# Quantifying Change: A Data-driven Exploration of Police Defunding's Impact on Use of Force and Arrests

By Lucy Civitella, Annika More, Isaac Tabb, and Abby Vece

## Table of Contents

# Abstract

This study investigates if Burlington, VT's defunding of the police had the intended impact of reducing use of force incidents and alleviating use of force and arrest disparities amongst racial groups. Poisson regression models comparing use of force in Burlington to a control city, Norfolk, VA, observed that no matter the demographic, Burlington's defunding of the police prevented a reduction in use of force. A geospatial analysis examining use of force by area of Burlington found there was a difference in the counts of use of force incidents in different wards but not in different areas. Using Burlington arrest data, two logistic regression models were fit to investigate the effect of demographic factors on use of force. These models found that Black individuals had higher odds of being affected by use of force when compared with White individuals. Separate logistic regression models found that age, race, and gender were insignificant predictors for both weaponless and firearm-pointed force types but that the odds of both types increased after defunding. Overall, this study uncovers trends which suggest that Burlington's defunding of the police did not achieve the intended improvements that Burlington aimed to address, potentially causing the opposite effect.

# Introduction

On June 30th 2020, the Burlington City Council passed a resolution to reduce funding for typical police officer positions in the Burlington Police Department (Siegel, 2021). This was done in response to national unrest following the police killings of Breonna Taylor and George Floyd. The plan resulted in the number of police dropping from 95 to 64 (Siegel, 2021). Notably, not until six months after this resolution passed did the Burlington City Council vote to reallocate funds towards 10 community service officers and 3 community support liaisons (Siegel, 2021). Around the same time that Burlington, VT defunded their police department, many other cities jumped into similar plans that involved cutting police budgets, but evidence of this strategy's direct impact on reducing use of force incidents or crime rates was, and still is, limited and inconclusive. The need for more rigorous research and evaluation of such sweeping policy changes is still needed.

The goal of this study is to investigate the impact of Burlington defunding the police on use of force and arrest statistics after the policy was implemented. Generally, the intended effect of defunding the police is to reduce use of force counts and alleviate any disparities in use of force and arrests amongst racial groups. For the analysis, there are two data sets from the city of Burlington, a catalog of use of force data (City of Burlington, 2023) and a catalog of arrest data (City of Burlington, 2023). To first analyze the basic effects of Burlington's defunding of the police, Burlington's use of force counts were compared against counts in a control city, Norfolk, VA, which did not defund their police force

(Norfolk Budget & Strategic Planning, n.d.). Additionally, use of force counts separated by racial group were examined to unearth insights on the impact of Burlington's defunding on use of force against different racial groups. It was also important to investigate if the impact of Burlington's defunding on use of force counts differed by wards or areas. Finally, research was conducted to uncover whether individual factors of an arrest could predict use of force both before and after defunding, and whether these factors could predict the type of force. Overall, the goal of this study is to investigate if Burlington's defunding of the police had the intended impact of reducing use of force counts and alleviating disparities in use of force and arrests amongst racial groups.

# Methods

Datasets Used for Analyses

There were three total datasets used for this project. The first two datasets were for Burlington, VT and came from the City of Burlington Data Hub. The first dataset was a catalog of use of force incidents in Burlington, VT (City of Burlington, 2023) while the second dataset was a catalog of arrests in Burlington, VT (City of Burlington, 2024). The third dataset that was used for this project was a catalog of use of force incidents (Norfolk Police Department, 2023) in Norfolk, VA, the control city for the segmented Poisson regression analyses. The Norfolk, VA dataset originally includes both use of force incidents and citizen complaints but was reduced to only the use of force incidents.

Impact of Burlington Police Defunding on Use of Force Incident Counts

To analyze the effect of Burlington defunding the police on use of force incident counts, a segmented Poisson regression model was employed. The model fit four separate lines, one for before and one for after the Burlington defunding time point (Quarter 3, 2020) in both Burlington and a control city. To conduct the Poisson analysis, Burlington's use of force data was compared to Norfolk, VA, a control city which never defunded their police force. Norfolk, VA was selected as the control city because it was the only city that could be found which had not defunded their police force and had sufficient available use of force data.

Each row in the original Burlington use of force dataset represented a single individual affected by use of force. For the Poisson analysis, a separate Burlington dataset was formed, containing the count of individuals affected by use of force for each quarter of each year. For the Norfolk use of force dataset, each row represented a single incident of use of force and could contain multiple individuals. The

Norfolk dataset was converted to have the same structure as the original Burlington use of force dataset, where each row represented an individual rather than an incident. For the Poisson analysis, a new Norfolk dataset was formed containing the count of individuals affected by use of force for each quarter of each year. The Burlington and Norfolk datasets were eventually subsetted to only the quarters where the two datasets overlapped (2019 Q1 to 2023 Q2). The two datasets were then merged by quarter. The final dataset used for the Poisson analysis contained four columns. The first column was Quarter, representing the quarter and year. The second column was City, an indicator variable where 0 represented Norfolk, VA (control) and 1 represented Burlington, VT (defunded). The third column was called BeforeAfter, an indicator variable which was 0 for before the defunding time point (Quarter 3, 2020) and 1 if from after the defunding time point. The fourth and final column of the Poisson dataset was Count, holding the count of use of force incidents for the corresponding quarter/city pairing.

To perform the Poisson analysis, a Poisson Regression model was built in R using the glm( ) function. The equation for the Poisson model was as follows:

$$Count \sim Poisson(\lambda_x)$$

$$\lambda_x = \beta 0 + \beta 1 * City + \beta 2 * Quarter + \beta 3 * BeforeAfter + \beta 4 * City * BeforeAfter + \beta 5 * Quarter * BeforeAfter + \beta 6 * City * Quarter + \beta 7 * City * Quarter * BeforeAfter$$

By including terms for City, Quarter, BeforeAfter, an interaction term between each pairing, and the interaction term between all three variables, the above model fits a Poisson regression line by quarter for both Norfolk, VA and Burlington, VT before and after the defunding event occurred in Burlington. To ensure that the results of the model were valid and interpretable, all Poisson regression assumptions were checked. Additionally, the parallel trends assumption was checked for the two cities' trends before the defunding time point. For more details on the procedure used to check assumptions, see Appendix A. Finally, after checking assumptions, the intercepts and factors of change from quarter-to-quarter for Norfolk, VA and Burlington, VT before and after the defunding time point were computed. This was done to best interpret and understand the effects of the defunding event.

The exact same segmented Poisson model described above was additionally run on two other datasets which contained counts for only one of the following two racial groups: Black and White. These separate models were created to investigate the effects of Burlington's police defunding on specific racial groups. In regards to the datasets used for the separate racial group models, the only difference was that the counts were only for the specific racial group the model was investigating. It is also important to note that the original Norfolk, VA and Burlington, VT use of force datasets contained incidents with Hispanic, Asian, and Native American individuals. It was decided that models would not be created for these racial groups since there was a lack of data, and thus the results may not have been accurate.

### Geospatial Analysis of Use of Force Incidents

A geospatial analysis was used to investigate the impact of Burlington defunding the police in various areas in Burlington. To properly do the geospatial analysis, a Wilcox-signed rank test was used to analyze the use of force in the areas and wards of Burlington before and after the defunding because the distribution of counts between the areas and wards was not normal. There are five areas of Burlington according to the Burlington Police Department: Downtown, Old North End, New North End, University Hills, and South End. There are also 8 wards numbered 1-8 (City of Burlington, n.d.), which are used for voting districts. The wards are being analyzed together, and then the areas are being analyzed together. A Wilcox-signed rank test was also used to see if the use of force counts in each of the wards, and each of the areas overall went down 20%. 20% was chosen because that was how much the use of force cases went down in total from before defunding (June 2017 - June 2020) to after (June 2020 - June 2023). The analysis on the wards was changed to a smaller window because of the redistricting in Burlington in January 2023, so the analysis on the wards goes from January 23, 2018, to January 23, 2023, and has 699 incidents. This part of the analysis also uses a wider range of data overall than the other parts and deals with counts, so there is an equal amount of time on both sides of the defunding to ensure a stronger analysis

### Predictive Analysis of Use of Force in Arrests

In addition to investigating changes in use of force incident counts, it was also important to examine if the proportion of arrest incidents which involved use of force changed before and after defunding. Proportions were gathered by merging the Burlington arrest dataset with the Burlington use of force dataset by incident number. Upon examining contingency tables for each variable of interest in relation to the presence of force, it became evident that race exhibited the most substantial variation. This analysis aimed to determine the impact of defunding the police on the proportions of use of force within race groups as well as to determine if use of force varied significantly across race groups from before to after defunding. Data before spanned January 1, 2019 to June 30, 2020, while data after began July 1, 2020 and ended June 1, 2023. Since proportions were used, having a longer period of time after the defunding did not invalidate/skew the results. In addition to proportions tests, logistic regression models were fit as well. The logistic regression models gave priority to critical variables such as age, gender, and race. As many arrests only had data on whether or not use of force was administered, any arrest records lacking information for at least one of the demographic variables (age, gender, or race) were excluded from the dataset before conducting analyses. Through this process, 275 entries were removed in the data before defunding the police, and 73 entries were eliminated in the data after defunding the police. Overall, 1749 arrests were obtained before the defunding, while 2862 arrests were obtained after the defunding. Weighted random imputation was used to replace 96 missing values in the dataset before defunding the police (race: 3, gender: 92, age: 1) and 70 missing values after defunding the police (race: 7, gender: 56, age: 7).

