

IDS 701 Unifying Data Science Final Project

April 2022



Estimating the Impact of Support Team Assisted Response in Denver, Colorado

Dauren Bizhanov
dauren.bizhanov@duke.edu

Cindy Chiu
yuting.chiu@duke.edu

Sydney Donati-Leach
Sydney.donatileach@duke.edu

Aarushi Verma
Aarushi.verma@duke.edu



Motivation

Traditionally, when one calls 9-1-1 in the United States, the operator will dispatch the appropriate emergency agency to one's location. These emergency services can be either the police, the fire department, or an ambulance. However, in Denver, Colorado, there is another service known as the Support Team Assisted Response (STAR). The STAR experiment was implemented on June 1, 2020 to provide assistance when someone calls 9-1-1 for something like a mental health crisis, substance use issue or even incidents like homelessness or public intoxication. The 9-1-1 operators can dispatch a STAR team to the scene which includes a licensed behavioral health professional and a paramedic. The program began with only two teams of professionals in two vans. In November 2021, the program expanded to six teams with three vans. Finally, in February 2022, Denver approved \$1.4 million for STAR to increase to ten teams and six vans, running seven days a week from 6 a.m. to 10 p.m. (Schmelzer, 2022).

The idea behind the program was to provide better access to mental health services and ideally decrease the amount of violence that can often occur when a police officer is called to the scene. Police officers are not trained to handle mental health crises in the same way a specially trained psychologist or social worker is. Officers may resort to violence against the person in mental distress, and often arrest them when the situation does not garner it. Mental health crises make up almost half of 9-1-1 calls and more than 60% of the jail inmates have mental health issues (Jame et al., 2006).

Denver is not the only city to implement this kind of program. Portland, Oregon, San Francisco and Oakland, California, Durham, North Carolina, and Austin, Texas have also funded pilot programs. Each city is citing growing success with its programs which means more cities will surely follow suit. The purpose of this analysis is to corroborate these reports of success by looking at how STAR has impacted non-violent arrest rates in Denver, Colorado.

Research Design

In order to determine if the STAR program was effective in decreasing non-violent arrest rates in Denver, we used the causal inference approach of difference-in-difference. Through difference-in-difference, we will compare the non-violent arrest rates of our treatment county, Denver, to that of our control counties within the same time period. The time period we are analyzing is January 2019 until December 2020. We will choose the control counties based on proximity to Denver county, which will ensure a similar city layout and population which will minimize any baseline differences. To assess this difference-in-difference we will also need to normalize these trends by each county's respective population and get the arrest rates per county rather than just a count of arrests per county. This will allow us to isolate the effects of the treatment, the STAR program, which was only implemented in Denver county, and compare it against other counties in the state where the STAR program was not implemented.

If the STAR program is effective, we expect Denver's non-violent arrest rates, post-implementation of the program to be different than the other counties that did not implement the program. The difference-in-difference approach helps us minimize the potential state-wide or nationwide impact on non-violent arrest rates that could have taken place and evaluate the effect of the program alone.

Datasets

We obtained data for both national crime and arrests to calculate the number of non-violent arrests and the number of non-violent crime incidents that took place within our timeframe of interest. Since there is no



single source data that contains both crime and arrest rates, we obtained the data from two different sources and conducted analysis separately on both data.

For crime data, we obtained the dataset from the National Incident-Based Reporting System (NIBRS). Each row represents an incident, and the only location identifier is agency ID. We then obtained the county FIPS code by joining the agency ID to the intermediate file we generated by aggregating the agency ID to FIPS. We joined the offense type code to offense type description to determine what kind of crimes were committed during each incident. The purpose of this was to categorize incidents as either violent or non-violent (Kaplan, 2021). The STAR program only assists non-violent crimes such as homelessness or public disorder, so we needed the ability to filter out violent crimes. We discovered that one incident can involve both violent and non-violent crime. In order to eliminate the chance of counting an incident in both categories, best practice is to look at the most serious associated offense committed in the incident. Therefore, if there is at least one violent crime documented in the incident, the entire incident is considered violent. After joining all the tables together, we aggregated the data to the county level by month and calculated the crime rate per 100,000 people to account for the county population.

