CITADEL

Datathon Championship 2023

From Cradle to Community to Courtroom: An Empirical Analysis of Solutions to and Factors Influencing Crime, Arrests, and Sentencing in Chicago and Sister Cities Across the United States

Team 19 | Tristan Brigham • Drake Du • Yuqing Jian • Matthew Rui

Executive Summary

Research Motivation

In her April 2019 campaigns, former Chicago Mayor Lori Lightfoot promised sweeping reforms of the Cook County criminal justice system. Yet, four years later, racial tensions are higher than ever in the city, and homicide rates have increased since she took office [1]. Why has this been the case?

We seek to become the first study to perform an in-depth, end-to-end analysis of some of Chicago's overlooked reforms, shedding light on the importance of institutional support and infrastructure to the success of such policy changes.

In particular, we evaluate these and other reforms using various quantitative methods to provide insights into their impact and effectiveness. In doing so, we take the first steps in the crucial journey to understanding how such reforms can **shape a more equitable and just criminal justice system, both in Chicago and across the United States.**

Key Findings

We employed random forest models, synthetic control methods (SCM), difference-in-difference (DID) modeling, various statistical significance tests, and regression analysis to derive 5 main insights.

Covert Policing Quotas Persist

Despite quotas being outlawed by most police departments in the United States, **we find statistically significant evidence of continued heightened arrest patterns toward the start and end of police reporting cycles, which we call the *serial-quota* effect**. While our analysis sheds light on the quantitative evidence supporting the existence of quotas and a serial-quota effect, future research can capture the nuances of this phenomenon further.

Retention Impacts Representation

We find strong evidence that Chicago's police force is not representative of the people they serve. **In particular, our analysis provides some of the first quantitative evidence for a difference in retention rates among underrepresented groups in the CPD, contributing to this disparity in representation.** Future research, in particular, can explore the nuances of this finding and quantitatively ascertain the largest influences on this disparity in retention rates.

Judges Have an Outsized Impact

We find that judges hold significant power in sentencing decisions, influence that, when affected by bias, has the capacity to ruin lives. Through machine learning analysis and D'Agostino-Pearson analysis of sentencing data, we uncovered the extent of judicial discretion and its susceptibility to bias. **Alarmingly, around 10% of Chicago's judges exhibit unusual plea deal patterns, signaling potential prejudice.**

Comprehensive Policies are Crucial

In assessing nationwide crime rates, we highlight the double-edged nature of LEAD diversion programs. **Without comprehensive policy support, such programs risk allowing low-level, non-violent offenders to fall through the cracks. However, when integrated within a robust policy framework, they demonstrate a potent capacity to curb crime.** While our analysis fails to find significant effects on economic growth and measures of judicial equity, future research can examine these longer-term impacts.

Policy Recommendations

Drawing from both policy research and our statistical analyses, we find that comprehensive policy solutions to criminal justice inequities are crucial. In particular, programs focused on stemming inequity in the criminal justice system have the potential to yield significant benefits, but when implemented alone they can have a minimal or even detrimental effect.

Given Chicago's status as a leader in the space of criminal justice reform efforts, the city inherently implements a wide range of policies to combat crime. **We seek to shed light on some such policies and inform steps going forward.**

Literature Review

From Prohibition to the War on Drugs, extensive research has been conducted on criminal justice policies and crime as a whole throughout the history of the United States. **We seek to expand the extant literature twofold by tackling several gaps within the space.**

First, while many of the causes and risk factors for crime — such as poverty, barriers to education, and unemployment — are well-documented and well-known, research on other, less straightforward factors are comparatively sparse. As such, in this report, **we focus on examining overlooked influences on crime in addition to arrests and sentencing.** Specifically, we will analyze the impact of covert arrest quotas, the disparate retention of police of color, and the inequitable demographic distribution of plea deals to increase a holistic understanding of how criminal behavior is manifested not only in the city of Chicago but also in similar, “sister” cities.

Second, while much of the current literature disjointedly examines causes and solutions to crime and sentencing outcomes, we seek to be one of the first studies to take our discussion and analysis from cradle to community to courtroom, and **investigate aspects from different stages of the criminal justice system in line with taking a holistic, end-to-end approach.** As one example, to assess the efficacy of arrest diversion programs such as LEAD, we consider the importance of requisite infrastructure on the ability of such programs to be successful.

Exploratory Data Analysis

To analyze the intricacies of the criminal justice system in Chicago, we must first have a strong understanding of broader trends in the space.

Naturally, we first explore the types of crimes that are being committed as well as where they are being committed to inform our holistic review. Looking at the crimes committed in Chicago from 2001 to 2023, we obtain the following barplot.

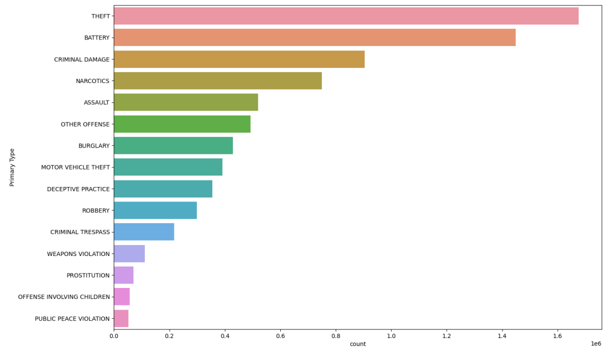

Figure 1: Distribution of Crime Types

We see that theft is the most common crime, followed by battery and criminal damage. Narcotics and assault also represent a significant proportion of crimes. Burglary, motor vehicle theft, deceptive practices, and robbery have moderate frequency while criminal trespassing, weapons violations, and prostitution are less common.

Next, we examine the frequency of crimes by location.

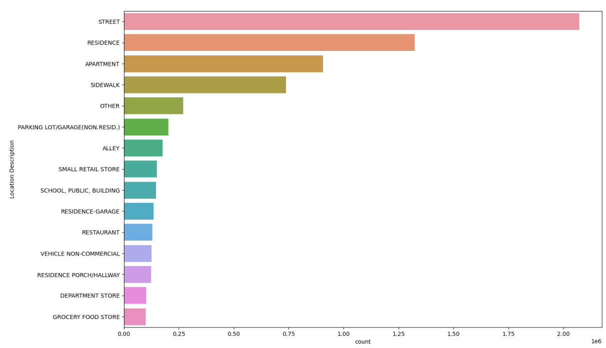

Figure 2: Crimes by Location Description

We find that the most common location for criminal activity is the street, followed by residences and apartments. Sidewalks and other unspecified locations also see a smaller, yet still significant number of incidents. Lastly, non-residential parking lots/garages, alleys, and small retail stores see relatively few instances of crime.

We now look at the trend in the frequency of crime in Chicago over time.

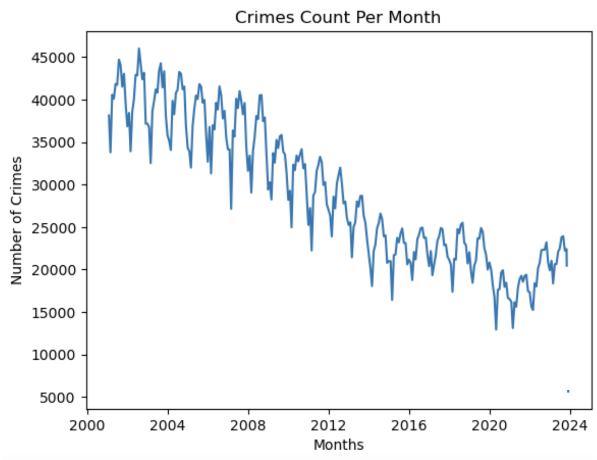

Figure 3: Crime Counts by Year

From this time-series, we see a **clear downward trend in crime as well as an inter-month cyclicality** over the course of each year, which can be explained by various studies on the seasonality of crime [4].

We then construct heatmaps for each year to visualize how the concentration and frequency of crime across Chicago change over time. We see that the **overall downward trend in crime seems to be evenly distributed across the entire city, with the exception of the downtown area**, a finding that aligns with anecdotal reports on the longstanding challenges of combating crime in this region.

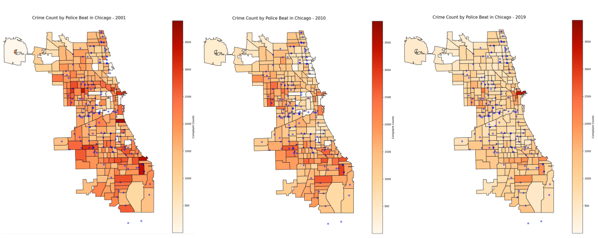

Figure 4: Crimes Count Per Month (Social Service Centers Overlaid as Dots)

With this understanding of crime in Chicago, to improve the rigor of our findings, we must contextualize any analysis against other similar "sister" cities. As such, we introduce data from the FBI's Universal Crime Reporting Program. Started in 1929, this database contains every criminal incident in reporting agencies' jurisdiction. As of the beginning of 2020, 48.9% of the U.S. population was covered by a UCR-reporting agency [5]. In situations where data is missing or incomplete, we were able to manually draw down data from sister cities' open data portals.