Chi square tests were used to assess whether there was an association between race and use of force before and after defunding. Two-sample proportion tests were employed to assess if there were significant differences in proportions across race groups in each time period, as well as when comparing proportions within the same race group across both datasets. Additionally, two-sample proportion tests were employed to determine if use of force proportions before defunding the police were significantly different than after defunding the police. Assumptions for Chi-square and two sample proportion tests were tested and satisfied.

To evaluate whether use of force in arrests could be predicted by demographic factors such as race, age, and gender, two logistic regression models were fit to model the patterns before and after defunding. Assumptions for logistic regression models were satisfied through observing correlation matrices among predictors, as well as doing standard checks of the data. For both models, a validation set approach was conducted by splitting the datasets into training and test sets with test sets making up 25% of the dataset. Both regression models were fit on the training data using the glm() function to predict use of force based on race, gender, and age. Race was dummy coded to three levels (White as the reference level, African American, and Other[American Indian or Alaska Native]). Gender was dummy coded to three levels (Middle-aged [ages 31 - 50] as the reference level, Young [ages 12-30], and Elderly [ages 50+]). Gender was dummy coded to three levels (Female as reference level, Male, and Other [Transgender, Non-Binary, and other]). The equation for the both models using data from before and after the defunding was as follows:

$$P = \frac{e^X}{1+e^X}$$

Where P = probability of a 1 (presence of force)

and $X = \beta 0 + \beta 1 * raceBlack + \beta 2 * raceOther + \beta 3 * genderMale + \beta 4 * genderOther + \beta 5 * ageYoung + \beta 6 * ageElderly + \epsilon$

Arrests were classified at a threshold of 0.5. Predicted values P greater than or equal to 0.5 were classified as arrests where use of force was administered and values less than 0.5 were classified as arrests where use of force was not administered. To assess the accuracy and fit of both models, McFadden's Pseudo $R^2$, Hosmer-Lemeshow Goodness-of-Fit, Area Under the ROC Curve (C statistic), and Confusion Matrices were used.

Predictive Analysis of Type of Force by Time Period

To analyze the relationship between the predictors; age, race, gender, substance use, emotional distress, and subject action, and the likelihood of Weaponless and Firearm Pointed force being used, several logistic regression models were performed. Logistic regression was chosen because the different types of use of force are binary. Four different models were created, investigating Weaponless and Firearm Pointed types of force before and after 2020 Q2. Collinearity was investigated among the variables to note if there were any relationships which could drastically alter conclusions made. In order to do this a Chi-squared test was performed looking for an association between each type of use of force and the predictor variables. If an association was found, then further analysis was conducted. This included removing the predictor from the full model and creating interaction terms. The aim of this was to see if there was a change in the coefficient estimates or p-values, which could indicate collinearity. The purpose of this overall analysis was to investigate whether variables were important for predicting the force used, and to determine if there was evidence of collinearity.

Afterwards, it was investigated whether using the logistic model was appropriate. This was done by testing for Goodness-of-Fit using the Hosmer-Lemeshow test. The purpose of this test was to assess how well the logistic model fit the observed data by determining whether the predicted probabilities match the observed outcomes. Furthermore, in order to assess the performance of the logistic regression model, Area Under the Receiver Operating Characteristic Curve, AUROC, was used. This measured the overall performance of the binary classification model, and the models ability to distinguish between the binary cases of whether the type of use of force was present or not. The McFadden's Pseudo $R^2$ was conducted to compare the fit of the main effect model and the full model. It is crucial to evaluate the model's effectiveness in predicting the different types of force based on the predictor variables.

# Results

<u>Impact of Burlington Police Defunding on Use of Force Incident Counts</u>

Table 1a: Race demographics (Data USA, n.d. (a)) and population (Data Commons, n.d.) of **Burlington** from 2019-2023. Est. indicates estimated value.

| Year | % White | % Black | Population |
|---|---|---|---|
| 2019 | 82.90 | 5.47 | 42838 |
| 2020 | 82.60 | 4.95 | 44873 |
| 2021 | 83.80 | 4.25 | 44781 |
| 2022 | 85.02 (est.) | 3.65 (est.) | 44595 |
| 2023 | 86.28 (est.) | 3.13 (est.) | 44408 (est.) |

Table 1b: Race demographics (Data USA, n.d. (b)) and population (World Population Review, n.d.) of **Norfolk** from 2019-2023. Est. indicates estimated value.

| Year | % White | % Black | Population |
|---|---|---|---|
| 2019 | 43.40 | 40.50 | 243581 |
| 2020 | 43.10 | 40.10 | 237591 |
| 2021 | 42.90 | 40.10 | 235089 |
| 2022 | 42.70 (est.) | 40.10 (est.) | 232995 |
| 2023 | 42.50 (est.) | 40.10 (est.) | 230930 |

Tables 1a and 1b depict the percentages of the populations in Burlington, VT and Norfolk, VA that identify as either White or Black. There is currently only data available for 2019-2021. The percentages for 2022-2023 were estimated based on the increase/decrease in percentages observed from 2020 to 2021. Tables 1a and 1b show that the vast majority of Burlington's population is White while Norfolk, VA is far more balanced in terms of racial demographics. Tables 1a and 1b also clearly show that Norfolk's population is consistently between 5 and 5.5 times the size of Burlington. There is not a significant amount of data available on gender demographics in either of these cities, but for the purpose of the following analyses, the ratio of males to females is assumed to be relatively close to 1:1.

Table 2: Intercepts and factors of change from quarter-to-quarter (FOC) of predicted use of force counts before and after defunding in Norfolk, VA and Burlington, VT. Four fitted lines (before/after, Norfolk/Burlington) were created for the overall Poisson regression model along with models subsetted to only black or white individuals.

| City (*Demographic*) | Intercept: Before | FOC: Before | Intercept: After | FOC: After |
|---|---|---|---|---|
| Norfolk (*Overall*) | 92.48 | 0.995 | 66.73 | 0.929 |
| Burlington (*Overall*) | 47.08 | 0.968 | 30.26 | 1.031 |
| Norfolk (*White*) | 17.38 | 1.006 | 17.37 | 0.907 |
| Burlington (*White*) | 35.29 | 0.969 | 20.17 | 1.046 |
| Norfolk (*Black*) | 75.10 | 0.992 | 49.44 | 0.935 |
| Burlington (*Black*) | 11.78 | 0.966 | 10.26 | 0.992 |

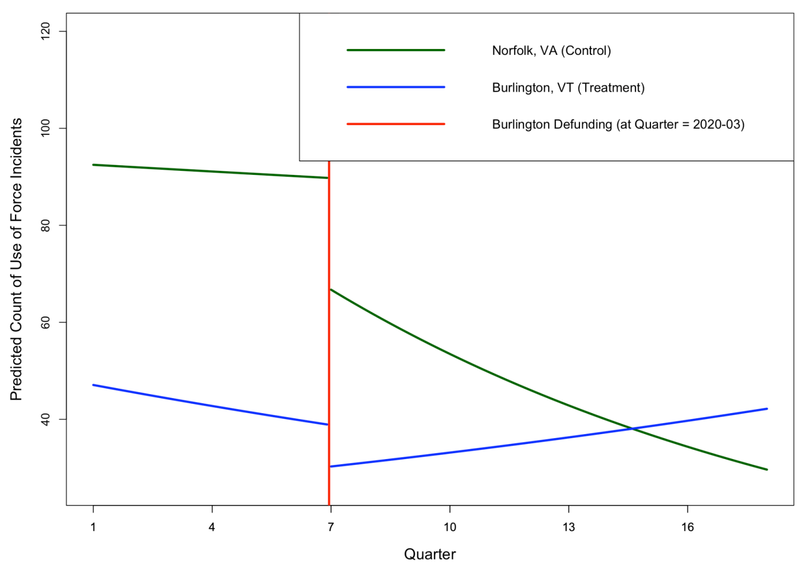


Figure 1: Fitted lines of predicted use of force counts for overall Poisson regression model for both Norfolk, VA and Burlington, VT before and after the defunding time point. Quarters range from the first quarter of 2019 (Quarter=1) to the second quarter of 2023 (Quarter=18).