For arrest data, we obtained the dataset from Uniform Crime Reporting (UCR) Program Data. This data contains the number of arrests within each county on a monthly basis from 1974 to 2020. The STAR program was launched in June 2020, so we filtered the data to monthly arrest rates from 2019 to the end of 2020, which would give us roughly six months of post-treatment data. The methodology behind choosing two full calendar years was to account for any seasonality that may take place with crime. The data is composed of more than 1,200 columns. The first few columns contain the information of geographic locations such as agency name, state, county Federal Information Processing Standards (FIPS) code and population of the county. The other columns represent a different type of crime by demographic such as age, gender, and racial subgroups. Each row in the data was pre-aggregated according to these demographic subgroups. Since we did not have individual level data, we focused on extracting these aggregated arrest numbers for each crime and agency. Again, we split arrests into violent and non-violent and then aggregated arrests on a county and monthly level. Finally, we found the arrest rate by dividing the total arrest number by the county population and calculated the monthly arrest rate per 100,000 people.

Comparison Counties

The goal of our analysis is to determine if the implementation of the STAR program had an impact on the non-violent arrest rates (outcome variable) in the county where it was implemented (treatment) as compared to the counties where it was not implemented (control). Before doing so, we need to ensure in our selection process that the treatment and the control counties are similar in nature.

The two criteria we used for similarity were the population of the counties and the geographical location of the counties within Colorado. We suspected that counties in close proximity to Denver would mimic the crime rate and arrest rate in Denver. In addition, these counties close to Denver are also metropolitan as opposed to other counties in Colorado, and therefore would have similar cultures of city life. While the population did vary significantly for some of the counties, our analysis was based on a normalized arrest rate with respect to the population, so we decided to proceed with these counties in our control group. The table below represents the counties in our treatment and control groups along with their FIPS codes.



Table 1. Treatment Assignment and FIPS Codes for Counties Included in Analysis

County Name	FIPS Code	Treatment/Control
Denver	08031	Treatment
Adams County	08001	Control
Arapahoe	08005	Control
Boulder	08013	Control
Broomfield	08014	Control
Douglas	08035	Control
Jefferson	08059	Control

Table 1 shows all the FIPS codes for each county we selected in Colorado. FIPS code uniquely identifies the geographic area. The red row is our treatment county, and the blue rows are our control counties.

Figure 1. Map of the State of Colorado and the Treatment/Control Counties Included in Analysis