Using this extended dataset, we were able to construct the following time-series graph of crime trends over time across relevant cities.

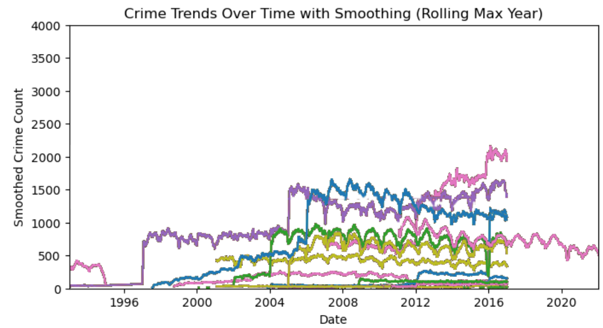

Figure 5: City-Wide Crime Rates Over Time for Cities with Requisite Mahalanobis Distances

Throughout our report and analysis, we only consider cities that have a low Mahalanobis distance to Chicago. The Mahalanobis distance is a measure of the distance between a point and a distribution, considering the correlation between variables and the spread of the distribution. This distance accounts for correlations and different variances in multiple dimensions, ensuring that compared cities are in similar environments and at similar stages in their respective economic, criminal, and social cycles.

Finally, to understand how patterns in crime may vary among sub-populations of Chicago, we graphed arrests by age groups from 2011 to 2022. We found that from 2012 through 2014, **there was a steep dropoff across the board in the number of people being punished for crimes**. The effect was especially heightened in young people – those less than 20 years old.

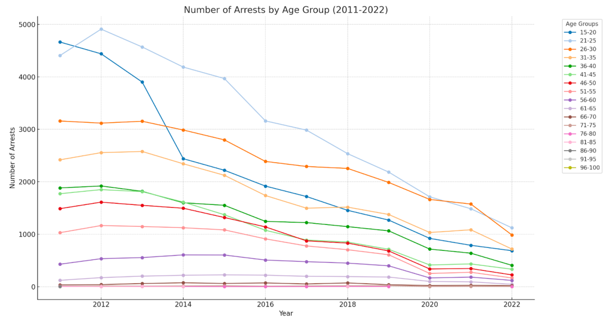

Figure 6: Distribution of People Sentenced by Age in Chicago

Honing in on groups that saw the largest decreases, we find that young people have experienced the sharpest absolute decline in criminal justice proceedings while consistently making up the majority of incidents. There exists a weak inverse relationship between the age of the person and the likelihood they experience the criminal justice system in a given year.

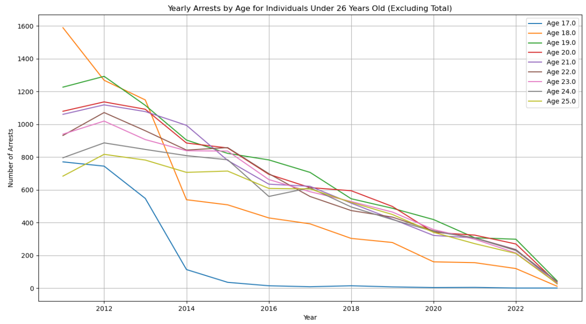

Figure 7: Distribution of People Sentenced by Age in Chicago (Selected)

The data exhibits remarkable predictability in the year-over-year drawdown in incidents, and begs the question of why there are two local minima at the young end of the distribution as well as around the age of 45.

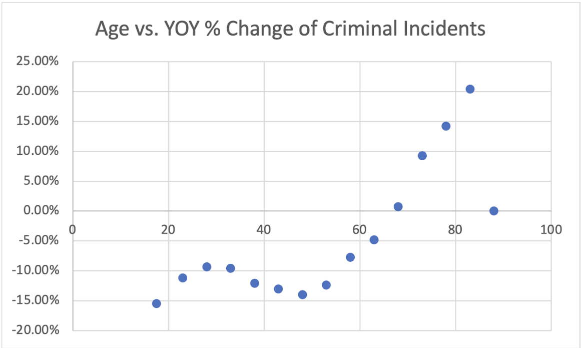

Figure 8: Average Year Over Year Change in Crime Incidents Across Provided Data

Relevant Insights

Based on our initial data analysis, we drew several key conclusions and hypotheses.

1. There has been a serious downward trend in the amount of incidents occurring across beats as well as age groups – especially incidents involving young people across the dataset. **Our exploratory data analysis therefore provides evidence confirming well-recognized academic research on how programs geared toward decreasing crime rates among youth in Chicago have largely succeeded.**

2. As shown by our analysis of cities with similar environments to Chicago, seemingly analogous implementations of policies have different impacts on crime rates. **We hypothesize that there may be exogenous factors at play influencing the success of such policies.**

3. In Chicago, areas with high concentrations of social services seem to experience the quickest and most effective drawdown in crime rates. **This signals the possible existence of an inverse relationship between crime rates and the amount and quality of social services.**

Inquiry Background

The existing literature is replete with studies on the importance of vocational opportunities and educational institutions in combating crime. What role, however, do law enforcement agencies themselves play as imperfect, and often, flawed policing institutions?

We seek to investigate one factor behind arrests that, thus far, has not seen rigorous analysis in the extant literature. Namely, we investigate how insidiously enforced quotas drive spurious arrests.

Perhaps most famously, the New York Police Department has faced years of allegations, including from police officers themselves, of a "quota" system that requires officers to log a certain number of arrests and summonses each month — or face retaliation [6]. Indeed, in 2017, after seven years of litigation, New York City agreed to pay a $75 million settlement to more than 900,000 people who received summonses that were later tossed on insufficient legal grounds [6].

Still, the quota system may linger. In particular, countless anecdotal accounts reveal that the productivity of cops is measured in monthly cycles that are compared against that of the previous year [7]. In this system, officers are pressured to reach numbers of arrests comparable to that of the prior corresponding cycle. As such, **we would expect that officers would give out more citations and carry out more arrests at the beginning of each cycle** when counts are reset, and at the end of each cycle, when pressure to meet productivity benchmarks comes to a head.

To investigate this, we analyze almost a decades-worth of data on arrests in Chicago. In particular, we use a Student's t-test in addition to a Mann-Kendall test with a Yue-Wang modification to examine such data. **We find significant differences in arrests at the beginning and end of each monthly cycle**, providing some of the first quantitative evidence for this U-shaped arrest distribution, which we call the serial-quota effect.

To see how this serial-quota effect may vary among cities, we then do the same for New York City, a sister city of Chicago that has also seen many anecdotal reports of arrest quotas. A Mahalanobis-distance test confirms the similarity between the cities with respect to crime rates, demographic breakdown, and economic activity.

Lastly, we evaluate the efficacy of quota-centered reform, detailing an illustrative case study on the impact one police commissioner's campaign has had on spurious arrests. Perhaps ironically, this example **serves to reinforce anecdotal research on the flaws of law enforcement agencies** and the difficulties of fundamentally reforming them.

Statistical Methodology

While Chicago has no official policy on arrest quotas, we investigated how arrest patterns in Chicago changed throughout the month. We hypothesized that if there existed a quota system, formal or informal, we should see a consistent spike in arrests at similar times every month.

Using Chicago arrest data from January 2014 to December 2022, we aggregated the number of arrests made on each day of each month. To account for different lengths of months, we normalized the data for the 29th (for leap years), 30th, and 31st days so that our arrest data would not dramatically decrease at the end of the month, by virtue of there being fewer months with 30 and 31 days than those with 28 or 29 days.

This gave us the following time-series graph.

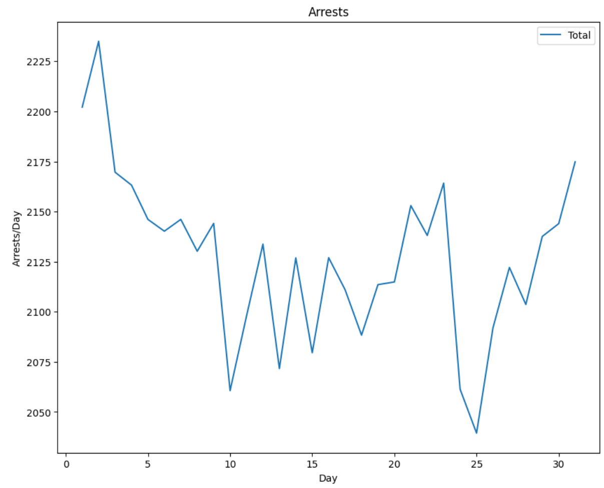

Figure 9: Distribution of Arrests by Day of Month in Chicago

From this, **we can clearly see a U-shaped distribution of arrests by day of month**.