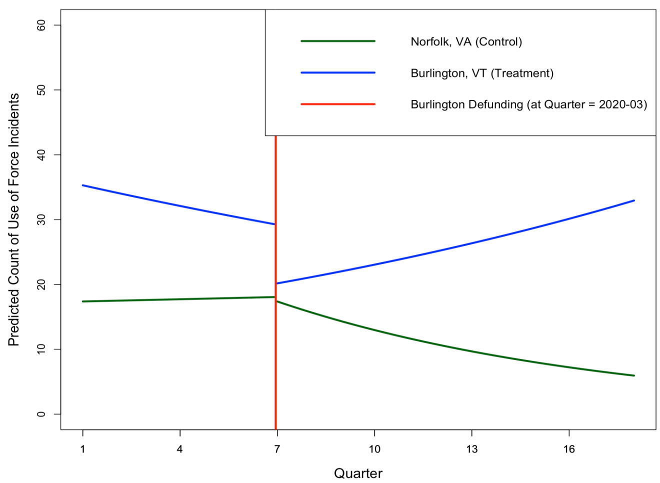


Figure 2: Fitted lines of predicted use of force counts for Poisson regression model subsetted to **White** individuals for both Norfolk, VA and Burlington, VT before and after the defunding time point. Quarters range from the first quarter of 2019 (Quarter=1) to the second quarter of 2023 (Quarter=18).

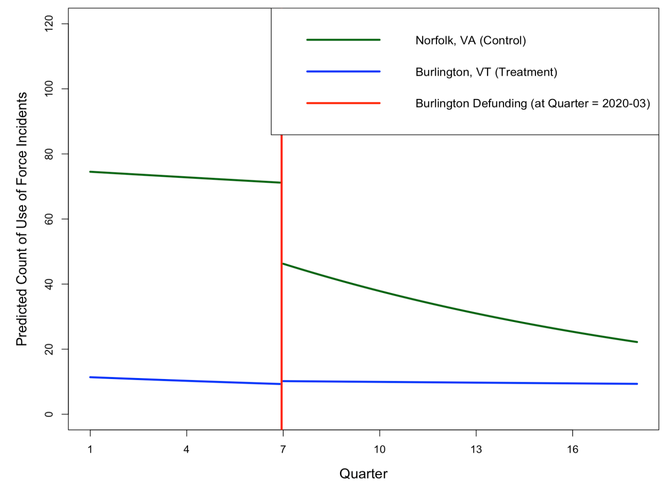


Figure 3: Fitted lines of predicted use of force counts for Poisson regression model subsetted to **Black** individuals for both Norfolk, VA and Burlington, VT before and after the defunding time point. Quarters range from the first quarter of 2019 (Quarter=1) to the second quarter of 2023 (Quarter=18).

Using the results from Table 2, the intercepts and factors of change from quarter-to-quarter were computed for each of the four fitted lines from the overall segmented Poisson regression model. For Norfolk, VA before the defunding time point, the intercept was 92.48 use of force incidents. In other words, for the first quarter of 2019, the expected count of use of force incidents in Norfolk, VA was 92.48. The factor of change from quarter-to-quarter for Norfolk, VA before the defunding time point was 0.995, meaning that for every additional quarter before the defunding time point, the expected count of use of force incidents for Norfolk, VA decreased by a factor of 0.995. For the third quarter of 2020, which was the first quarter after the defunding time point in Burlington, the expected number of use of force incidents for Norfolk, VA was 66.73. For every additional quarter after the defunding time point, the expected incident count in Norfolk decreased by a factor of 0.929.

In regards to Burlington VT, the expected count of use of force incidents in the first quarter of 2019 was 47.08, and for every additional quarter before the defunding time point, the expected count of incidents decreased by a factor of 0.968. For the first quarter after defunding, the expected count of use of force incidents in Burlington, VT was 30.26, and for every additional quarter after defunding, the expected count of incidents increased by a factor of 1.031.

The intercepts and factors of change from quarter-to-quarter for the four fitted lines from the overall Poisson model are depicted visually in Figure 1. Through the visualization, it is clear that the factors of change from quarter-to-quarter of the fitted lines for Norfolk and Burlington are similar before the defunding time point. After the defunding time point, the factor of change from quarter-to-quarter for Norfolk becomes clearly negative while the factor for Burlington becomes clearly positive.

Referring back to Tables 1a and 1b, one should note that the difference in use of force incident counts between Norfolk and Burlington does not line up with the two cities' difference in population. Despite Norfolk's population being 5 to 5.5 times the size of Burlington, the expected total use of force incident counts for Burlington begin to exceed Norfolk's in the third quarter of 2022 (see Figure 1).

—

Table 2 additionally depicts the results of the Poisson regression model which investigated use of force counts specifically affecting White individuals. Table 2 shows that for the first quarter of 2019, the expected count of use of force incidents affecting White individuals in Norfolk, VA was 17.38. For every additional quarter before the defunding time point, the expected count of incidents affecting White individuals in Norfolk, VA increased by a factor of 1.006. For the first quarter after the defunding time point, the expected count of incidents affecting White individuals in Norfolk was 17.37. For every additional quarter after the third quarter of 2020, the expected count decreased by a factor of 0.907.

For Burlington, Table 2 shows that the expected count of use of force incidents affecting White individuals in the first quarter of 2019 was 35.29. For each additional quarter before the defunding time point, the expected count in Burlington decreased by a factor of 0.969. For the first quarter after the defunding time point, the expected count of incidents affecting White individuals in Burlington, VT was 20.17. For every additional quarter after defunding, the expected count of incidents increased by a factor of 1.046.

The intercepts and factors of change from quarter-to-quarter for the four fitted lines from the Poisson model which focused on White individuals are depicted visually in Figure 2. One can see that the factors of change from quarter-to-quarter for Norfolk and Burlington are relatively similar before the defunding time point (see Appendix A for parallel trends test). After the defunding time point, the factor of change from quarter-to-quarter for Norfolk becomes negative while the factor for Burlington becomes positive.

As was similarly noted with the total use of force incident counts, the difference between Norfolk and Burlington in counts of use of force incidents against White individuals does not line up with the two cities' difference in population (see Tables 1a and 1b). Even with the difference in percentage of White individuals in Burlington compared to Norfolk, Norfolk still has a White population that is approximately 2.5 times the size of Burlington's. Despite Norfolk's larger White population, Burlington consistently has a higher count of use of force incidents against White individuals (see Figure 2).

—

Table 2 also shows the coefficient estimates for the Poisson regression model which investigated counts of use of force incidents affecting Black individuals. Table 2 shows that in Norfolk, VA, the expected count of use of force incidents affecting Black individuals in the first quarter of 2019 was 75.10. For every additional quarter before the defunding time point, the expected count decreased by a factor of 0.992. For the first quarter after the defunding time point, the expected count of incidents in Norfolk was 49.44, and for every additional quarter, that expected count decreased by a factor of 0.935.

Table 2 exhibits that for Burlington, the expected count of use of force incidents affecting Black individuals in the first quarter of 2019 was 11.78. For each additional quarter before the defunding time point, this expected count decreased by a factor of 0.966. For the third quarter of 2020, the expected count of use of force incidents affecting Black individuals in Burlington was 10.26. For each additional quarter after the defunding time point, the expected count decreased by a factor of 0.992.

The factors of change from quarter-to-quarter and intercepts for the four fitted lines from the Poisson model which focused on Black individuals are depicted visually in Figure 3. The factors of change from quarter-to-quarter of the fitted lines for Norfolk and Burlington are very similar before the defunding time point. After the defunding time point, the factor of change from quarter-to-quarter for Norfolk becomes clearly negative while the factor for Burlington remains relatively flat.

Although Burlington generally has smaller counts of use of force incidents against Black individuals, the difference between Burlington's counts and Norfolk's counts is disproportionate to the difference between the size of the two cities' Black populations. Norfolk's counts for 2021 (Quarters 9-12 in Figure 3) hover between 30 and 40 incidents while Burlington's hover slightly above 10. Based on Tables 1a and 1b, Norfolk's Black population in 2021 was around 94270, while Burlington's was around 1900. Although Norfolk's Black population was approximately 49.6 times the size of

Burlington's in 2021, the use of force counts towards Black individuals in Norfolk were only 3 to 4 times the size of Burlington's.