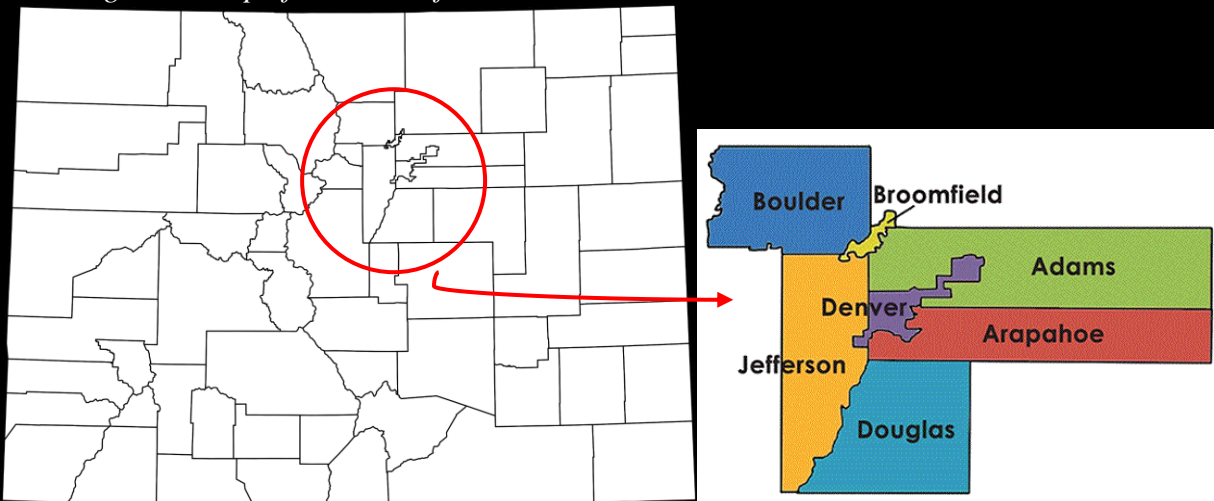


Figure 1. shows an outline of all the counties in the state of Colorado. The colored section is a larger view of the of the seven counties we include in this analysis.

Summary Statistics

Before we performed our analysis, we examined the summary statistics of the data to understand the distribution of the outcome variable. Before the STAR program was implemented, the treatment group had an average arrest rate of 400.32 per 100,000 people per month. The control group had an average total arrest rate of 283.49. After June 2020, Denver had a mean arrest rate of 241.48. The decrease in mean total arrest rate might be due to COVID-19 because we also saw the decrease in the control group to 174. Furthermore, we saw a higher trend in male arrest rates as compared to females in both control and treatment groups.



Table 2. Average Arrest Rates of Treatment and Control Counties

Treatment group	Time period	Average total arrest rate (per 100,000 people)		
		Total	Male	Female
Control	Before treatment	283.49	197.18	86.31
Treatment	Before treatment	400.32	297.23	103.08
Control	After treatment	174.00	125.95	48.05
Treatment	After treatment	241.48	181.43	60.05

Table 2. shows the average arrest rate of different genders of both control and treatment group. The arrest rates are broken into before the STAR program is implemented, which is before June 2020, and after the STAR program is launched.

Trend Analysis

Before we perform the difference-in-difference regression, the parallel trend assumption needs to be satisfied. As shown in Figure 2, the average total arrest rate (per 100,000 people) between the control and treatment groups is similar before June 2020. This means we can accurately examine the causal effect of the STAR program. If the trends are no longer similar after 2020, that would indicate that the STAR program is having an impact, either positively or negatively, on non-violent arrests.