To confirm our suspicion, we then grouped our data into two groups (arrests that occurred either before the 5th day of each month or after the 25th day of each month, and arrests that occurred between the 5th and 25th day of each month) and ran a Student's t-test on this data. We also conducted a Levene Test, which calculates the equality of variances, to determine whether the Student's t-test's assumption of no heteroskedasticity is met.

To further confirm our findings, we utilized a Mann-Kendall test to detect trends in this arrest data, as such tests can specifically detect increasing/decreasing trends that need not be linear in time series data [8]. Most crucially, unlike our previous Student's t-test, Mann-Kendall tests do not assume that the data is normally distributed. For our study, we implemented a Mann-Kendall test with a Yue-Wang modification, which further implements a variance correction method to account for serial autocorrelation, which we would expect to see in our data given the vast anecdotal reports on contiguous periods of heightened arrest frequency [9].

T-Test and Mann-Kendall Results

After conducting a Levene test on the two sets of data, we failed to reject our null hypothesis that heteroskedasticity does not exist ($p = 0.48$). With this key assumption met, we ran our Student's t-test and obtained a p-value of 0.01, indicating **very strong statistical significance of our findings**. Additionally, when we ran a Yue-Wang augmented Mann-Kendall test on the Chicago arrest data, the model found that **our hypothesized trend of a U-shaped distribution was statistically significant** with $p = 0.01$ as well, confirming our robust results.

From this, we draw some of the first quantitative evidence indicating that a serial-quota effect exists, supporting anecdotal research into the matter.

Statistical Methodology

To further understand how quotas can impact arrest distributions, next we studied sister cities of Chicago and their arrest trends. In particular, we examined New York arrest data from 2006 to 2022. After aggregating data and adjusting for data on the 29th, 30th, and 31st day of each month, we found that **arrests similarly start off very high at the beginning of each month before dramatically falling off toward the end of each month.**

In particular, we found the following time-series graph.

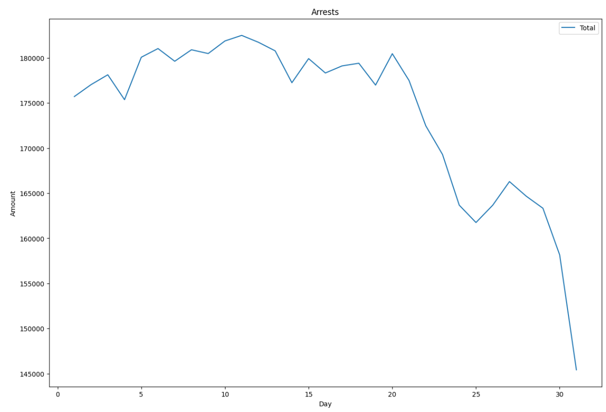

Figure 10: Distribution of Arrests by Day of Month in New York City

Interestingly, New York City does not seem to see a spike in arrests at the end of the month. This can be explained by vast anecdotal reports of the New York Police Department's informal policy of pushing heavily toward arrest quotas primarily at the beginning of each month [5].

To find the significance of our results, we again separated arrest data into two groups (arrests that occurred before the 20th and after the 20th of each month) and ran a Student's t-test and a Yue-Wang augmented Mann-Kendall test on our data after again verifying requisite conditions through a Levene test on heteroskedasticity.

T-Test and Mann-Kendall Results

After conducting a Levene test, we failed to reject the null hypothesis that heteroskedasticity doesn't exist ($p = 0.46$). With this key assumption of the Student's t-test met, **we found the rapid decline in arrests at the end of each month to be statistically significant** with $p < 0.01$. Additionally, applying the same Mann-Kendall model with the Yue-Wang modification as before, we again were able to conclude the existence of this trend with $p = 0.005$.

This significant drop in arrests aligns with anecdotal research and **provides strong evidence that a quota is in place**, with officers pushing heavily toward productivity benchmarks at the beginning of each month before tapering off once this has been largely achieved.

As such, while Chicago has yet to pay any $75 million settlement over arrest quotas, **our models were able to readily detect the effect of New York City's quota on its arrest data**, and, in finding similar statistically significant results for Chicago, **we posit that the city may have analogous policies that incentivize high arrests at the beginning and end of each month.**

From this, we provide some of the first quantitative support indicating that a serial-quota effect exists in Chicago in addition to showing that the exact impact quotas have on arrest distributions may vary from city to city, highlighting the importance of anecdotal and first-hand research to understand the scope of the issue.

Case Study

Now that we have evidence pointing toward the existence of unjust and corrupt quotas, what can be done about them?

In 2016, incoming New York Police Department Police Commissioner James O'Neill vowed to end the practice of arrest quotas, pledging "wrath" on any perpetrators [7]. How effective, though, has his particular campaign been since its inception?

To visualize any changes in arrest distributions, we separated our data into pre- and post-2016 arrest numbers to more clearly ascertain the influence of O'Neill on arrests.

Then, to account for the overall decline in raw arrest numbers after 2016, we standardize our parameter of interest to be percent above/below the given monthly average.

This gave us the following time-series graph.

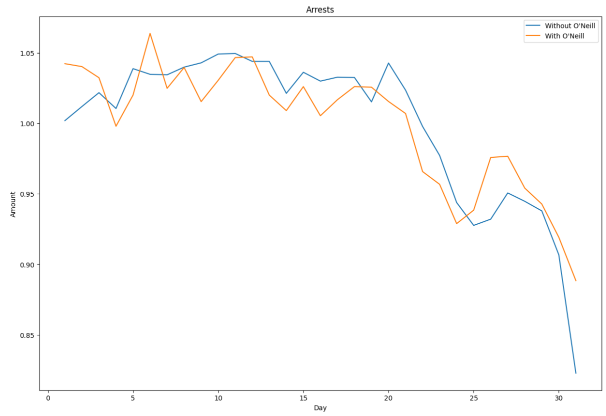

Figure 11: Distribution of Arrests by Day of Month in New York City Before and After Neill's Ascension

We can see that **the distribution of arrests does not seem to have changed significantly** from before O'Neill became police commissioner to after. Indeed, we find that the correlation between the two distributions is markedly high at 0.90.

These results tell us that any impact O'Neill had, if at all, does not seem to have significantly altered the way in which officers satisfy quotas. That the New York Police Department Police Commissioner himself could not influence significant change is **evidence itself of the flaws of law enforcement agencies and the difficulties of fundamentally reforming them.**

Ultimately, while our analysis sheds light on the quantitative evidence supporting the existence of quotas and a serial-quota effect, future research can further capture this phenomenon's nuances.

Inquiry Background

The productivity benchmark pressures law enforcement agencies face as a whole can lead to misaligned incentives. How, though, do institutional problems manifest among the police force itself?

One way is through disparities in retention. As described by the U.S. Equal Employment Opportunity Commission, law enforcement agencies struggle to attract individuals from underrepresented communities, let alone retain them [9].

This is a significant problem as decades of research confirm that **it is only when members of the public believe their law enforcement organizations represent them, understand them, and respond to them that we see the trust necessary for defusing tension and solving crimes.** Disparities in retention among underrepresented communities therefore erode this representation, rendering law enforcement less responsive to the residents they serve.

Despite these harms, the current literature largely lacks statistical analysis on the disparities in retention among races. As such, we seek to provide preliminary insight and quantitative evidence for this trend. Specifically, **we first verify the existence of a mismatch between the demographics of police and those they police** using simple proportions. Next, we leverage the Kaplan-Meier survival analysis method to show that **rates of retention are perennially lower for police of color than for their white counterparts.**

Statistical Methodology

To determine how representative Chicago police are with respect to the communities they serve, we compiled demographic data from the Chicago Police Department (CPD) and the city of Chicago as a whole. The CPD data included the race and gender of officers, while the city's demographic data included the distribution of races within the population.

For each of these sets of data, we then computed the demographic distribution of the data. Lastly, we compared these distributions to search for discrepancies in the representation of the general population among police forces.

Simple Proportions Result

Our work yielded the following pie charts.