Geospatial Analysis of Use of Force Incidents

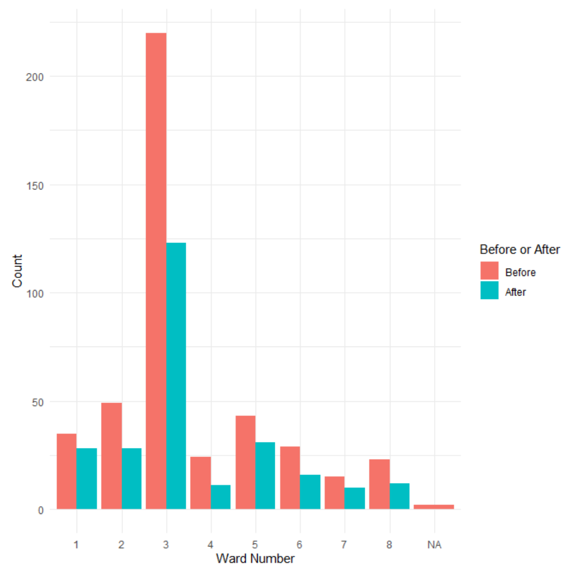


Figure 4: Counts of use of force incidents before and after the defunding of the police in each ward in Burlington

* Wilcox signed rank test with Ho: The number of use of force incidents in wards before defunding equals the number of use of force incidents after, p = 0.01
* Wilcox signed rank test with Ho: The true difference in the use of force counts in wards from before to after is down 20%, p = 0.02

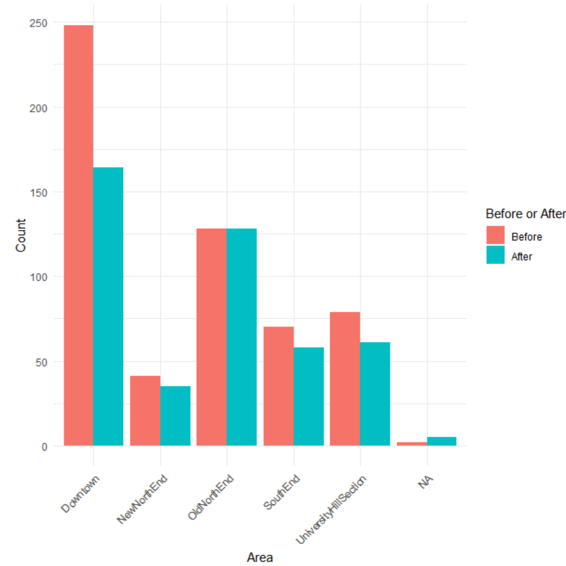


Figure 5: Counts of use of force incidents before and after the defunding of the police in each area in Burlington

* Wilcox signed rank test with Ho: The number of use of force incidents in areas before defunding equals the number of use of force incidents after, p = 0.10
* Wilcox signed rank test with Ho: The true difference in the use of force counts in areas from before to after is down 20%, p = 0.06

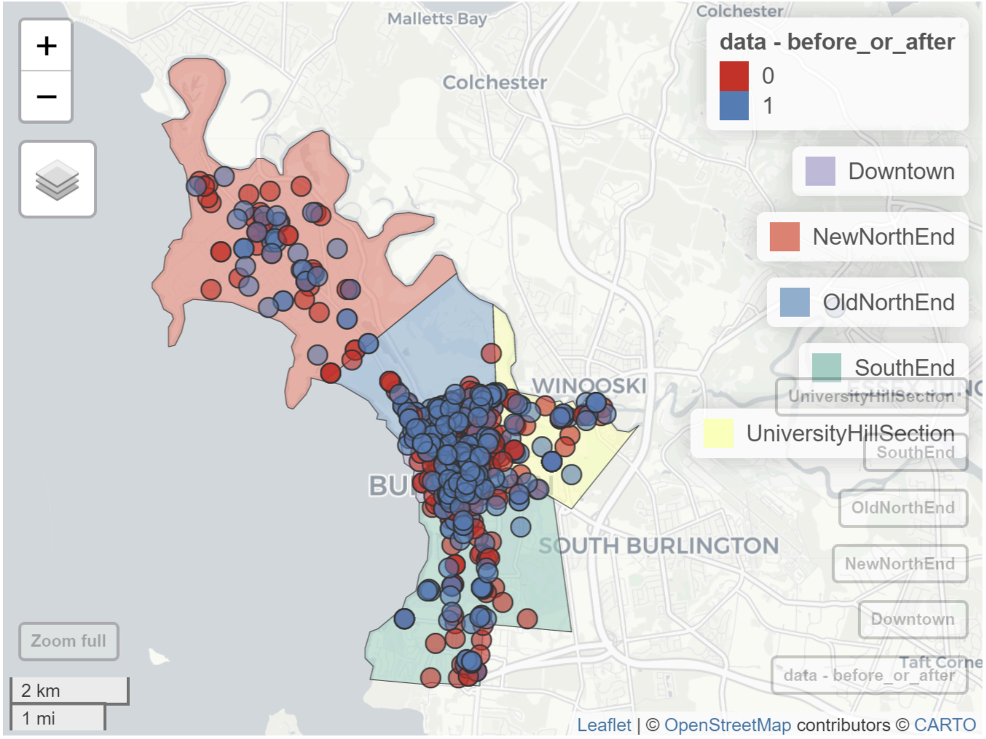


Figure 6: Map of Burlington and the use of force incidents by area, with a hyperlink to interactive map

After running a Wilcox-signed rank test to test the differences in counts of use of force before and after the defunding in the wards, it was found that there was a statistical difference ($p = 0.01$) in the use of force by wards before and after defunding the police. In each ward, the number of use-of-force incidents went down between 20% and 54% after defunding (Figure 4). After running another Wilcox-signed rank test on the areas, there was not a statistical difference ($p = 0.10$) in the use of force before and after defunding the police. All the areas had a decrease in incidents except the Old North End, which had the same number of cases before and after. The biggest decrease was in Downtown, which had a 33.9% decrease (Figure 5).

Overall, the number of incidents of use of force went down 20.5% after defunding. Another Wilcox signed rank test was run to test if the true percent decrease in the use of force counts equals 20% in both all wards combined and all areas combined. It was found that the percent decrease in use of force counts in all wards combined was not equal to 20% ($p = 0.02$), but in all areas combined the percent decrease was equal to 20% ($p = 0.06$).

This analysis might seem contradictory to the previous analysis, but that is because this analysis looks at an even time frame from before and after defunding. There is data from three years before (June 2017-June 30, 2020) and three years after (July 1, 2020 - June 30, 2023). The previous analysis only starts looking at data in 2019. Here we are looking at the total use of force counts for the three years before and the total use of force counts for the three years after, and not the trends. This decrease in the use of force counts comes from the initial decrease after the defunding, which made the overall counts post-defunding lower.

Delving deeper into differences in the wards, it is important to note that the count of use of force incidents is highest in Ward 3 as well as downtown, as shown in Figures 4 and 5. Both areas are where most people are concentrated. Ward 3 includes most of the downtown area, so it makes sense that for before and after it is the area with the most incidents. The Burlington Police Station is also located in Ward 3, and right on the edge of the Old North End and downtown areas. It was originally hypothesized that the proportion of use of force incidents would increase in the areas around the police station, but Ward 3 accounted for 50.2% of use-of-force incidents before defunding, and 47.5% after. The proportion of use of force incidents that occurred Downtown also decreased, but it increased in the Old North End. However, looking at the map of incidents, both before and after defunding, incidents are less concentrated the further from Downtown you get, possibly because this is where the population is less dense, as shown in Figure 6.

<u>Predictive Analysis of Use of Force in Arrests</u>

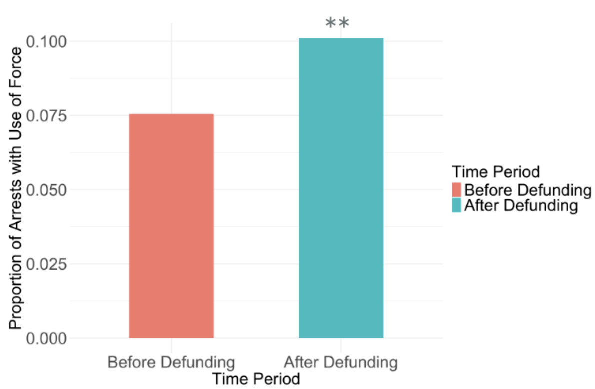


Figure 7: Proportion of arrests that use force before and after defunding the police. ** represents a statistically significant difference.

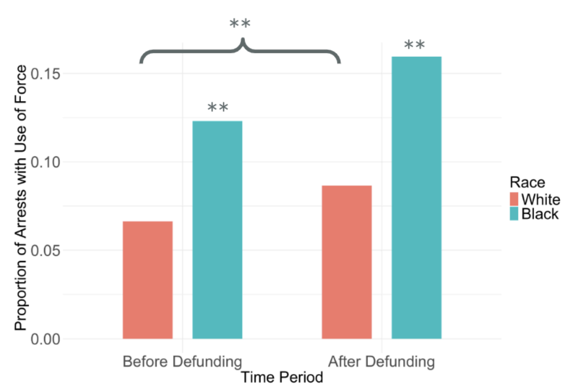


Figure 8: Proportion of arrests with use of force among different races for before and after the defunding of the police. ** represents a statistically significant difference.