Figure 2. Average Arrest Rates for Treatment and Control Counties

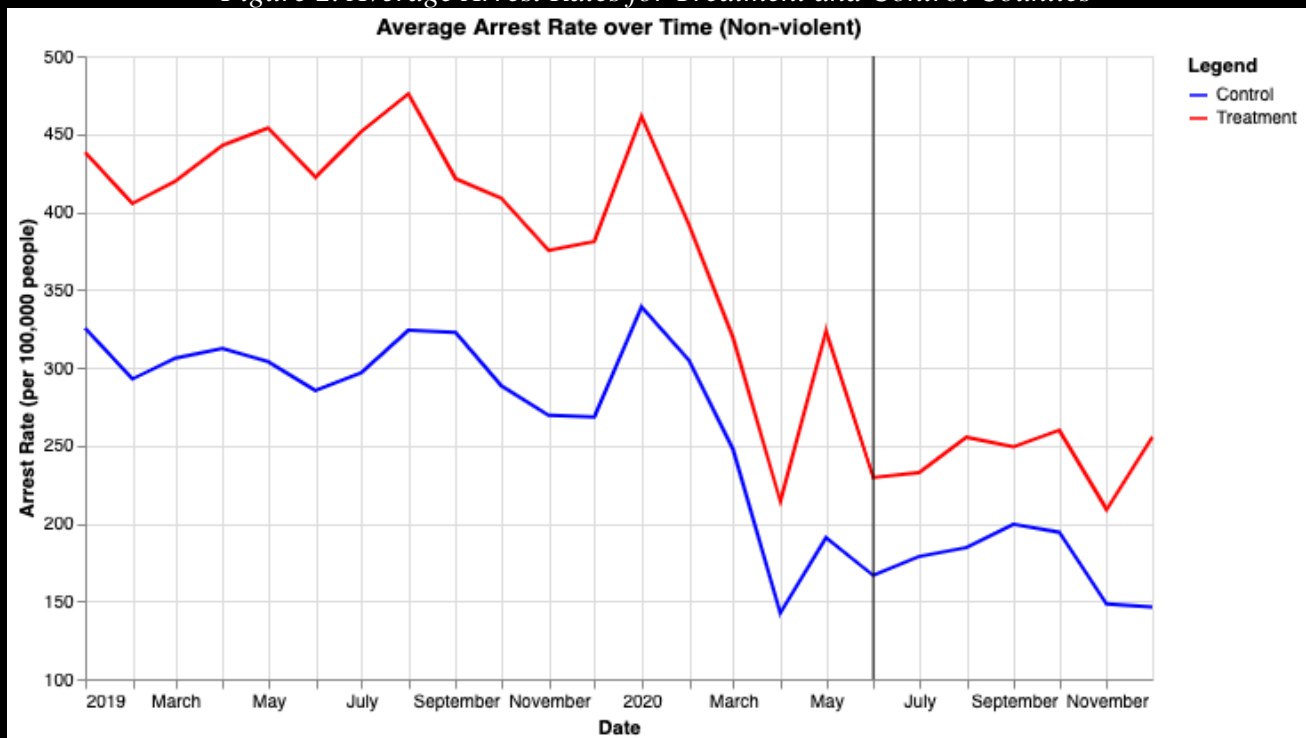


Figure 2 shows the average non-violent arrest rates for both control and treatment counties by month. The blue curve is the control counties, and the red curve is the treatment county. The black line denotes the launch of the STAR program, which is June 2020. The trend curves were parallel before the start of STAR.

We also examined the non-violent crime rate trend (per 100,000 people) before and after the STAR program was implemented. The purpose of analyzing non-violent crime incidents is to evaluate if there are



any underlying changes in the crime rate that might cause the arrest rates to increase or decrease accordingly. In Figure 3, we can see the non-violent crime rate increased after the program's start date. This means that if we see any decrease in non-violent arrest rates, it cannot be attributed to a decrease in non-violent crime.