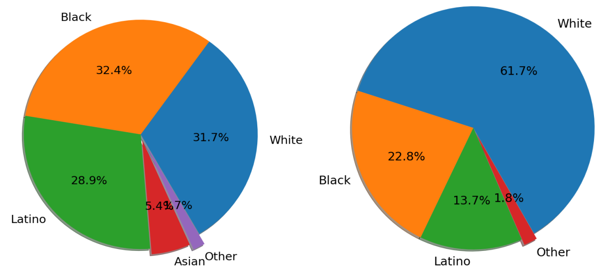

Figure 12: Percentage Breakdown of Race in Chicago (Left) and the Chicago Police Department (Right)

For the city of Chicago, Black and White individuals are nearly equal in proportion, with Black individuals slightly surpassing White individuals by 0.7%. However, within the CPD, **White officers are significantly overrepresented compared to Black officers**, with a difference of 38.9 percentage points.

Additionally, for the city of Chicago, Latino and White individuals are nearly equal in proportion, with White individuals slightly surpassing Latino individuals by 2.8%. However, within the CPD, **White officers are significantly overrepresented compared to Latino officers**, with a difference of 48.0 percentage points.

Both of these disparities verify a lack of representation within the CPD as compared to the city's demographics, where **White officers are about 1.95 times more prevalent in the department than in the city population, while Black and Latino officers are both underrepresented** in the CPD by nearly a half when compared to their representation in the city's population.

Statistical Methodology

Having verified the disparities in demographic representation between the CPD and the city as a whole, we now consider why this disparity exists.

Much of the extant research on this topic focuses on the factors that discourage underrepresented groups from joining the police force to begin with.

But what about those who are already in the police force? We seek to be one of the first studies to rigorously examine how disparities in retention rates among different demographics contribute to the overall difference in the makeup of the CPD as compared to Chicago as a whole.

We began by engineering a comprehensive dataset from the records of the CPD that includes both the demographics and employment durations for officers of various races and genders.

We then performed extensive data cleaning to ensure accuracy. This process involved removing duplicate records and correcting inconsistencies as well as handling missing and incomplete data entries.

Next, to analyze retention rates, we utilized the Kaplan-Meier survival analysis method, a non-parametric statistic used to estimate the survival function from lifetime data.

We defined "survival" in the context of an officer's tenure with the CPD. Thus, the "event" of interest in our analysis was the departure of an officer from the police force, whether due to retirement, resignation, or other reasons.

Lastly, we stratified these survival curves by race and gender to compare the retention rates across different demographic groups.

Kaplan-Meier Curve Result

Our analysis yielded two graphs.

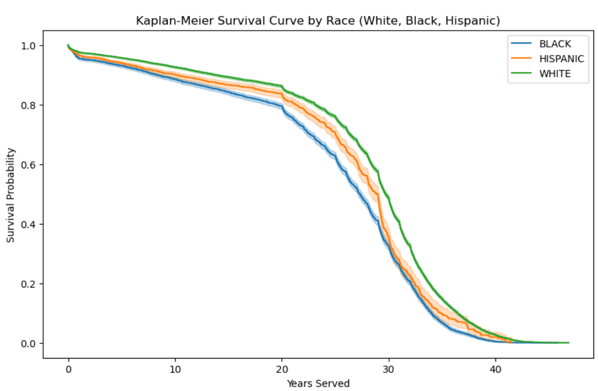

Figure 13: Kaplan-Meier Survival Curve by Race

In this graph, we can see that, unsurprisingly, **the survival curves for officers of color declined faster than that for white officers, providing evidence of a lower retention rate.** We can also see that, across all years served, this retention rate is lower for Black and Latino officers than for their white counterparts.

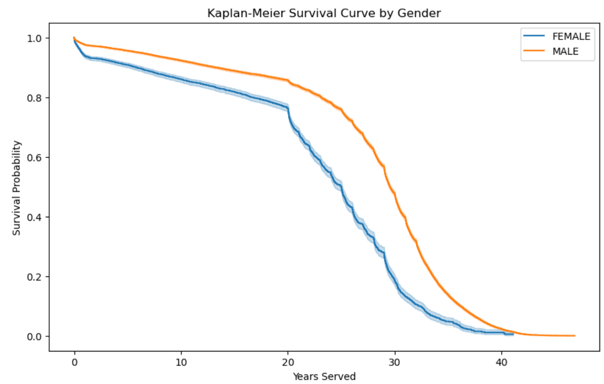

Figure 14: Kaplan-Meier Survival Curve by Gender

Similarly, in this graph, we can see that **the survival curve for female officers declined faster than that for male officers, providing evidence of a lower retention rate.** We can also see that, across all years served, this retention rate is lower for female officers than for their male counterparts, especially at 20 – 30 years of service.

Our analysis therefore provides some of the first **quantitative evidence for a difference in retention rates among underrepresented groups in the CPD**, contributing to a police force unrepresentative of the city through its policies.

Future research, in particular, can explore the nuances of this finding and quantitatively ascertain the largest influences on this disparity in retention rates.

Inquiry Background

In Chicago and its sister cities alike, those who make arrests often do not represent the communities they police. What happens, though, after an arrest has already occurred? More specifically, what factors influence the length of sentencing, how does this vary from judge to judge, and how has this changed over time?

The current literature is rife with quantitative evidence extending back decades on the existence of a racial disparity in criminal sentencing [11]. However, few studies have sought to analyze the strength of this phenomenon over time. Additionally, while this racial disparity in criminal sentencing is widely recognized, the current understanding of it is far from complete as “inequity may arise from several sources” [12].

To investigate these unknowns, we first seek to employ statistical analysis to **demonstrate the importance of race as a factor in criminal sentencing over time**. In particular, we utilize random forest classifiers to identify influential features in criminal sentencing and map how these change over time. Then, we aim to leverage rigorous methods to help reveal one source of inequities in sentencing, namely, the disparate demographic distribution of plea deals. In particular, we leverage a D'Agostino-Pearson test to show how this inequity can arise from individual judges themselves.

Statistical Methodology

As described anecdotally, race seems to play a role in both the length of a sentence and the specific charge associated with it. Correspondingly, this should suggest that race can serve as a predictor for a given offender’s total sentence time as well as a given offender’s type of charge.

To test this, we utilize a random forest classifier – a machine learning algorithm often utilized to determine important features of data. In particular, random forest models combine decision trees from random samples of the training set to assign importance values and make predictions.

For our random forest classifier, we first quantified the combined sentence time by adding the product of discrete commitment units and commitment terms for each offender.

Next, based on anecdotal research, we selected the relevant features of title of charge, whether this charge is the primary change in an incident, age at incident, the judge overseeing the sentence, race of offender, gender of offender, and the number of prior charges.

We then encoded categorical variables, handled missing values, and split the data set into training data and testing data.

Lastly, we utilized a random forest classifier to identify features most predictive of sentencing outcomes for each year over the past 2 decades to see how the importance of these factors changed over time.

In order to run the random forest regression, we ensured that our data satisfied 3 conditions: negligible multicollinearity, negligible heteroskedasticity, and negligible autocorrelation. To address multicollinearity, we performed a simple VIF calculation and removed data columns with high values from our input data. Next, we conducted a Levene test and failed to reject the null hypothesis that there is no heteroskedasticity in our data with a p-value of 0.4. Lastly, we employed a simple Durbin-Watson test and received 2.203, indicating that our data display low negative autocorrelation.

Random Forest Results

Our random forest classifier yielded the following insights.

Race Impacts Sentencing Length

We were able to construct a random forest classifier that could predict the length of prison sentences with an accuracy of 79.32%.

We found that **the particular judge assigned to a case has the single greatest impact on sentencing** length with a predictive power of 38%. We will return to this factor later in our report to better understand this subjectivity of sentencing and how it can manifest as unequal outcomes for marginalized communities.

Age played the second-largest role in sentencing with a predictive power of 28%. This follows directly from policies themselves that depend on the age of the offender as well as proxies for age, such as the number of prior offenses, a factor that also has an impact on sentencing length.

We also found that the final disposition or outcome of the charge had a predictive power of 24%. This makes sense as a given conviction would give us an associated range of sentence lengths.

We then see a steep drop off in predictive power with the next most impactful predictor being charge count with a predictive power of 4.31%. Intuitively, this makes sense due to the existence of ratchet policies that increase sentence length based on relevant prior convictions.