After running the descriptive analysis on proportions of arrests that used force before and after June 30, 2020, it was discovered that there was a significant difference in use of force after defunding when compared with before defunding. Before defunding the police, 7.5% of arrests used force. After defunding the police, 10.1% of arrests used force. Using a two-sample two-sided proportion test, a p-value of 0.043 ($<0.05$) was calculated. The proportion of use of force after defunding the police is greater than the proportion of use of force before defunding the police. (Figure 7).

To determine if relationships exist between demographic variables and use of force proportional to arrests, chi square tests were conducted. For both before and after defunding, age and gender did not yield significant p-values ($p > 0.05$). However, for both time periods there was an association between race and use of force ($p = 0.0017$ [before], $p < 0.0001$ [after]). To delve deeper into this relationship, prior to police defunding, the percentage of arrests involving force among Black individuals stood at 12.3%, contrasting with 6.63% among White individuals. Using a two-sample proportion test, a p-value of 0.0011 ($<0.05$), indicates there is a significant difference in the proportion of use of force in arrests among Black people when compared with White people. After defunding the police, the proportion of arrests that used force in Black individuals was 15.91% while it was 8.65% in White individuals. A p-value $< 0.001$ ($<0.05$) indicates there is also a significant difference in the proportion of use of force in arrests among Black people when compared with White people after defunding the police. To determine if defunding the police affected the proportions of use of force in Black people, another two-sample proportion test was conducted. At a p-value of 0.177 ($>0.05$), there is no significant difference in the proportion of use of force in Black individuals before vs. after defunding the police. To determine if defunding the police affected the proportions of use of force in White individuals, a p-value of 0.033 ($<0.05$) was calculated. The proportion of use of force in White individuals is significantly higher after vs. before defunding the police. These relationships can be seen in Figure 8. Since the sample size for Other (Alaska Native or American Indian) is fairly small for both datasets, it was excluded from the analysis.

Table 3: Coefficient estimates and std. Error for logistic regression model for data **before** and **after** defunding the police. Models fit to predict use of force based on age, gender, and race. (*,**, and *** indicates statistical significance at 0.05, 0.01, and 0.001 level respectively)

| Term | Coef. Estimate (**Before**)<br>Std. Error in parenthesis | Coef. Estimate (**After**)<br>Std. Error in parenthesis |
|---|---|---|
| B0 Intercept | -2.8547*** (0.24064) | -2.33817*** (0.16080) |
| B1 raceBlack | 0.62320* (0.24514) | 0.65707*** (0.17035) |
| B2 raceOther | -0.42258 (0.74167) | 0.12940 (0.41083) |
| B3 genderMale | 0.10481 (0.25265) | 0.03545 (0.17851) |
| B4 genderOther | -12.71133 (485.13257) | 0.30173 (0.64169) |
| B5 ageYoung | 0.39047 (0.2233) | -0.03903 (0.16118) |
| B6 ageElderly | 0.01013 (0.37833) | -0.26086 (0.25411) |

Table 4: Confusion Matrix displaying observed vs. predicted use of force in the testing set <u>before</u> defunding the police. Predictions made from the logistic model predicting use of force by age, race, and gender.

| | Observed No Use of Force | Observed Use of Force |
|---|---|---|
| Predicted No Use of Force | 405 | 33 |
| Predicted Use of Force | 0 | 0 |

Table 5: Confusion Matrix displaying observed vs. predicted use of force in the testing set <u>after</u> defunding the police. Predictions made from the logistic model predicting use of force by age, race, and gender.

| | Observed No Use of Force | Observed Use of Force |
|---|---|---|
| Predicted No Use of Force | 644 | 72 |
| Predicted Use of Force | 0 | 0 |

While Chi-square tests did not uncover any associations between use of force and age or gender, a point of focus remained on whether demographic factors could serve as predictors for use of force. Regression coefficients for a logistic model with training data before the defunding was calculated (Table 3). In this model, the reference variable across the three predictors is White people, females, and middle-aged individuals. The only statistically significant variable was Black individuals. Holding values for age, and gender constant, the odds of use of force being administered was 86.49% higher for Black individuals than it was for White individuals.

The model's Goodness-of-Fit and the explanatory power of its predictor variables were assessed using a Hosmer-Lemeshow Goodness-of-Fit test and by calculating McFadden's R-squared metric. The Goodness-of-Fit test yielded a p-value effectively 0, indicating a lack of fit in the model. Moreover, McFadden's R-squared value revealed that 26.43% of the variation in the model can be explained by gender, age, and race.

To evaluate the model's accuracy, predictions were made on a testing set comprising 488 arrests using a threshold of 0.5. Analysis of the confusion matrix (Table 4) showed a recall value of 0 and a specificity value of 1.0. These findings suggest that the model encountered difficulties in accurately predicting arrests involving the use of force based solely on demographic factors. Furthermore, the area under the ROC curve was 0.6086, indicating poor predictive/discriminatory performance of the model.

A second logistic regression model was performed on training data after defunding the police to determine if predictive power of demographic factors on use of force differed between datasets. Again, regression coefficients were outputted using reference levels of White for race, middle-aged for age, and female for gender (Table 3). Similar to before defunding the police, the only statistically significant variable was Black people. Holding values for age, and gender constant, the odds of use of force being administered was 92.91% higher for Black people than it was for White people.

Results from the Goodness-of-Fit test were comparable to the first model yielding a p-value of effectively 0, indicating a lack of fit in the model. Additionally, McFadden's R-squared value calculated that only 1.22% of the variation in the model can be explained by gender, age, and race.

Predictions were made on a testing set comprising 716 arrests using a threshold of 0.5. Analysis of the confusion matrix (Table 5) showed a recall value of 0 and a specificity value of 1.0. These findings line up with our previous model and suggest that the model encountered difficulties in accurately predicting arrests involving the use of force based solely on demographic factors. Furthermore, the area under the ROC curve was 0.5728, indicating very poor predictive performance of the model.

Predictive Analysis of Type of Force by Time Period

Table 6: Weaponless force before and after 2020 Q2 with predictors, coefficient estimate, and p-value for the full logistic regression model (*indicates statistical significance at 0.05 level)

| Variable | Coefficient Estimate (**Before**) P-value in parentheses | Coefficient Estimate (**After**) P-value in parentheses |
|---|---|---|
| Age | -0.0086286 (0.53020135) | 0.0098181 (0.5702412) |
| Race: Black | -0.0347296 (0.93073734) | -0.0921321 (0.8243145) |
| Gender: Male | -0.4316475 (0.29183460) | 0.0696276 (0.8806869) |
| Alcohol Presence | 0.8608599 (0.01310049)* | 0.0267487 (0.9544496) |
| Drugs Presence | 0.8215021 (0.16755755) | 1.2706421 (0.1120400) |
| Emotional Distress Presence | -0.0347929 (0.92219771) | 0.2591942 (0.5949803) |
| Subject Action: Assaultive | 0.0237043 (0.94776361) | -0.4697026 (0.2799833) |
| Subject Action: Compliant | -18.3230931 (0.98342070) | -3.1005945 (<0.0001)* |
| Subject Action: Deadly Force | -1.6943492 (0.1586828) | 14.2045243 (0.9889427) |
| Subject Action: Passive Resistance | -0.1786818 (0.7950789) | -1.8570438 (0.00051417)* |
| Subject Action: Verbal Non-Compliance | -1.4194359 (0.2340505) | -2.7236296 (0.0173345)* |

Table 7: Firearm Pointed force before and after 2020 Q2 with predictors, coefficient estimate, and p-value for the full logistic regression model (*indicates statistical significance at 0.05 level)

| Variable | Coefficient Estimate (**Before**) P-value in parentheses | Coefficient Estimate (**After**) P-value in parentheses |
|---|---|---|
| Age | 0.0306312 (0.09999966) | 0.0079529 (0.6970052) |
| Race: Black | 0.1890180 (0.7153514) | 0.4239175 (0.4002906) |
| Gender: Male | 1.8790139 (0.01197150)* | 0.5714166 (0.3373188) |
| Alcohol Presence | -0.8182395 (0.08127491) | -0.2288084 (0.6739988) |
| Drugs Presence | 0.2856203 (0.6660936) | 1.2556081 (0.04842239)* |
| Emotional Distress Presence | -0.2478469 (0.6094473) | 0.4013438 (0.4699297) |
| Subject Action: Assaultive | -1.3319382 (0.04643073)* | -0.2248996 (0.7005924) |
| Subject Action: Compliant | 3.5348417 (< 0.00001)* | 4.9879445 (< 0.00001)* |
| Subject Action: Deadly Force | 3.6959295 (0.008185940)* | 2.1719002 (0.1472691) |
| Subject Action: Passive Resistance | 2.2438294 (0.003263851)* | 2.7307734 (< 0.00001)* |
| Subject Action: Verbal Non-Compliance | 0.0395828 (0.9740687) | 2.1458144 (0.02140614)* |

After performing the logistic model, several conclusions can be made. Table 6 depicts the results for the logistic regression full model in which the outcome being investigated is Weaponless force before 2020 Q2. The only statistically significant predictor in this model is alcohol presence. It can be concluded that holding all other predictors constant, the odds of weaponless force being used is 2.3652 times higher when alcohol is involved compared to when alcohol is not involved.