Figure 3. Average Crime Rates for Treatment and Control Counties

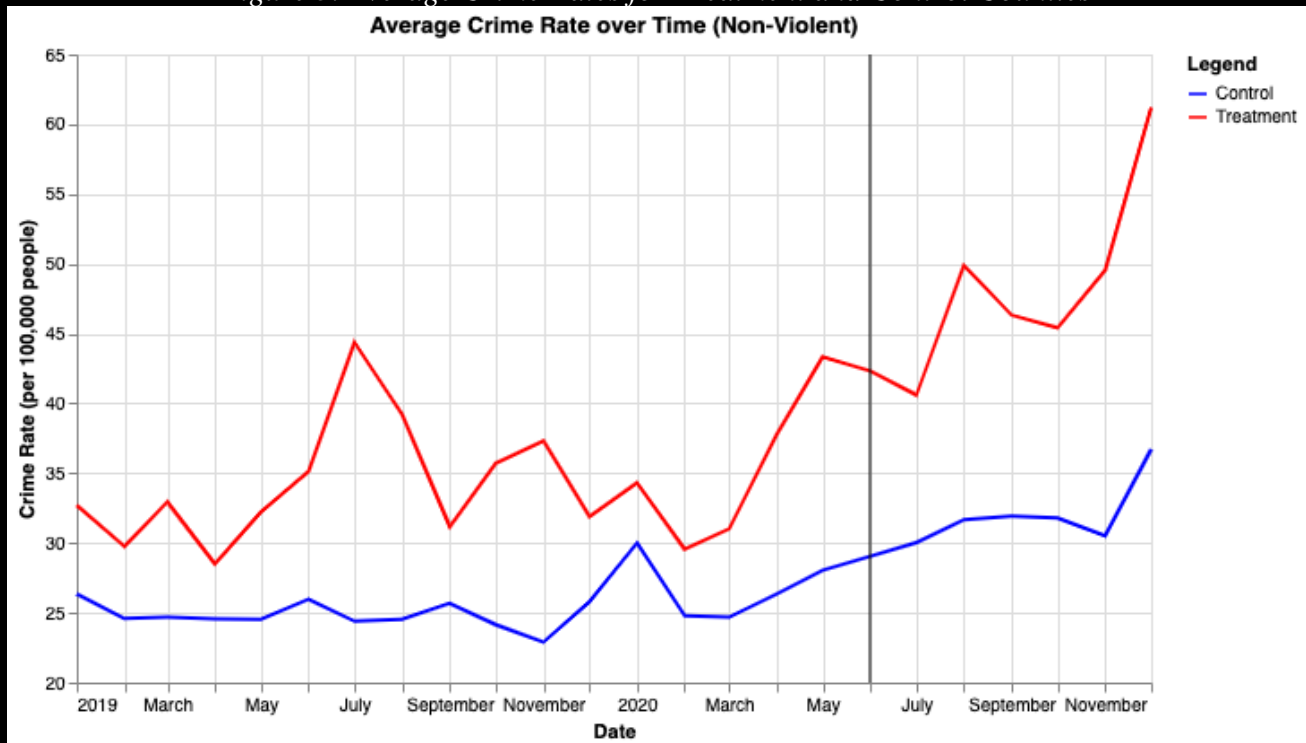


Figure 3. shows the average non-violent arrest rates for both control and treatment counties by month. The black line denotes the launch of the STAR program, which is June 2020. The average non-violent crime rates are not decreasing after STAR was launched. Therefore, we can exclude the effect of decreasing crimes from our analysis.

Finally, we wanted to examine if the STAR program influenced violent arrest rates in any way. In many cases, when police are dispatched to a non-violent situation, it can escalate into a violent situation. This can be due to the mental health of the person committing the crime, or the police officer choosing to take violent action. It is possible the STAR program is decreasing the likelihood of that occurring because they are trained to de-escalate situations and provide appropriate mental health support. However, as Figure 4 shows, there is no clear indication of any effect taking place on the violent arrest rate.