Next, **we see that race has a predictive power of 3.82%,** confirming our anecdotal research on the existence of an impact of race on sentencing length. We also see this to a lesser extent with gender at 0.9%, and primary charge at 0.6%. These make sense given gender-based bias in the judicial system in addition to the fact that a primary charge can predict a large part of sentencing length conditional on the event that a primary charge becomes a conviction.

```
Feature Importances in Descending Order:
SENTENCE_JUDGE: 0.38044953975875034
AGE_AT_INCIDENT: 0.28065162796561277
DISPOSITION_CHARGED_OFFENSE_TITLE: 0.241824388285535
CHARGE_COUNT: 0.043064225754725446
RACE: 0.038189230317956555
GENDER: 0.009387584229906107
PRIMARY_CHARGE_FLAG: 0.0064334036875138174
```

Figure 7: Factors Predicting Sentencing Length

Thus, our research validates what the extant literature has found: namely, that race can play a role in criminal sentencing.

Race Affects Ultimate Disposition

Next, utilizing our same data set, we constructed a random forest classifier that predicted the final disposition of a given offender with an accuracy above 90%.

We again found the particular judge to be the greatest factor influencing final disposition with a predictive power of 31%.

We also find that charges themselves can predict the final disposition with a predictive power of 26%. This is not surprising since, if a charge becomes a conviction, this thereby tells us the final disposition. Interestingly, however, we note the remarkable result that the judge has a larger impact on the final disposition than the initial charge itself.

Age playing a role in charge type with a predictive power of 24% is also unsurprising as certain policies grant lesser degrees to youth.

Interestingly, we see that charge count has a higher predictive power at 11.8%. This follows from the fact that ratcheting charges means that the number of charges influences the severity of the conviction, impacting the final disposition itself. It makes sense that charge count has higher predictive power for disposition than sentencing length, too, as sentencing lengths may vary for a given conviction but disposition does not vary for a given conviction.

We see that primary charge is next with a predictive power of 3.4%. For the same reasoning as with charge count, since sentencing lengths may vary for a given conviction but disposition does not vary given a conviction, it is not surprising that primary charge has more predictive power for disposition than for sentence length.

Lastly, we again see that **race has a predictive power of 3% and that gender has a small predictive power of 1%,** verifying the anecdotal reports we discussed above.

```
Feature Importances in Descending Order:
SENTENCE_JUDGE: 0.3074710437237113
DISPOSITION_CHARGED_OFFENSE_TITLE: 0.25698008682422563
AGE_AT_INCIDENT: 0.24248215130930512
CHARGE_COUNT: 0.1179421081314661
PRIMARY_CHARGE_FLAG: 0.034147348488511704
RACE: 0.03097022969299434
GENDER: 0.010007031829785778
```

Figure 7: Factors Predicting Final Disposition

As such, **we have provided quantitative evidence in support of the widely recognized impact of race on the charge side of criminal sentencing**.

Race as a Predictor Has Weakened

Lastly, by leveraging our random forest classifier on data for each year for the past 22 years, we can see that **the importance of race as a factor influencing criminal sentencing has declined over time.**

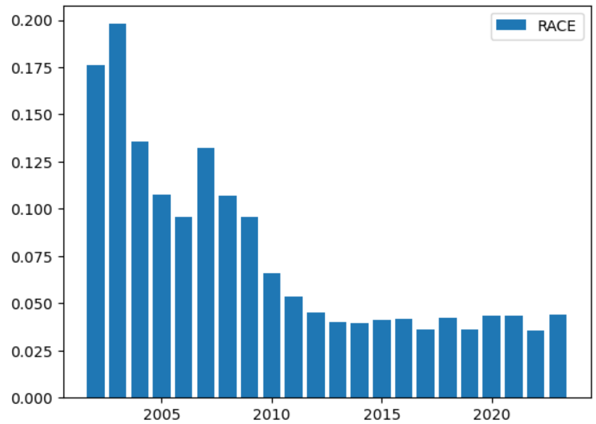

Figure 15: Influence of Race over Time

Through our random forest classification, we have thus confirmed research on the influence of race on criminal sentencing, while also contributing new evidence showing that this effect has been decreasing over time.

Statistical Methodology

Having identified judges as having the single largest impact on the outcome of a case, we now return to this phenomenon by examining judges and their sentencing practices in Chicago to better understand the subjectivity of sentencing, and how this can manifest as unequal outcomes for marginalized communities.

Notably, **the vast majority of criminal trials are resolved with plea deals**. In fact, from January 1992 to August 2023, 94.6% of criminal trials in Chicago were resolved by the defense accepting a deal offered by the prosecution. This begs the question, however, of **whether the judge assigned to a case affects the plea deal outcomes of individual cases.**

To determine this, we first filtered judges to only examine those who had overseen at least 500 cases. From here, we calculated and examined the proportion of their cases that ended in a plea deal.

For each judge, we also further examined the racial breakdown of the defendants who took plea deals and the racial makeup of the defendants who were sentenced at trial. To analyze our results for any statistically significant insights, we implemented a D'Agostino-Pearson test, which measures the skewness and kurtosis of a dataset to determine if it follows a normal distribution.

Simple Proportion and D'Agostino-Pearson Results

After calculating the proportion of defendants accepting plea deals per judge, we were able to plot the frequency of each judge's plea deal rate to get the following graph.

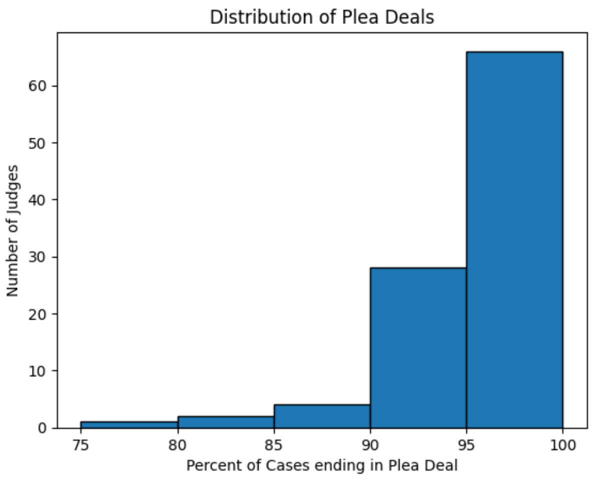

Figure 16: Judges by Proportion of Cases Overseen Ending in Plea Deals

This distribution makes sense as judges in Chicago are allocated to cases based on their municipal district, with each municipal district handling different types of crimes. As such, we would expect to see differences in density among the rates of plea deals, as each judge is likely to only oversee certain types of cases, each with their own average likelihood of ending in plea agreements.

For each judge, we then examined the groups of defendants who took a plea deal and those who were sentenced by trial. We determined the percentage of white defendants in each group for every judge and plotted the difference below, with positive values indicating that the percentage of white individuals in the plea deal group was higher than the percentage of white individuals in the trial group.

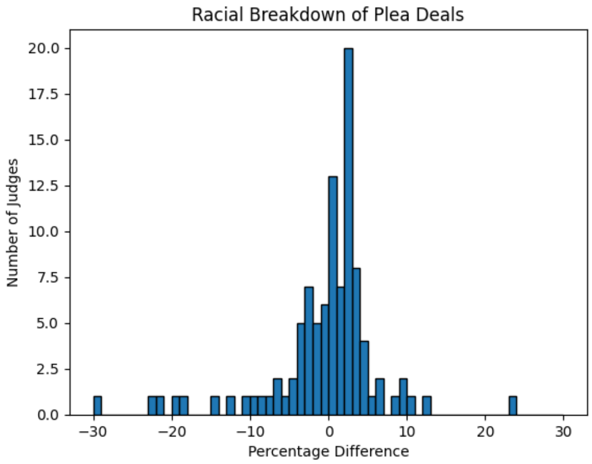

Figure 17: Judges by Difference in Racial Breakdown between Plea Cases and Trial Cases

As expected, the vast majority of judges had a roughly equal racial breakdown between their plea deal and trial groups. **However, 12 of the 102 judges had substantial discrepancies** (greater than 10%) **between the racial breakdown of these two groups**, which given that most of these judges had overseen thousands of cases, is unlikely to all be explained by simply variance.

To confirm this, we utilized a D'Agostino-Pearson test, which confirmed our suspicion that this racial discrepancy is not normally distributed, implying that the large discrepancy of the 12 judges is likely not due to variance with p = 0.01.

This demonstrates that **judge selection has a clear impact on plea deals, specifically the racial breakdown of the groups that plea or don't plea,** further reinforcing our prior feature importance results from the random forest modeling.

Inquiry Background

Racial disparities in sentencing make it resoundingly clear that our criminal justice system is far from perfect. Coupled with the fact that time behind bars can actually increase the likelihood that certain individuals will re-offend [13], how, then, can we prevent non-violent drug offenders from falling further into a cycle of addiction and recidivism?

One solution is to rethink the War on Drugs entirely. After all, by most measures, the largely ill-informed draconian campaign has failed, with more than a million Americans having passed away from overdoses since the government started counting in 1999 [14].