Table 6 also depicts the results for the logistic regression full model in which the outcome being investigated is Weaponless force after 2020 Q2. There are several statistically significant predictors, which are all subject actions: Compliant, Passive Resistance, and Verbal Non-Compliant. It can be concluded that holding all other predictors constant, the odds of experiencing weaponless force for compliant individuals are 0.04502 times lower than for non-compliant individuals. In addition, it can be concluded that holding all other predictors constant, the odds of experiencing weaponless force for individuals passively resisting are 0.1561 times lower than for those who are not passively resisting. Lastly, holding all other predictors constant, the odds of experiencing weaponless force for verbal non-compliant individuals are 0.0656 times lower than for those who are not verbal non-compliant.

Table 7 depicts the results for the logistic regression full model in which the outcome being investigated is firearm pointed before 2020 Q2, there are several statistically significant predictors, which are gender, and the subject actions; Assaultive, Compliant, Deadly Force, and Passive Resistance. It can be concluded, holding all other predictors constant, that the odds of a firearm being pointed are 6.5471 times higher for men than they are for women. In addition, holding all other predictors constant, the odds of experiencing firearm pointed force for assaultive individuals are 0.263965 times lower than for non-assaultive individuals. Furthermore, holding all other predictors constant, the odds of experiencing force for compliant individuals are 34.2896 times higher than for non-compliant individuals. Additionally, holding all other predictors constant, the odds of a firearm being pointed for individuals who used deadly force are 40.283 times higher than for those who did not use deadly force. Lastly, It can be concluded that holding all other predictors constant, the odds of a firearm being pointed for individuals passively resisting are 9.4294 times higher than for those who are not passively resisting.

Table 7 also depicts the results for the logistic regression full model in which the outcome being investigated is firearm pointed after 2020 Q2. Similar to before defunding, there are several statistically significant predictors being Drug presence, and the subject actions; Compliant, Passive Resistance, and Verbal Non-Compliance. It can be concluded, holding all other predictors constant, that the odds of a firearm being pointed for individuals under the influence of drugs are 3.5099 times higher than those not using drugs. Additionally, holding all other predictors constant, the odds of experiencing force for compliant individuals are 146.635 times higher than for non-compliant individuals. Moreover, holding all other predictors constant, the odds of a firearm being pointed for individuals passively resisting are 15.344 times higher than for those who are not passively resisting. Finally, holding all other predictors constant, the odds of experiencing weaponless force for verbal non-compliant individuals are 8.549 times higher than for those who are not verbal non-compliant.

To evaluate the models' accuracy and fit, collinearity among the predictor variables was tested. In addition, a Hosmer-Lemeshow Goodness-of-Fit and AUROC statistic was computed. A Chi-squared test was performed looking for an association between different predictors and the two most prevalent types of force, weaponless and firearm pointed. If an association was found, potential collinearity was investigated. In the firearm pointed before 2020 Q2 model, there was evidence of collinearity of subject action on alcohol. This was the only evidence of collinearity among the models. Interaction terms, such as subject race times subject action, alcohol times subject action, and subject gender times subject action, were added to the model, but they were not statistically significant.

Table 8: Output from the Hosmer-Lemeshow Goodness-of-Fit Test.

| Model | P-value |
|---|---|
| Weaponless Before 2020 Q2 | 0.7326 |
| Weaponless After 2020 Q2 | 0.6567 |
| Firearm Pointed Before 2020 Q2 | < 0.0001 |
| Firearm Pointed After 2020 Q2 | 0.828 |

It is important to analyze the fit of the logistic models used, otherwise any conclusions drawn from statistical analyses are not applicable. To prove the effectiveness of how well the logistic model fit the observed data, Hosmer-Lemeshow Goodness-of-Fit Tests were conducted. A p-value greater than 0.05 indicates that there is no statistically significant lack of fit, therefore the model fits the data. As shown in Table 8, all logistic models fit well, with an exception being Firearm Pointed Before 2020 Q2. This is unsurprising considering in earlier analysis it was discovered that subject action and alcohol are confounded with each other. To further measure the performance of the logistic models, the Area Under the Receiver Operating Characteristic Curve, AUROC, was calculated. For weaponless force before defunding the police, the AUROC for the full model is 0.7521, indicating that the model has low to moderate predictive discriminatory power. Additionally, for weaponless force after 2020, the AUROC for the full model is 0.8069, indicating that the model has moderate predictive discriminatory power. Lastly, for firearm pointed force both before and after 2020, the AUROC for the full models are 0.873 and 0.8899, respectively. These indicate that both models have high predictive discriminatory power.

# Discussion

Through the multiple analyses conducted, many of the same insights on the impact of Burlington's defunding can be drawn. Although Burlington's defunding of the police had good intentions, the policy did not appear to lead to a lasting reduction in use of force incidents or alleviate use of force or arrest disparities amongst racial groups. While there was an initial decrease in use of force incidents immediately after defunding occurred in Burlington, the segmented Poisson analysis showed that this decrease was not sustained. By the second quarter of 2023, Burlington's use of force counts returned to the same level they had been pre-defunding. Had Burlington not defunded the police, one would have expected a continuous decrease in use of force counts similar to the trend seen in Norfolk, VA. Since

Burlington's use of force counts eventually returned to the same level as pre-defunding, it appears that Burlington's defunding did not have the reduction effect on use of force counts that was originally intended. Additionally, Burlington's defunding did not lead to a sustained reduction in use of force counts for either White or Black individuals, yet again showing that the policy did not have its intended effect. Furthermore, proportional analyses from arrest data also reveal that defunding did not lead to any changes in Burlington's disproportionate use of force against Black people compared to White people, showing again that the policy failed to alleviate disparities among racial groups. When exploring the changes in use of force counts in different regions of Burlington, geospatial analysis uncovered a decrease in counts of use of force incidents in all wards and almost all areas of Burlington after the police were defunded. It is important to note that there was a significant difference in the use of force in wards but not in areas. This finding may be because wards are constructed politically, and are used as voting districts. Further research investigating possible gerrymandering, where individuals with similar characteristics are lumped together into wards, could explain the statistically significant decrease in use of force counts found in all wards. The redistricting of wards that was done in 2023 may have been a viable solution to balance these disparities, but further analysis would be needed to investigate the effects of the redistricting as data from post-redistricting was not included in this analysis. There was also less data used when analyzing the wards, which could also contribute to the fact that the differences in wards after defunding were found to be statistically significant. It is important to note that the apparent contradiction in results between the geospatial analysis and the comparative analysis with Norfolk may be due to the inclusion of incident counts from 2017-2018 which were not used in segmented Poisson regression models. Notably, a geospatial analysis of trends as a function of time was not explored. The geospatial analysis only looked at the total counts from the three years before defunding compared to the total counts in the three years after defunding in different regions. The reduction in use of force counts was a result of the initial decrease in use of force incidents post-defunding, so while overall there are less use of force counts after defunding, the decrease was not sustained. By 2023, counts returned to the same as pre-defunding. This would explain why the results of the geospatial analysis appear to contradict those of the Poisson analysis.

To further enforce points made in the count analyses, logistic regression models also gave reason to believe that Burlington's defunding measures were ineffective. In the model which investigated Burlington arrest data, race was a significant factor in predicting the presence of use of force in an arrest both before and after defunding. This analysis revealed that the odds of force being used on Black individuals was higher when compared to White individuals. This finding highlights the persistent disparity in the use of force across racial demographics. While the statistical significance of race as a predictor highlights broader systemic issues, the models' lack of fit and predictive power also emphasizes the complexity involved in understanding the dynamics of use of force within law enforcement interactions. In contrast to race being a significant predictor in the predictive analysis of the general presence of force, one finding is the insignificance of subject age, race, and gender on weaponless and firearm pointed force incidents, both before and after the defunding of the police. This highlights the possibility that alternate factors impact weaponless and firearm pointed force in particular. For both types of force, incidents involving certain subject actions were at an increased risk of use of force after the defunding of police, providing further evidence of the possibility that the goal of defunding the police to reallocate resources towards de-escalation measures was ineffective in reducing risk of use of force.

One hypothesis for why Burlington's defunding of the police did not lead to improvements in use of force or racial disparities could be that the policy caused additional issues with the quality of police work. As mentioned earlier, although Burlington passed a resolution on June 30th, 2020 which led to a reduction in officers from 95 to 64, it was not until six months later that the community service officer and community support liaison positions were created. Thus, there were multiple months where the Burlington police force was acting with a reduced staff and no new social service positions to make up for it. One can hypothesize that in this situation, a police force may not have been able to perform optimally. This situation likely also contributed to officers feeling undervalued within their community (Siegel, 2021), which could explain why Burlington's defunding of the police did not lead to a reduction in use of force incidents nor alleviate racial disparities in use of force or arrests.