Figure 4. Average Violent Arrest Rates for Treatment and Control Counties

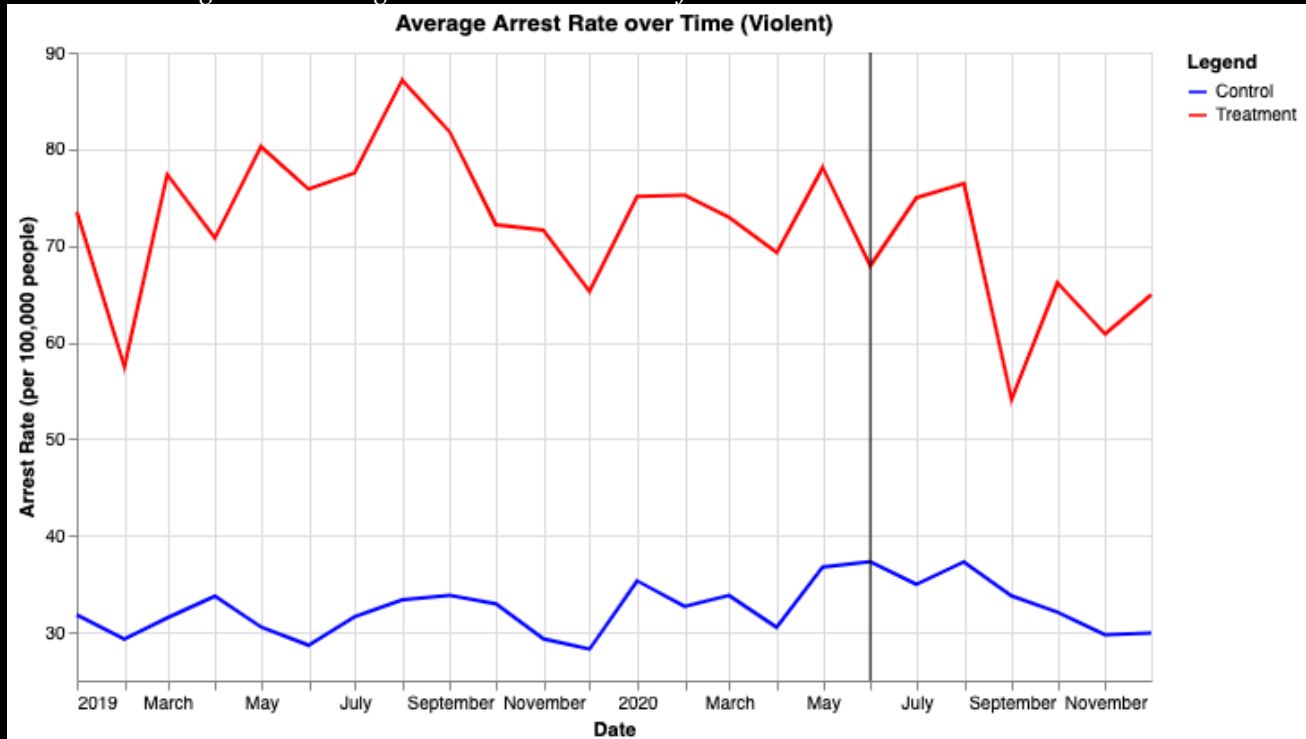


Figure 4. shows the average violent arrest rates for both control and treatment counties by month. The black line denotes the launch of the STAR program, which is June 2020. The average violent arrest rates seem random before and after the STAR is launched. Therefore, we can exclude violent arrest from our analysis and focus on non-violent arrest rates.

Results

The results are separated into two different strategies: a placebo test and a difference-in-difference regression model. In both strategies we employed multiple regressions where the independent variable was a demographic subgroup. This was necessary because of the structure of the arrest data. The data did not provide arrests on an individual level, but instead were pre-aggregated into demographic subgroups. Therefore, we could not add these variables as features in one model, and instead opted to model each of them separately. We also included county and month fixed effects into each regression.

The idea behind the placebo test was to further explore the parallel trend assumption within each demographic subgroup. We limited the data to only the pre-implementation time period of January 1, 2019, to May 30, 2020. The simulated implementation date we chose was January 1, 2020. The results of the placebo test can be seen in Table 3. For the overall non-violent arrest rate, the difference between the control and the treatment groups is not significant. This is also true for white, male, and female subgroups. This result gave us additional confidence that the assumption is met for the abovementioned subpopulations. However, other groups had significant differences between the treatment and the control. Therefore, any results we obtain using the difference-in-difference method should not be credible for those subgroups as the main assumption is violated.



Table 3. Placebo test results

Coefficient	Standard Error	T	P-Value	Lower CI	Upper CI	Independent Variable
-33.945	36.966	-0.918	0.361	-107.29	39.403	Overall
-17.841	4.232	-4.216	0.000	-26.237	-9.445	Black
-15.304	34.321	-0.446	0.657	-83.404	52.795	White
1.022	0.571	1.789	0.077	-0.112	2.155	Asian
-1.806	0.287	-6.293	0.000	-2.376	-1.237	American Indian
-26.096	25.196	-1.036	0.303	-76.09	23.898	Male
-7.849	12.072	-0.65	0.517	-31.803	16.105	Female

Table 3. shows the regression results from the placebo test. The placebo test was implemented using Jan 2020 as treatment date and performed a diff-in-diff regression using that date. The p-value of the overall non-violent arrest rate is not significant, which means we don't see a difference between control and treatment group before and after Jan 2020. However, some racial groups have significant p-values and might violate the model assumption in those groups.

