risk score, because of the perceived objectivity of the algorithm they might begin to develop a negative bias against youth people and start to perceive younger people as more dangerous and prone to criminal activity (Ferguson 2017).

The Chicago Police Department looks at any individual with an SSL score over 250 as high risk and almost everyone in the dataset under the age of 40 has a score above this threshold and no one older than 40 has a score that is deemed to be high risk (Posadas 2019). On top of this, the dataset seeks to determine those who are most likely to be involved in a crime as either perpetrator or victim but has no method for distinguishing between the two. The major issues with this arise to varying degrees based on the level of transparency that is afforded to those who are interacting with the scores along with how they are being used. Firstly, it is unlikely that those who are using this list in the Chicago Police Department are aware of the large role that age plays in calculating an individual's score as the decision support system that generates the list is the intellectual property of the Illinois Institute of Technology ("Strategic" 2017). While age has long been a factor that police officers and judicial systems use when assessing risk, the lack of transparency afforded by decision support systems means that officers cannot determine for themselves how a factor like age should be weighted because it has become "legal truth" (Bayamlıoğlu 2018; Mayson 2018). Additionally, it is unknown whether or not the CPD was aware that these values include both victim and perpetrator risk assessments. This is problematic because if the officers are not aware that a high score could present someone as a victim, they may treat those individuals as a threat rather than people to be protected. More broadly, because the CPD does not track exactly how these scores are used it is very difficult to determine their effects which makes finding problems in the system a largely impossible task (Posadas 2019). Even in the case where every institutional body that interacted with the list fully understood the limitations and strengths of the decision support system as well as the realities of what the algorithm produces, it is still concerning to use a risk assessment score like this one as can reinforce existing trends in the data in the case that the coefficients were determined from some previous dataset and it takes the age out of crime theory and turns it into a fact (Bayamlıoğlu 2018, Mayson 2018). The transparency of the algorithm significantly impacts the extent of the dangers it provides, and without a full explanation from the CPD we cannot enumerate precisely the flaws in the system.

II. Demographic Analysis

Next in the analysis, I wanted to use census data to compare the demographics of those on the Strategic Subject List with the communities they come from to see if there were any large discrepancies, as well as compare the average scores across the different census tracts. From



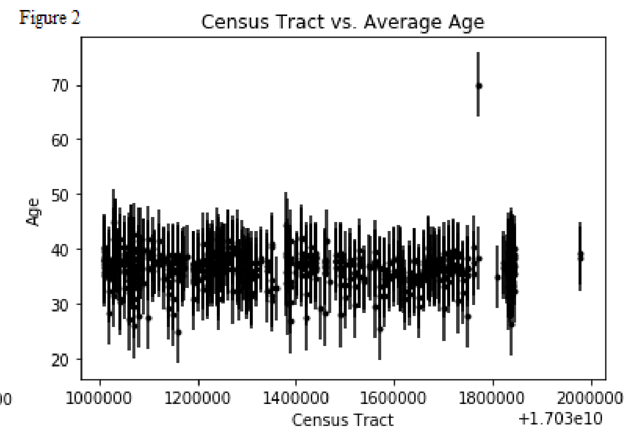


Figure 1 and Figure 2 we can further see that SSL score and age are highly correlated. These graphs show the tract's average values with a 95% confidence interval. There appears to be one census tract with an average age of 70 and an average score of around 170 which stands out from the majority which have an average score from 250 to 325. It is also interesting to note that given the threshold at which an individual is deemed to be high risk is 250, it is worrying that from the figure we can see that the vast majority of the census tracts have an average SSL score above that mark.