One may ask why it is that Norfolk, VA experienced a decrease in incidents despite not having a treatment applied to their police force. It is important to note that the Black Lives Matter protests following George Floyd's death had occurred throughout June 2020. The decrease in use of force incidents in Norfolk, VA despite a lack of treatment could be partially attributed to the aftermath of these protests. If Norfolk's decrease in counts was indeed due to the protests, one could infer that Burlington's struggles with staffing and lack of social service positions were enough to counteract the effects of the same protests on Burlington's use of force counts. It is important to note that all of these explanations are only hypotheses. Further research would need to be done on the effects of both the Black Lives Matter protests and the defunding of police on use of force incidents and racial disparities in other cities to unearth any insights on common trends.

# Conclusions

Overall, Burlington's defunding of the police did not have the intended effects of reducing use of force incidents and alleviating disparities in use of force and arrests amongst racial groups. The defunding event appeared to have led to an initial decrease in use of force incidents, but by the middle of 2023, use of force counts reverted back to their previous state from before defunding. Additionally, the disproportionate use of force in arrests against Black individuals compared to White individuals persisted after defunding. Lastly, the risk of both weaponless and firearm-pointed force based on certain subject actions increased after defunding. In short, Burlington's defunding of the police was generally ineffective at achieving its original goals.

It should be noted that there were a few key limitations to this research. The logistic regression models which used Burlington arrest data found race to be a significant predictor, but one must be careful drawing conclusions from these results since both models exhibited poor fit and lacked the capability to accurately predict instances of excessive force. Similarly, it is also important to note the poor fit with the logistic regression model which looked at relationships between subject characteristics and firearm pointed force before police defunding. Yet again, this poor fit should be taken into account when drawing conclusions from the model.

In regards to the Poisson regression models, using Norfolk, VA as a control city poses a few issues. As explained in the methods section, Norfolk, VA was chosen as the control city because it was the only city found which had not defunded the police and had a public use of force dataset. Unfortunately, Norfolk, VA and Burlington, VT are not incredibly similar. While they are both liberal cities on the east coast, Norfolk, VA is far larger and far more diverse. Since Norfolk is not the perfect control group for Burlington, the trends that occurred in Norfolk may not directly match up with the trends that would have hypothetically occurred in Burlington had the city not defunded the police. Although many conclusions were made from this study, additional research is needed to investigate if the observed trends were also seen in other cities throughout the United States. One possible way further research could be carried out is to examine use of force and arrest trends in additional control and defunding-treated cities. If similar findings were uncovered in additional cities, one could be more certain that the conclusions drawn from this research reflect the true effects of defunding the police.

# References

City of Burlington. (2023). *Use of force* [Data set]. https://data.burlingtonvt.gov/datasets/9dd9dca94be44e86b0c8f8dd7db5d4a4_0/explore

City of Burlington. (2024). *Arrests* [Data set]. https://data.burlingtonvt.gov/datasets/2781dfa492424452a5f498e12436562a_0/explore

City of Burlington (n.d.) 2022 Redistricting Retrieved April 14, 2024, from https://www.burlingtonvt.gov/cityplanning/btvstat/redistricting

Data Commons. (n.d.). *Statistical Variable: Count_person, About: Burlington, VT*. Retrieved April 14, 2024, from https://datacommons.org/browser/geoId/5010675?statVar=Count_Person

Data USA. (n.d. (a)). *Data USA: Burlington, VT*. Retrieved April 14, 2024, from https://datausa.io/profile/geo/burlington-vt/

Data USA. (n.d. (b)). *Data USA: Norfolk, VA*. Retrieved April 14, 2024, from https://datausa.io/profile/geo/norfolk-va

Norfolk Police Department. (2024). *Police use of force and citizen complaint incidents* [Data set]. https://data.norfolk.gov/Public-Safety/Police-Use-of-Force-and-Citizen-Complaint-Incident/fcqe-uvnb/about_data

Norfolk Budget & Strategic Planning Department. (n.d.). *Budget documents.* The City of Norfolk. https://www.norfolk.gov/908/Budget-Documents

Siegel, E. R., Rappleye, H. (2021, December 19). Burlington decided to cut its police force 30 percent. Here's what happened next. *NBC News.* https://www.nbcnews.com/news/us-news/burlington-vermont-defunded-police-force-s-happened-rcna8409

World Population Review. (n.d.). *Norfolk, Virginia Population*. Retrieved April 14, 2024, from https://worldpopulationreview.com/us-counties/va/norfolk-city-population

# Author Contributions

Annika

Analyzed proportions of use of force across individual factors in arrest data. Performed logistic regression modeling to predict use of force in an arrest based on the individual's age, race, and gender. Differences in coefficients for two time period models were used to assess the impact that defunding the police had.

Isaac
Researched and chose Norfolk, VA as control group. Performed data cleaning on Norfolk, VA use of force dataset. Built segmented Poisson regression models for overall use of force incident counts along with for Black and White individuals only. Compared the outcome of the defunding the police on incident counts in Burlington, VT to trends in Norfolk, VA (where defunding did not occur) over the same time period.

Abby
Did data cleaning to make the data usable for analysis. Created a map of Burlington to better visualize the uses of force before and after the defunding. Did analysis of the wards and areas to see if there was a statistical difference.

Lucy
Performed general analysis of use of force data before and after the defunding of police. Conducted logistic models for weaponless and firearm pointed use of force before and after 2020 to predict use of force based on age, race gender, substance use, emotional distress, and subject actions. Investigated associations and potential collinearity among the variables.

# Acknowledgements

We would like to express our gratitude to Dr. Jeff Buzas, for his statistical expertise and guidance throughout this project. We would also like to thank Dr. Maria Sckolnick for her invaluable help in providing statistical consulting services. This research would not have been possible without them.

# Appendix A

**Segmented Poisson Assumption Checking**
For the segmented Poisson regression models, both the parallel trends assumption and the Poisson regression assumptions were checked. For Poisson regression, the following five assumptions outlined were evaluated:

1. Dependent variable consists of counts
2. One or more independent variables
3. Independence of observations
4. Distribution of counts follow a Poisson distribution

5. Mean and variance of model are identical

To check the fourth assumption, four separate graphs were created. These four graphs compared the similarity of the Poisson model's predicted counts to the observed counts for each quarter. There were four total graphs which corresponded to both Norfolk, VA and Burlington, VT before and after the defunding time point. To check the fifth assumption, the dispersiontest function from R's AER library was used.

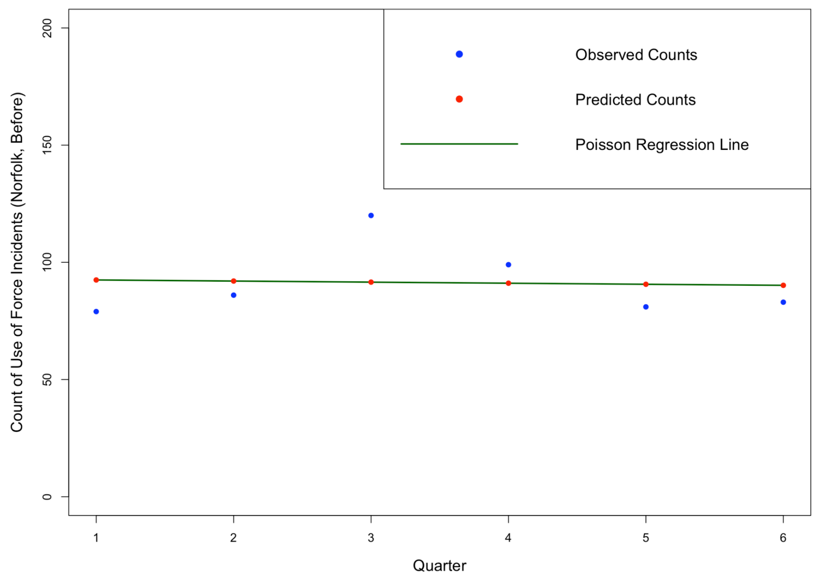


Figure 9a: Predicted counts from overall Poisson model compared to observed counts for Norfolk, VA before the defunding time point