By contrast, after Portugal removed criminal penalties for possessing some drugs in 2001 and instead shepherded individuals into treatment as well as had them appear before dissuasion commissions, overdose deaths and HIV rates fell in subsequent years, and public drug markets disappeared [15].

The outstanding caveat, however, was the existence of accessible and affordable addiction services in Portugal. Indeed, when Oregon, a state with just half the addiction services it needs, decriminalized certain drugs in 2020, overdose deaths in Oregon rose sharply, far outpacing the increase recorded nationally [15]. This makes sense as policies cannot help those in need when the necessary infrastructure isn't in place.

As such, **we seek to analyze the efficacy of one often overlooked solution to the drug epidemic — law enforcement assisted diversion (LEAD) — specifically for cities with robust support.**

LEAD programs redirect low-level, non-violent offenders to community-based support services and treatment programs as opposed to jail and prosecution, providing individuals with the tools they need to reintegrate into society [16].

Indeed, for many individuals, **LEAD programs have been credited for "turning their life around" and giving them a chance at support, sobriety, and stability** [17]. Yet, while countless anecdotal accounts affirm the benefits of such programs, the current literature has yet to see substantial statistical analysis of them, in part due to their relative infancy.

We present some of the first rigorous evidence demonstrating the efficacy of LEAD programs in cities with requisite services. In particular, in employing difference-in-differences (DID) models with the synthetic control method (SCM), **we see that both violent and non-violent crime tend to decrease in cities that implement LEAD programs and have the necessary infrastructure, supporting prior anecdotal findings.**

On the other hand, however, **we observe that when such programs are implemented without the requisite support and without coincident policy measures, they do not have significant benefits and may even induce outsized harms.**

As such, we remain cautiously optimistic about the potential for Chicago's 2021 LEAD program to see positive results in the coming years. While such programs in general have a track record of success, comprehensive implementation is key, and only time can tell whether the city's efforts will come to fruition.

Statistical Methodology

To understand the impacts of LEAD programs, we leveraged DID models using SCM to simulate control cities given our small sample size and potential exogenous factors affecting variance in individual comparison cities' features. Specifically, we sought to measure the effect that LEAD programs had on rates of crime, GDP growth, and judicial equity scores — our variables of interest.

The average control technique utilized by SCM limited the need for individualized placebo modeling, which helped us understand and evaluate whether the model adhered to the key parallel trends assumption.

Before creating our synthetic controls, we performed the following steps to ensure that our model was appropriate.

1. AIC value optimization: This reduced the complexity of the model to improve explainability.
2. White's test: This ensured that the data considered didn't exhibit heteroskedasticity. Across all cases in this

report, we failed to reject the null hypothesis that the data doesn't exhibit heteroskedasticity.

3. Breusche-Pagan test: In cases where we rejected the null hypothesis with a p-value under 0.2, we ran this to confirm our results. We used White's test first as it is better adapted for variance across multiple variables.
4. Durbin-Watson test: The measured auto-correlation in time-series data residuals. We found that there was not a statistically significant autocorrelation component in the residuals.
5. Multicollinearity test: We tested this by constructing a simple correlation matrix. Given the nature of economic and criminal justice data, there are some examples of inherently elevated correlations between the variables. While we could have set up our DID model using principal component analysis, to ensure explainability, we left this as an open question and noted multicollinearity.
6. Propensity score matching: Since this is effectively a more generalized version of the DID model, we used this as a sanity check for our DID model as we assume that there is no covariate imbalance.

In order to construct synthetic controls for our DID model, we designated the city for which we were looking to assess the impact of a LEAD program implementation as our treated city. Next, we calculated Mahalanobis distances to find close-substitute, sister cities of the treated city and constructed a control city using a weighted average of these substitute cities with respect to their Mahalanobis distance. In order to ensure the highest confidence in our analysis as possible, we used the Mahalanobis distance instead of the general least squares error statistic to account for the correlation of data. This mitigated the effect of the aforementioned multicollinearity introduced by wide-reaching social and macroeconomic reform. We experimented with allowing for substitute cities originating from years other than the years considered for the treated city by including a penalty term in the Mahalanobis distance to incorporate the difference in general economic trends, but we found that this did not significantly affect results.

It is important to note that due to the availability of data on GDP growth and judicial equity scores, our ability to train models was limited. It is also important to recognize that in general DID models struggle with impacts that may include longer time horizons, such as the impact of fewer non-violent incarcerated individuals on GDP growth. We assume no hidden variables for the purpose of testing, and our results support this conclusion.

Difference-in-Differences Results

Utilizing the above adaptations to our specific study, we constructed our DID model. A few illustrative cases and general commonalities are highlighted.

Growth Unaffected in Short Term

Looking at economic impact, we find that the implementation of LEAD programs has little to no effect on short-term GDP growth across all cities considered.

We can look to Albany for a prime example of this result. After the city introduced a LEAD program, we did not see statistically significant changes to short-term GDP growth compared to that of Albany's synthetic control.

In particular, we obtained the following time-series graph using our customized DID model.

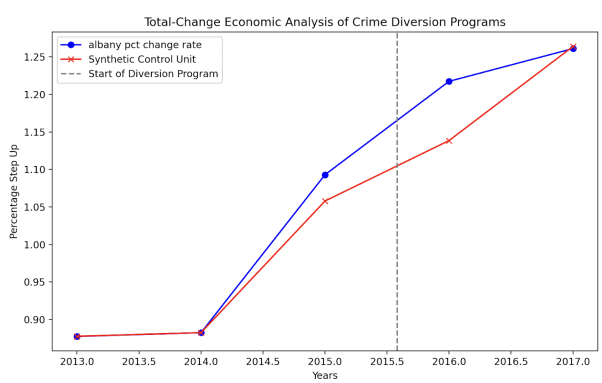

Figure 18: Albany Percentage GDP Growth over Time

We can see that after implementation of a LEAD program, t**he difference in GDP growth as compared to the synthetic control is marginal**. As such, we fail to reject our null hypothesis that such programs have no meaningful effect on short-term GDP growth.

This result is unsurprising as we previously discussed how the DID model struggles with impacts that may include longer time horizons, such as the impact of fewer non-violent incarcerated individuals on GDP growth. Indeed, in this case, **we were not able to detect any discernible effects in the short term**. Future research, however, can attempt to identify potential longer-term ramifications.

Racial Inequities Persist

With regard to ameliorating racial inequities, we found that the LEAD programs are inconsistent across our DID models.

To explain this finding, we investigated the demographics of those granted support under LEAD programs.

Unsurprisingly, we found that Black and Hispanic individuals were significantly less likely to receive support under LEAD programs as compared to their white counterparts.

Specifically, we obtained the following side-by-side barplot through our statistical analysis.

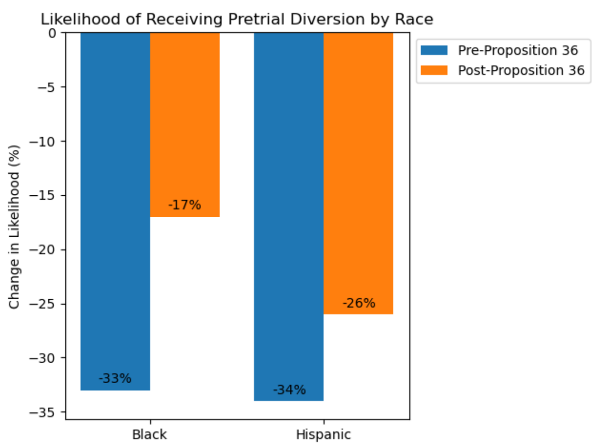

Figure 19: Likelihood of Receiving LEAD Program Support by Race

Thus, **even when LEAD programs are implemented, there is no guarantee that we see improvements in judicial equity as law enforcement assisted diversion programs by definition are administered by police officers, for which we can still witness observable bias.**

Indeed, our results confirmed separate research by the Prison Policy Initiative on how Black and Hispanic individuals are still less likely to receive a pretrial diversion than their white counterparts even after the introduction of established diversion programs, suggesting that inequities remain entrenched even after broad reforms are made [18].

Crime Sees Significant Decrease

Looking at the impact on crime, **we find that the implementation of LEAD programs decreased non-violent crime by a significant margin and violent crime by a small margin, leading to an overall deflation in crime rates across cities analyzed given requisite infrastructure.**

As one example of this result, we can look toward Sante Fe. After the city introduced a LEAD program in 2014 following New Mexico's statewide reform of its publicly funded behavioral health care system, we saw a statistically significant decrease in crime by 2.7% (p=0.041) [19].

In particular, we obtained the following time-series graph using our customized DID model.