In order to compare the different regions in the dataset to their real-world counterparts, I compared the demographics of the Strategic Subject List to their counterparts in the Chicago Community Area dataset ("Strategic" 2017; "Chicago Community"). Chicago is divided into 77 community areas, 75 of which have been officially defined since the 1920s, with the last changes to the boundaries completed in 1980 (Burgess). I decided first to look at the distribution of race in the strategic subject list as compared to the community areas. Figure 3 shows the percentages of each race in each community present in the Strategic Subject List back to back with the percentages of each race in the community overall. As can be seen from this graph, black people are overrepresented in the Strategic Subject List across many different communities. It is additionally interesting to note that there are is a very low percentage of Asian

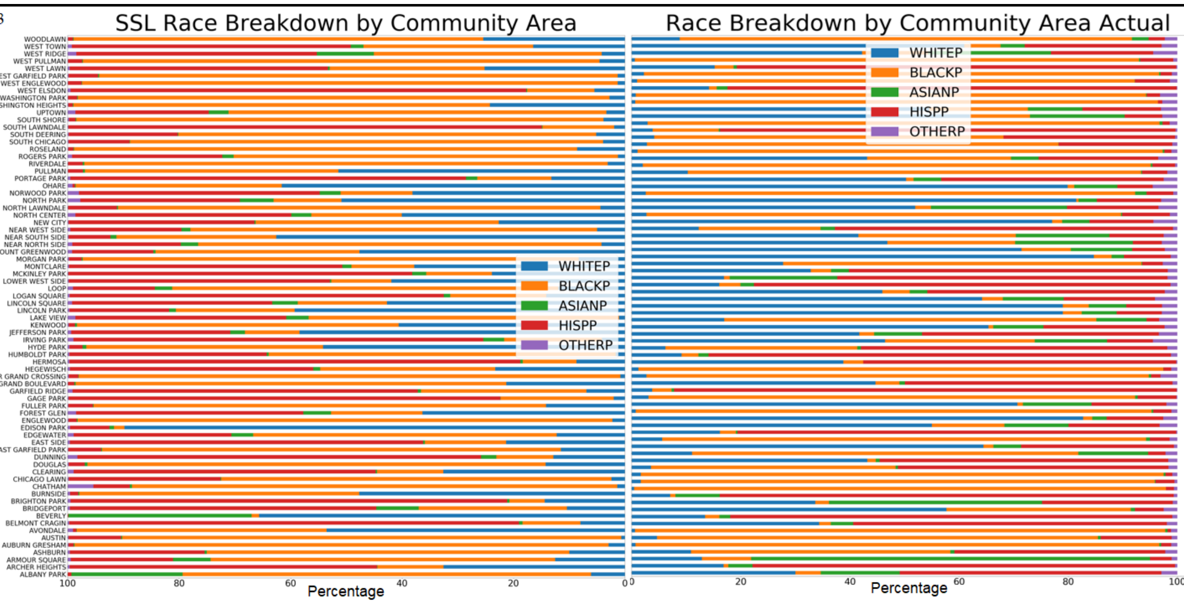


Figure 3

individuals in the Strategic Subject List despite some strong representation in the community areas. Significant also is the low rate at which white individuals appear in the SSL. Figure 4 highlights the percentages of each race in the community area known as North Park which has

Figure 4: North Park Community Percentages

	WHITE	BLACK	ASIAN	HISP	OTHER
Overall Community Demographics (percentages):	51.953084	2.871245	24.960195	16.7392	3.476276
SSL Community Demographics (percentages):	50.842697	12.219101	6.039326	28.651685	2.247191

one of the most prominent discrepancies in the data. In this community, the percentage of white people in the SSL matches with that of the community in general, but there is a difference of 10% between the observed proportion of black people and the actual proportion of black people. The proportion of Asians in the SSL is 18% lower than that in the community, and the proportion of Hispanics is 12% higher in the SSL. This is concerning because the dataset holds individuals who have been arrested, and this would suggest that black people are overrepresented in that population. If the decision support system behind the Strategic Subject List was developed experimentally through existing data, this is a sign of a problem wherein the algorithm would be more likely to produce discriminatory scores as a result of the discrepancies in the data it is fed (Ferguson; Mayson 2018). Figure 5 holds the race overall proportions for the community area dataset as compared to the SSL dataset ("Chicago Community"; "Strategic" 2017). As can be