The difference-in-difference (DID) regression results are shown in Table 4. The null hypothesis in our analysis is that the effect of the STAR program on non-violent arrest rates is not zero. In other words, we were testing if the STAR program either had a negative or a positive effect on the non-violent arrest rates in Denver, Colorado. The point estimate of the DID effect on the overall non-violent arrest rate is -49.34. This estimate by itself is practically significant as it is approximately 20% of the mean post-treatment arrest rate in the treatment group. However, the standard errors are high and the upper part of the 95% confidence interval crosses over zero. This means we cannot reject the null hypothesis at the 0.05 alpha level. The conclusion is the same for the demographic subgroups. Since our placebo test shows that we cannot rely on differential effects on these subpopulations, we should ignore any statistically significant results due to lack of credibility.

Table 4. DID regression results

Coefficient	Standard Error	T	P-Value	Lower CI	Upper CI	Independent Variable
-49.34	29.173	-1.691	0.093	-106.99	8.309	Overall
-27.332	6.753	-4.048	0.000	-40.676	-13.988	Black
-20.685	26.214	-0.789	0.431	-72.487	31.117	White
0.494	0.462	1.069	0.287	-0.419	1.406	Asian
-1.689	0.121	-13.95	0.000	-1.929	-1.45	American Indian
-44.569	19.002	-2.346	0.02	-82.119	-7.019	Male
-4.772	10.428	-0.458	0.648	-25.378	15.835	Female

Table 4. shows the diff-in-diff regression results. The coefficient of the overall non-violent arrest rate is negative, which indicates STAR has a negative impact on the non-violent arrest rate. However, the p-value of the overall non-violent arrest rate is not significant at 0.05 alpha level.

Interestingly, we found statistically significant results in male and female subgroups. These subgroups also passed the placebo test. In addition, the coefficient for the male group is around 24% of the average total non-violent arrest rate and the coefficient for the female group is only 6% of the average total non-violent arrest rate. This result needs further analysis and investigation but is out of scope for this report.

The final aspect of our results is a difference-in-difference plot. The overlay of the scatter plot on the linear regression shows that the data is in fact nonlinear; however, a linear regression was the best method to ensure the results of our model could be interpreted. The decreasing trend we see in the regression results can be attributed to the COVID-19 pandemic and the lockdowns that occurred. While the arrest rates do appear to



be lower after the implementation of STAR, we cannot draw any conclusions from this plot for the same reasons we cannot draw any conclusions from Table 4.