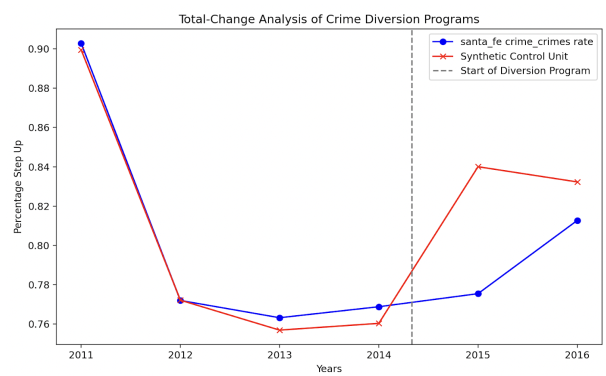

Figure 20: Santa Fe Percentage Crime Change over Time

As another example, after Denver, ranked among the places with the highest living standards in America by U.S. News & World Report and widely believed to have one of the most effective LEAD programs across the country [20], implemented such a program in 2017, the city saw a reduction in the number of violent crimes committed by over 45 percent (p=0.01).

In particular, we obtained the following time-series graph using our customized DID model.

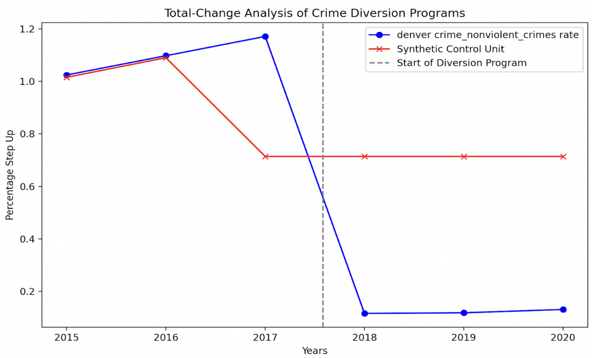

Figure 21: Denver Percentage Non-Violent Crime Change over Time

For these examples, Santa Fe's synthetic control is a weighted combination of San Diego, Milwaukee, and Cleveland; Denver is represented by a weighted combination of Albany, San Diego, and New York City.

From this, we reject our null hypothesis that LEAD programs have no meaningful effect on crime. In particular, we have evidence that LEAD programs can decrease crime within a city given the necessary support. Thus, our findings provide some of the first quantitative research validating vast anecdotal research on how redirecting low-level, non-violent offenders to community-based support services and treatment programs, as opposed to jail and prosecution, can help them rehabilitate and reintegrate into society.

Success Relies on Support

To test the other side of our theory, we expect that a LEAD program implemented without the right infrastructure would have no impact or potentially even a negative impact on crime.

As one example, we can look at the city of Seattle. When the city started its LEAD program back in 2011, Seattle ranked poorly for living standards with high inequality and inadequate social services, specifically in the domain of mental and behavioral health [22].

Unsurprisingly, after the city introduced a LEAD program, we saw a statistically significant jump in non-violent crime of 79.9% (p=0.017).

In particular, we obtained the following time-series graph using our customized DID model.

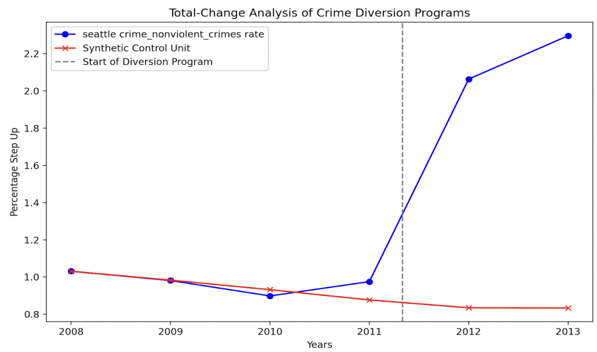

Figure 23: Seattle Percentage Non-Violent Crime Change over Time

Additionally, after the city introduced a LEAD program, we saw a statistically significant spike in overall crime of 80% (p=0.016).

Specifically, we obtained the following time-series graph using our customized DID model.

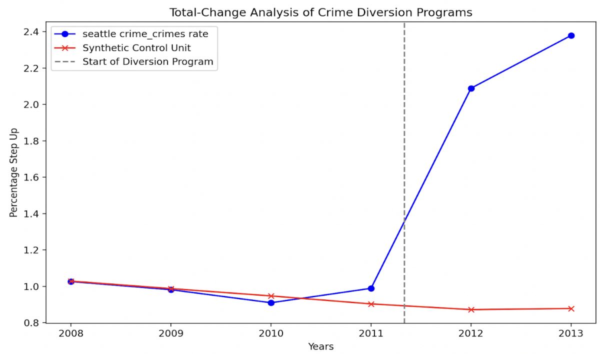

Figure 24: Seattle Percentage Crime Change over Time

Only for violent crime did we not see a statistically significant effect (p=0.063). In particular, we obtained the following time-series graph using our customized DID model.

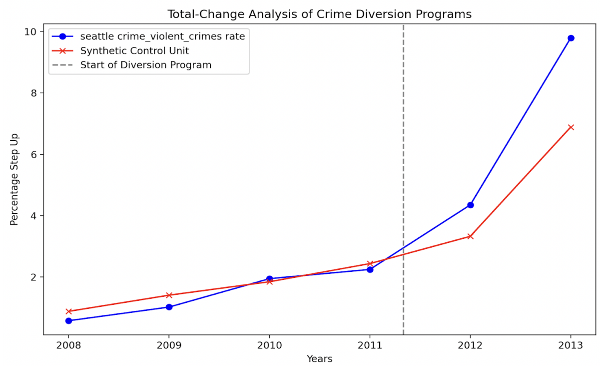

Figure 25: Seattle Percentage Violent Crime Change over Time

As our last example and focus of our report, following Chicago's 2012 violence reduction initiatives (a predecessor to its later 2021 LEAD program) and subsequent statewide drug abuse reform [21] we found that Chicago did experience the significant reduction in crime associated with a supported diversion program implementation. This makes sense – Chicago implemented the LEAD program without serious policy support, and therefore was not likely to experience the crime drawdowns.

In particular, we obtained the following time-series graph using our customized DID model.

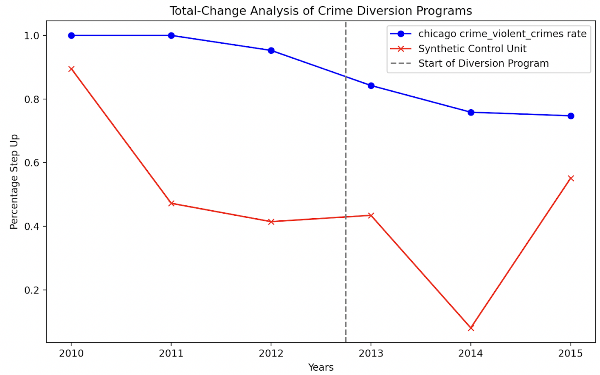

Figure 22: Chicago Percentage Violent Crime Change over Time

We note that Chicago's synthetic control is a weighted combination of New York, Indianapolis, and San Diego, and the synthetic control for Seattle is best represented using a combination of data from New York, Milwaukee, and Chicago.

Our results thus verify both anecdotal reports and existing research detailing the inefficacy of policy measures taken without requisite infrastructure and without directly interfacing with the communities they were meant to serve.

Chicago Has Yet to Crystallize

With this knowledge, we finally turn to Chicago's 2021 LEAD program. We see that, while an uptick in crime has been seen, a clear trend is not yet discernible, owing to the program's relative infancy.

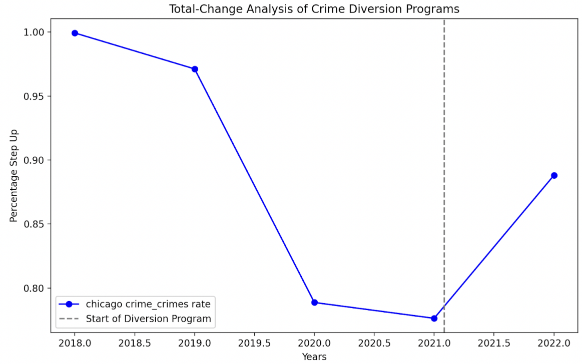

Figure 26: Initial Results from Chicago's LEAD Program

Ultimately, we maintain a guarded sense of optimism regarding the prospects of Chicago's LEAD program to yield positive outcomes in the years ahead. While initial results on Chicago's crime rates indicate a potential uptick in crime rates following this policy, given the compressed time horizon of collected data, we cannot yet draw any conclusions. Indeed, although similar programs in sister cities have demonstrated success, and though we provide some of the first quantitative evidence backing these anecdotal reports, careful implementation is crucial, and only time will reveal whether Chicago's endeavors will translate into tangible results.