Figure 5: Race Proportions

	White Percentage	Asian Percentage	Black Percentage	Hispanic Percentage	Other Percentage
Community Area Race Proportions	32.70529	6.155047	30.112605	29.014481	2.012577

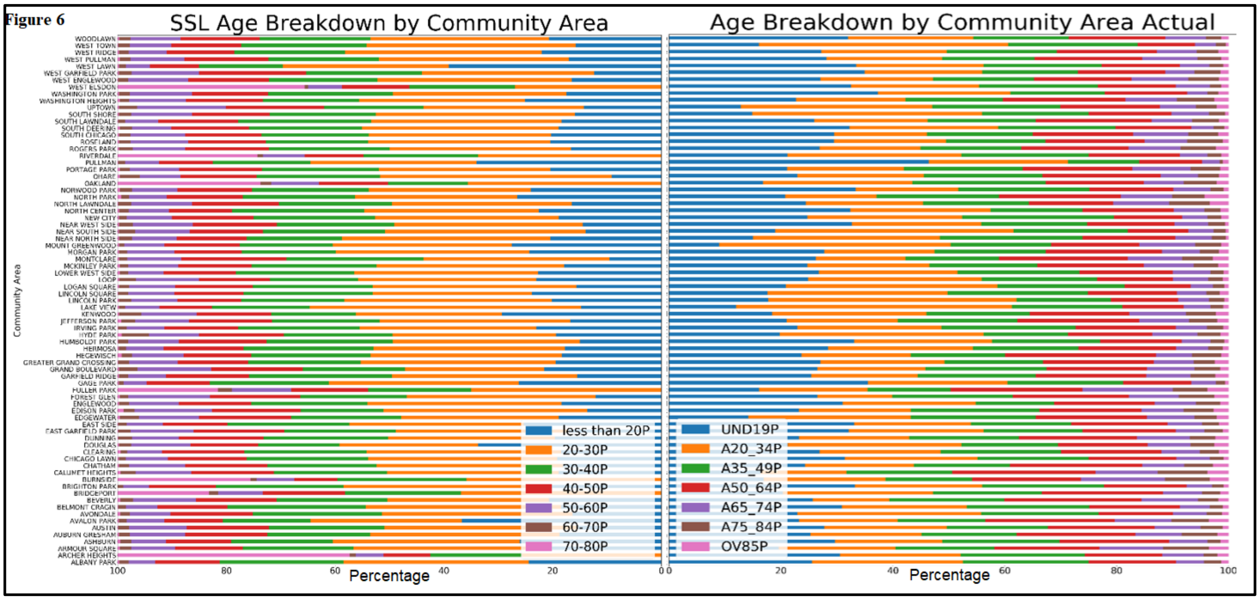
	White Percentage	Asian Percetage	Black Percentage	Hispanic Percentage	Other Percentage
Strategic Subject List Race Proportions	12.427206	1.127076	63.935404	22.154202	0.356112

seen in the figure, white people make up 33% of the overall population but only 12% of the people in the Strategic Subject List. At the same time, Asians make up 6% of the population but only 1% of those in the SSL. The biggest difference is present in the proportion of black people in each, with a population that is 30% black but a Strategic Subject List proportion of 64%. This is especially problematic given that over two-thirds of the people in the SSL have a score that is considered to be high risk (Posadas 2019). Thus this discrepancy between the community area and the SSL also manifests in a greater proportion of certain groups with high risk scores, which entrenches the discrimination already present in the system.