Figure 5. Difference-in-Difference Plot of Arrest Rates

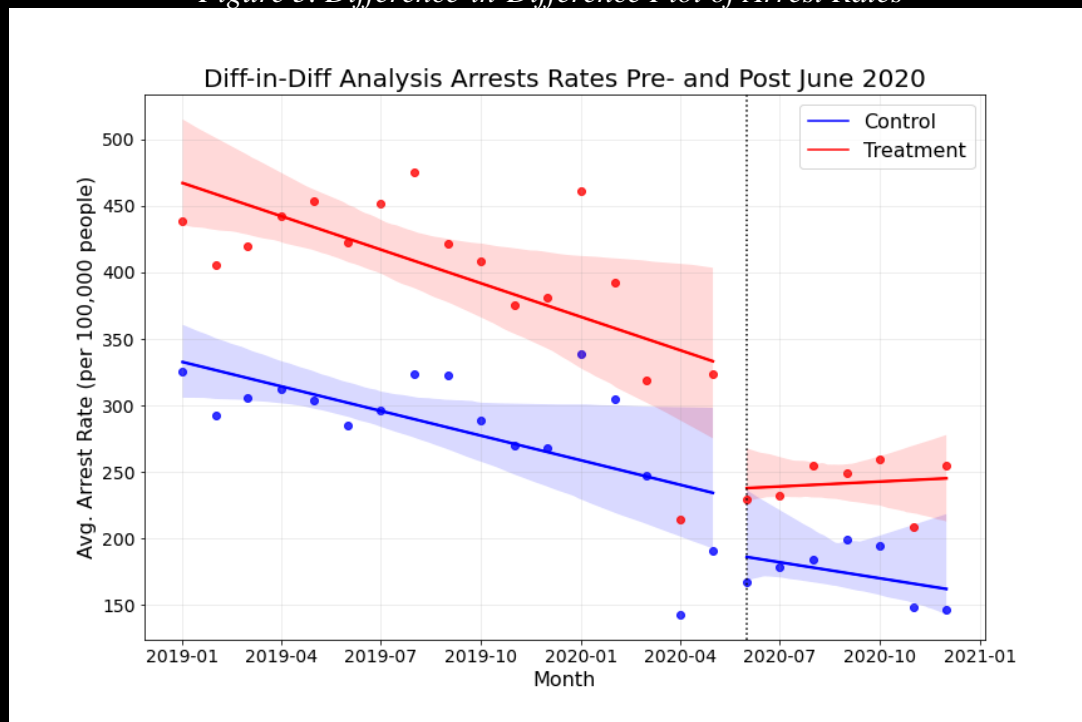


Figure 5 shows both the overall linear regression for the treatment and the control groups as well as a scatter plot of the average arrest rates. The vertical dashed line in the plot represents the date the STAR program was implemented. The decreasing trend before implementation is most likely due to COVID-19. Any trends after implementation are unclear as the results of the regression are insignificant.

Conclusion

Based on the results from our regression, we cannot assume that the STAR program decreased the amount of non-violent arrest rates. The overall trend from the difference-in-difference regression is lower than the counterfactual trend; however, this can be attributed to the fact that arrest rates were already decreasing before the program was implemented in June 2020, and there could have been ongoing effects from the pandemic lockdowns.

There were only two teams at the beginning of the STAR program so the likelihood that they were able to impact overall non-violent arrest rates is slim. We would need arrest data past December 2020 to evaluate if STAR was more successful as the number of teams and resources grew over time. It is promising that the program was successful as the point estimate is negative and could continue on that trend, but we cannot say for certain if that is the case.

Another limitation of this analysis was the fact that the STAR program does not release information about which kinds of non-violent crime the 9-1-1 operators route to them. We do not know the training that 9-1-1 operators receive to appropriately route calls to STAR. Additionally, our analysis bucketed all non-violent arrests, but it is highly probable not all non-violent incidents are appropriate for the program. If the STAR program can provide specific crimes that they are focusing on, especially the offense type, future studies can implement the difference-in-difference analysis on relevant arrests and provide a more specific analysis.



Why does Denver believe this program is successful and why are so many other cities implementing similar programs too? We believe the answer lies in the racial justice movement that occurred in 2020 (Lo et. al, 2021). The movement was asking for alternatives to violent policing that was taking place across the county, and the STAR program provided exactly that. Just because the program does not have statistically significant results does not mean that the program has not made a significant impact. The difference can be seen in how people are treated in mental health crises and how those in need of support are given it.



Resources

1. James, D. J., & Glaze, L. E. (2006). Mental health problems of prison and jail inmates.
2. Kaplan, Jacob. Jacob Kaplan's Concatenated Files: Uniform Crime Reporting (UCR) Program Data: Arrests by Age, Sex, and Race, 1974-2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-09-27. <https://doi.org/10.3886/E102263V14>
3. Kaplan J (2021). Uniform Crime Reporting (UCR) Program Data: A Practitioner's Guide. <https://ucrbook.com/>
4. Lo, K., Pearl, B., Figgatt, S., & Parsons, C. (2021, November 3). *5 discussions that shaped the justice reform movement in 2020*. Center for American Progress. Retrieved April 27, 2022, from <https://www.americanprogress.org/article/5-discussions-shaped-justice-reform-movement-2020/>
5. Schmelzer, E. (2022, February 21). *Thousands of calls later, Denver's acclaimed program that provides an alternative to police response is expanding*. The Denver Post. Retrieved April 7, 2022, from <https://www.denverpost.com/2022/02/20/denver-star-program-expansion/>
6. United States Department of Justice, Federal Bureau of Investigation. 2019-2020. Crime in the United States. Retrieved April 1, 2022, from (<https://crime-data-explorer.app.cloud.gov/pages/home>).

GitHub Repository

<https://github.com/MIDS-at-Duke/uds-2022-701-team-5/>