Key Findings

We started by performing statistical modeling to understand if and how quotas exist in police departments. **Our findings provide some of the first statistically rigorous analyses to corroborate the continued existence of arrest quotas in police agencies.** Specifically, we validate anecdotal claims that, despite extensive legislative and personal efforts to eliminate the practice, implicit arrest quotas continue to exist and propagate within police forces not only in Chicago, but other similar metropolitan areas.

Next, we executed demographics-focused tests on the demographic breakdown of Chicago's police to validate whether it is representative of Chicago's broader population. **We found that Chicago's police force has implicitly instituted perverse incentives to steer the force away from representing the people that they serve.** Specifically, low and varying survival rates skew the mean Chicago police officer towards young, white men which limits the police force's ability to connect with and gain the trust of its constituents.

Following our analysis of Chicago's lack of sufficient representation amongst the police force, we used advanced machine-learning techniques to parse sentencing data to determine how sentences are decided. **We found overwhelming evidence that the decision as to the length and personal impact of a sentence on a defendant lies squarely in the judge's purview.** We found evidence that judge selection likely plays an outsized role in inequitable sentencing and disproportionately high plea deal rates. Specifically, in Chicago alone, over 10% of judges exhibit statistically significant deviations from plea deal norms, indicating potential bias.

At the same time, we also found evidence that the role that race plays in the sentencing process has substantially shrunk over time. Although the impact still is not zero, **race's waning utility as a predictive factor of sentence length and final disposition points to a successful and ongoing transition towards an equitable judicial system.**

Then, we used novel, advanced modeling techniques through synthetic control modeling within a difference-in-differences model to measure the exact delta values with respect to crime rates in cities across the United States. **We found that diversion programs such as LEAD are inherently risky: if they are implemented without the necessary support, such programs run the risk of low-level offenders falling through the cracks. However, given robust infrastructure and complementary policies, we find that diversion programs exhibit impressive crime reduction properties when part of a comprehensive approach.** While we found no statistically significant effect on economic growth and measures of judicial equity, future research can examine these longer-term impacts.

Taken together, **we have provided the first end-to-end evaluation of Chicago's criminal justice system, and found Chicago's various approaches to criminal justice reform to have varying impacts.**

Such insights, however, have the potential to inform work in other cities as well. Chicago doesn't exist in a vacuum, and nor do our study's findings. Indeed, our Mahalanobis testing showed remarkable symmetries between Chicago and a host of other cities.

Policy Recommendations

Chicago is unique in its status as a leader with respect to criminal justice reform. Through our statistical analyses of the impacts of the city's policies, **we found a strong preference for policies with concrete implementations and community elements over broad, top-down initiatives to drive change.**

Additionally, **our report has highlighted the outsized influence that potential bias can have on outcomes.** Cities like Chicago should evaluate emergent technologies and other innovative policy means to ensure the equity of the criminal justice system across the board [23].

Future Research Directions

Throughout our investigation, we came across several promising avenues for future work.

Youth-Focused Resources

We found quite early on in our exploratory data analysis that youth crime rates were particularly elevated during the early years of the time frame we considered. A full exploration of the effect of educational reform as well as resources like after-school enrichment and job training programs could reveal promising insights into the impact such policies have on crime rates.

Economic Consequences

Our difference-in-differences model paired with Mahalanobis distance testing and synthetic controls revealed that there was no correlation between criminal justice initiatives like diversion policies and economic growth. While our analysis fails to find significant effects on economic growth and measures of judicial equity, future research can examine these longer-term impacts.

1. Burnett, S. (2023, March 7). Chicago mayor's race dominated by concerns about city crime. AP News. https://apnews.com/article/politics-2022-midterm-elections-lori-lightfoot-chicago-1441bca5446f2eb30cd9dfe377f85fff

2. Oceguera, E. E. and R. (2023, May 16). Illinois criminal justice reform ends cash bail, changes felony murder rule. Injustice Watch. https://www.injusticewatch.org/news/2021/illinois-criminal-justice-reform-cash-bail-felony-murder/

3. Johnson, A. (2023, July 10). Here's why warm weather causes more violent crimes-from mass shootings to aggravated assault. Forbes. https://www.forbes.com/sites/ariannajohnson/2023/07/06/heres-why-warm-weather-causes-more-violent-crimes-from-mass-shootings-to-aggravated-assault/

4. Johnson, A. (2023, July 10). Here's why warm weather causes more violent crimes-from mass shootings to aggravated assault. Forbes. https://www.forbes.com/sites/ariannajohnson/2023/07/06/heres-why-warm-weather-causes-more-violent-crimes-from-mass-shootings-to-aggravated-assault/?sh=8bb99fa5ab38

5. FBI. (2020, November 10). 2019. FBI. https://ucr.fbi.gov/nibrs/2019

6. Joseph, G., Fernelius, K. J., Ewulonu, N., Behne, M. F., & Rhodes, A. M. (n.d.). NYPD commander's text messages show how the quota system persists. The Appeal. https://theappeal.org/nypd-commanders-text-messages-show-how-the-quota-system-persists/

7. Keshner, A., & Tracy, T. (2018, April 8). NYPD chief James O'Neill pledges wrath against quota-chasing cops. New York Daily News. https://www.nydailynews.com/2017/05/12/nypd-chief-james-oneill-pledges-wrath-against-quota-chasing-cops/

8. Mann-Kendall Test for monotonic trend. Design Trend Mann-Kendall. (n.d.). https://vsp.pnnl.gov/help/vsample/design_trend_mann_kendall.htm

9. Yue, S., & Wang, C. (n.d.). The mann-kendall test modified by effective sample size to detect trend in serially correlated hydrological series - water resources management. SpringerLink. https://link.springer.com/article/10.1023/B:WARM.0000043140.61082.60

10. Advancing diversity in law enforcement. US EEOC. (n.d.). https://www.eeoc.gov/advancing-diversity-law-enforcement

11. The relationship between race, ethnicity, and sentencing: Outcomes: A ... (n.d.). https://www.ojp.gov/pdffiles1/nij/grants/208129.pdf

12. Federal Criminal Sentencing – Race-Based Disparate Impact and Differential Treatment in Judicial Districts (2023). https://www.rand.org/pubs/external_publications/EP70192.html.

13. Rethinking prison as a deterrent to future crime - JSTOR DAILY. (n.d.-a). https://daily.jstor.org/rethinking-prison-as-a-deterrent-to-future-crime/

14. The Economist Newspaper. (n.d.). A record number of Americans died from drug overdoses in 2020. The Economist. https://www.economist.com/graphic-detail/2021/07/15/a-record-number-of-americans-died-from-drug-overdoses-in-2020

15. The Economist Newspaper. (n.d.-b). Oregon botches the decriminalisation of drugs. The Economist. https://www.economist.com/leaders/2023/04/13/oregon-botches-the-decriminalisation-of-drugs

16. Lead: Law enforcement assisted diversion. LEAD: Law Enforcement Assisted Diversion | National Institute of Corrections. (n.d.). https://nicic.gov/resources/nic-library/web-items/lead-law-enforcement-assisted-diversion

17. YouTube. (2023, October 14). “it changed my life”: Spokane County’s jail diversion program turns local man’s life around. YouTube. https://www.youtube.com/watch?v=NLzWTRTJjBI

18. Initiative, P. P. (n.d.). Racial disparities in diversion: A research roundup. Prison Policy Initiative. https://www.prisonpolicy.org/blog/2023/03/07/diversion_racial_disparities/

19. Willging, C. E., Search for more papers by this author, Semansky, R., Dr. Willging is with the Pacific Institute for Research and Evaluation, PS, H., CE, W., Al., E., H, W., M, G., & RM, S. (2014, August 1). State Mental Health Policy: Back to the future: New Mexico returns to the early days of medicaid managed care. Psychiatric Services. https://ps.psychiatryonline.org/doi/10.1176/appi.ps.201400102

20. U.S. News unveils the 2017 best places to live. US News. (n.d.). https://www.usnews.com/info/blogs/press-room/articles/2017-02-07/us-news-unveils-the-2017-best-places-to-live

21. The Opioid Crisis in Illinois. (n.d.-b). https://www.dhs.state.il.us/OneNetLibrary/27896/documents/OpioidCrisisInIllinois_051617.pdf

22. Mental Health. (n.d.). https://doh.wa.gov/sites/default/files/legacy/Documents/1000//SHA-MentalHealth.pdf

23. Tashea, J. (2017, April 17). Courts are using AI to sentence criminals. that must stop now. Wired. https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/