As the Strategic Subject List largely seems to be recreating age, I also did an age breakdown between the two datasets. It is important to note that the two datasets use different age categories, the SSL mostly consists of ten-year ranges while the community area data represents fourteen-year ranges. Despite this we can still compare the data relatively well. From

Figure 6 we can see that the distribution of ages is far more closely matched between the SSL and the community areas than the distribution of race, but there are still some important differences. The 20-30-year-olds in the Strategic Subject List represent a higher proportion of the total than 20-34-year-olds in the community areas, despite the fact that the community dataset includes four additional years in that category. In Figure 6 we can see that young people are overrepresented in the list as expected though not necessarily to the extent anticipated given the

Figure 6



theory that individuals age out of crime and the fact that the Strategic Subject List was generated through arrest data. We might have expected a much greater percentage of those arrested to be young people, but we can see that there are still a fair number of older individuals being arrested as well. In Figure 7 we can see a comparison of the overall age proportions between the

Figure 7: Age Proportions

Community Area Age Proportions	Under 19 Percentage	20-34 Percentage	35-49 Percentage	50-64 Percentage	65-74 Percentage	75-84 Percentage	Over 85 Percentage
	24.134147	27.386408	20.11054	16.704075	6.691095	3.53087	1.442792
Strategic Subject List Age Proportions	Under 20 Percentage	20-30 Percentage	30-40 Percentage	40-50 Percentage	50-60 Percentage	70-80 Percentage	
	18.162235	34.368784	20.800739	14.445763	9.383725	0.82509	

community area dataset and the SSL dataset, although again the age categories are slightly different in each dataset (“Chicago Community”; “Strategic” 2017). Again the 20-30 age range in the SSL has a higher proportion than the 20-34 range in the community areas, which shows some overrepresentation in the 20-30-year-old population in the Strategic Subject List. Additionally, the 75-84-year-olds and the over 85-year-olds in the community area hold a combined 5% of the population, while above 70-year-olds make up less than 1% of the people in the SSL.

III. Effect on Crime Rates

The final piece of my analysis was to determine the success of the list. Given the list was introduced as an intervention to help to lower crime in Chicago, particularly gun violence, I created a time series to analyze the success of that mission. Figure 8 shows the number of crimes per year from 2001 to today, where we can see a steady decrease in the amount of crime per year. This shows crime trending down in general, but it is important to compare the time series data for Chicago to that of a comparable city without a Strategic Subject List or similar program in order to accurately assess the success of the SSL as an intervention and avoid falsely attributing the decrease in crime to the list. I chose to compare the crime rate in Chicago to that in Cincinnati, Ohio, because these two cities have similar murder rates and overall crime rates, and Cincinnati did not

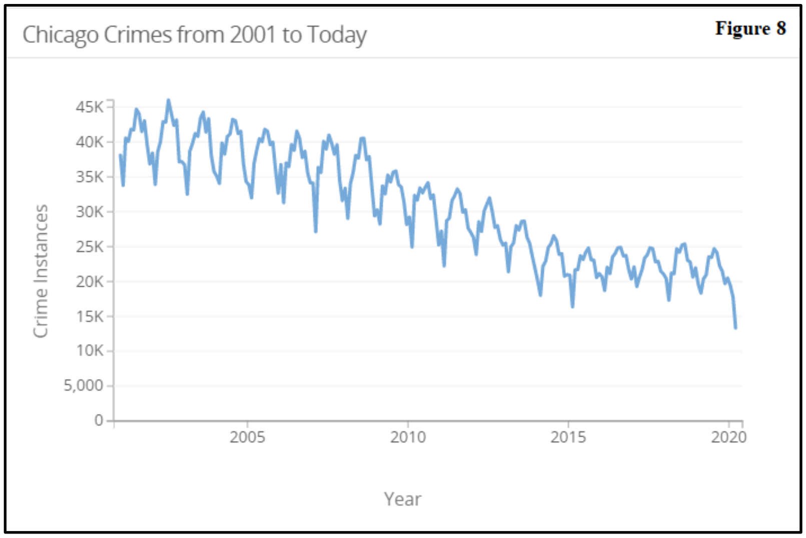


Figure 8

have a program similar to the SSL in this time period. ("FBI 2018"). Figure 9 and Figure 10

Figure 9

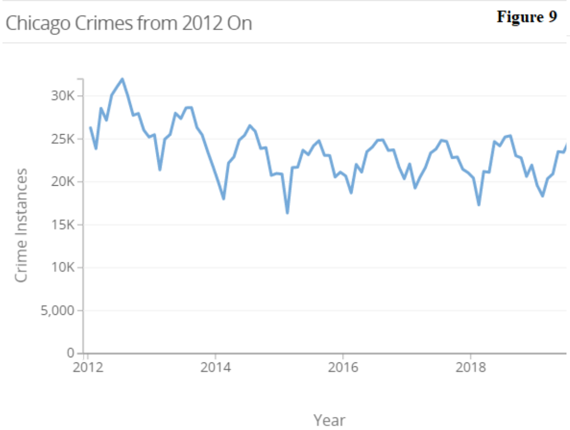
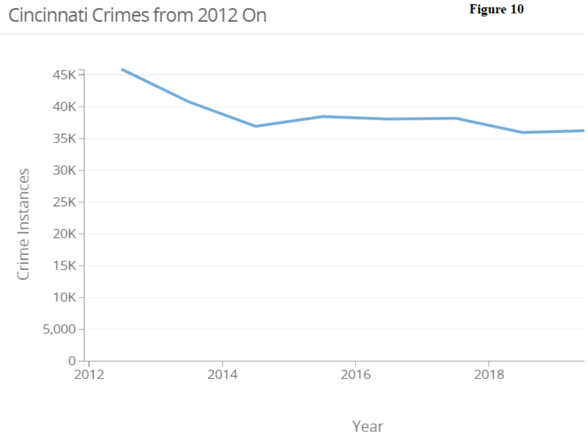


Figure 10



show the number of crimes per year from 2012 on. The Chicago crimes dataset amalgamates data by month whereas the Cincinnati crime dataset has it shown by year ("Chicago Crimes"; "Cincinnati"). However, from these datasets we can see that the dip in crime from 2012 to 2014, which we may have initially attributed to the Strategic Subject List which went into effect on August 1, 2012, appears in both cities. Furthermore, in analyzing the percentage decrease from the start of 2012 to the start of 2018, we can see that Cincinnati experienced a 21.51% decrease in crime while Chicago experienced a 20.27% decrease in crime. This would suggest that the Strategic Subject List may not be the cause of the reduction in crime in Chicago, which is significant because it means that the intervention does not appear to have accomplished the desired positive effect but is has resulted in increased tension between the citizens of Chicago and their police department as a result of the lack of transparency in its creation or usage, and for a time reinforced ideas about age and crime amongst those using these metrics which would suggest that the Strategic Subject List created more problems for the city of Chicago than it alleviated (Posadas 2019; Mayson 2018).

Conclusion

The role of decision support systems in the criminal justice system in the United States is a complicated issue, there are many arguments for and against the use of data-driven measures in general and they may vary widely based on how the decision support system is being used. Decision support systems can often create more problems than they appear to solve as they entrench bias in the data while reducing their transparency and can result in a theory becoming a fact of the criminal justice system. Chicago's Strategic Subject List does not appear to have caused the result that the city was hoping for as it does not seem to have decreased crime in the city more than would have been expected already, and as it largely captures age it could have created problematic understandings of the criminal population, depending on the extent to which it was used and the knowledge of the system of those working with it. The Strategic Subject List created animosity in Chicago as a result of the lack of transparency to the public. This analysis is important in that it combines a political scientific understanding of the criminal justice system with the technical knowledge of decision support systems and their associated strengths and limitations. Additionally, this work provides an in-depth understanding of the Strategic Subject List from Chicago and develops frameworks to analyze other decision-making systems in the future. Future research into other decision support algorithms in the criminal justice system, especially a comparative analysis would help to further our understanding of the variation of these systems and their impacts on those who use them as well as those who are affected by them. It would also be deeply interesting to compare the role of these systems in the criminal justice system with similar systems in other fields, including health care. This could be fascinating as health care in the United States also has interesting dimensions of bias and discrimination, and thus it would be fascinating to analyze how these ideas manifest in different contexts. Moreover, a study across countries in analyzing the role of decision making systems in the criminal justice system could help to illuminate underlying axioms and biases in the various systems, as well as potentially expand on ways in which these algorithms could be used that could be largely beneficial to society. Finally, research in the future as to how decision support systems could best be used in the criminal justice system would help to create utility from these algorithms rather than causing problems.

Bibliography

Andonova, E. (2019). *Cycle of Misconduct: How Chicago has Repeatedly Failed to Police its Police*. NLG Review.

Bayamlıoğlu İbrahim Emre. 2018. "The Presumption of Innocence's Janus Head in Data-Driven Government." In Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen, Amsterdam: Amsterdam University Press B.V., 58–61.

Brantingham, P. Jeffery. "The Logic of Data Bias and Its Impact on Place-Based Predictive Policing." Ohio State Journal of Criminal Law 15(473): 473–86.

Burgess, Ernest. Papers, Box 50, Folder 10, Special Collections Research Center, University of Chicago Library.

"Cincinnati Crime Incidents." 2020.

"Chicago Community Area CDS Data."

"Chicago Crimes 2001 to Present." 2020.

Conti-Cook, Cythia, and Glenn Rodriguez. "FAT* 2019 Implications Tutorial: Parole Denied: One Man's Fight Against a COMPAS Risk Assessment." In Atlanta.

Dastin, Jeffrey. 2018. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women." Reuters.

Danziger, Shai, Jonathan Levav, and Liora Avnaim-Pesso. 2011. "Extraneous Factors in Judicial Decisions." *Proceedings of the National Academy of Sciences* 108(17): 6889–92.

Diega, Guido Noto La. 2018. "Against Algorithmic Decision-Making." *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*.

Ensign, Danielle et al. 2018. "Runaway Feedback Loops in Predictive Policing." *Proceedings of Machine Learning Research* 81: 1–12.

Eterno, John A. 2017. Crime Numbers Game Management by Manipulation. CRC Press.

"FBI 2018 Uniform Crime Report."

Ferguson, Andrew Guthrie. 2017. "Police Are Using Algorithms to Tell Them If You're a Threat." Time.

Ferguson, Andrew Guthrie. 2019. RISE OF BIG DATA POLICING: Surveillance, Race, and the Future of Law Enforcement. S.l.: NEW YORK UNIVERSITY PRESS.

Hirschi, Travis, and Michael R. Gottfredson. "Age and the Explanation of Crime." *The Craft of Criminology*, 2017, pp. 143–166.

Imposition of a Sentence. 1984. *U.S. Code*. Vol. 18, sec. 3553, p. 715

Kopf, Richard G. 2015. "Federal Supervised Release and Actuarial Data (Including Age, Race, and Gender)." Federal Sentencing Reporter 27(4): 207–15.

Maxson, Cheryl L., et al. "Differences Between Gang And Nongang Homicides." *Criminology*, vol. 23, no. 2, 1985, pp. 209–222.

Oleson, J.C. 2011. "Risk In Sentencing: Constitutionally Suspect Variables And Evidence-Based Sentencing." SMU Law Review 64(4): 1329–1404.

Posadas, Brianna. 2019. "How Strategic Is Chicago's 'Strategic Subjects List'? Upturn Investigates." Medium.

Richardson, Rashida, Jason M Schultz, and Kate Crawford. 2019. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice." New York University Law Review: 193–228.

"Strategic Subject List Chicago." 2017.

Mayson, Sandra G. 2018. "Bias In, Bias Out." *128 Yale Law Journal*: 2262–98.

United States Census Bureau (2018). *Annual Estimates of the Resident Population for Incorporated Places of 50,000 or More, Ranked by July 1, 2018 Population: April 1, 2010 to July 1, 2018*.