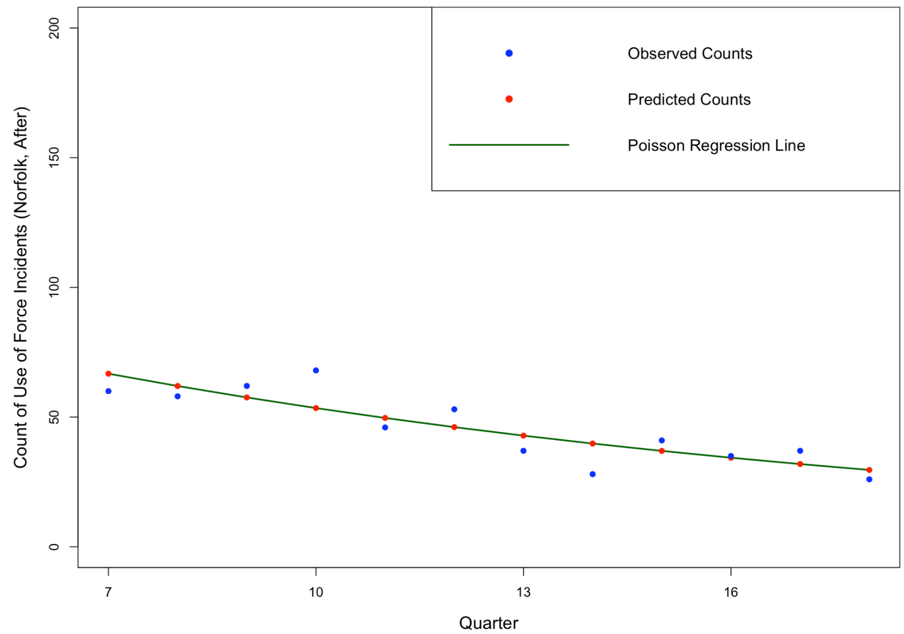


Figure 9b: Predicted counts from overall Poisson model compared to observed counts for Norfolk, VA after the defunding time point

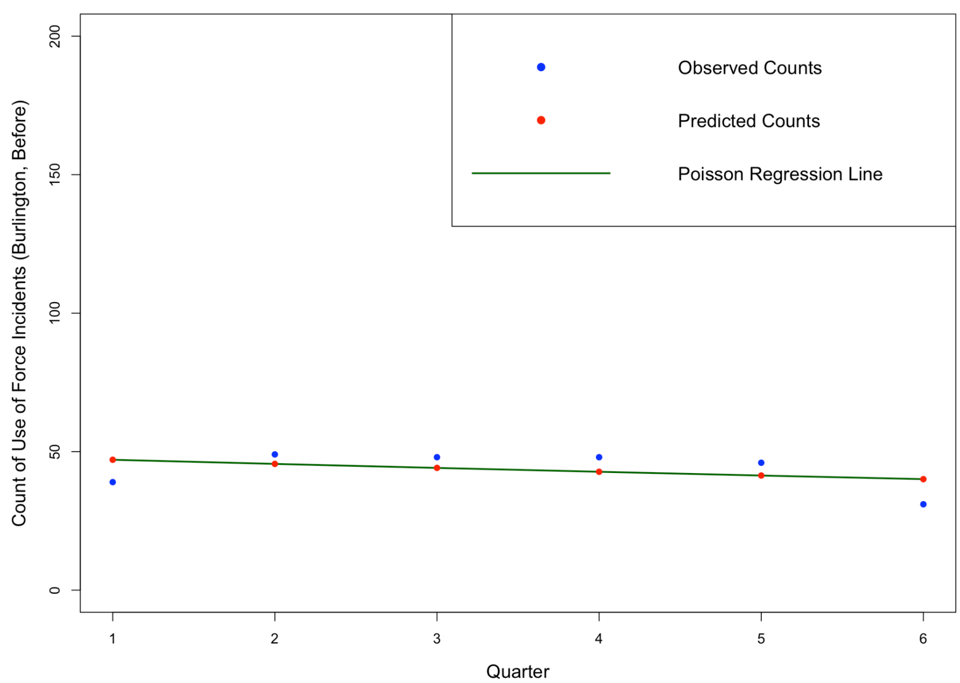


Figure 9c: Predicted counts from overall Poisson model compared to observed counts for Burlington, VT before the defunding time point

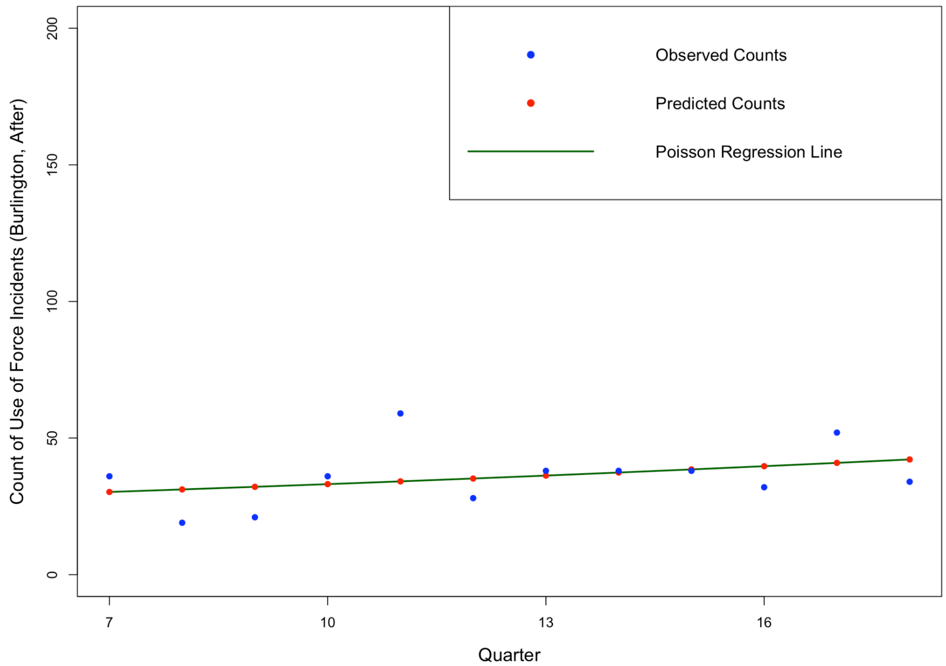


Figure 9d: Predicted counts from overall Poisson model compared to observed counts for Burlington, VT after the defunding time point

Before running any of the Poisson models, it was first necessary to check both the parallel trends assumption and Poisson regression assumptions for each model. To check the parallel trends assumption, the coefficient of interest was B6, the interaction between City and Quarter. This coefficient represents the difference between the pre-defunding slope for Norfolk, VA compared to Burlington, VT. As seen in Table 9 (see Appendix B), for the overall Poisson model (not separated by demographics), the coefficient for B6 was -0.027199 and was not statistically significant (p=0.53708). Thus, there was no evidence that B6 was not equal to 0 and the parallel trends assumption was satisfied.

In regards to Poisson regression assumptions, the first two are clearly satisfied as the dependent variable is the count of use of force incidents and there are three independent variables, City, Quarter, and BeforeAfter. The third assumption, independence of observations, was said to be satisfied but it was noted that quarterly incident counts are likely not entirely independent as one quarter's incident count is often similar to the previous quarter. The fourth assumption was evaluated using Figures 9a-d. The fourth assumption appears to be satisfied as the predicted and observed counts are generally very similar with the exception of a couple slightly larger differences. These slightly larger differences include quarter 3 for Norfolk, VA (Figure 9a) and quarter 11 for Burlington, VT (Figure 9d). Finally, the fifth assumption was satisfied as there was not statistically significant evidence of overdispersion at the alpha=.05 level (p=0.053).

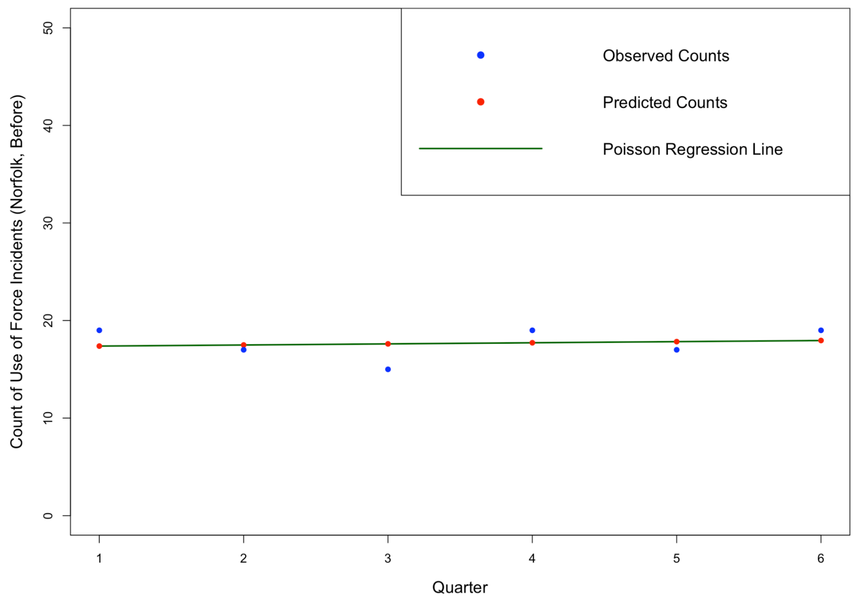


Figure 10a: Predicted counts from Poisson model (for White individuals) compared to observed counts for Norfolk, VA before the defunding time point

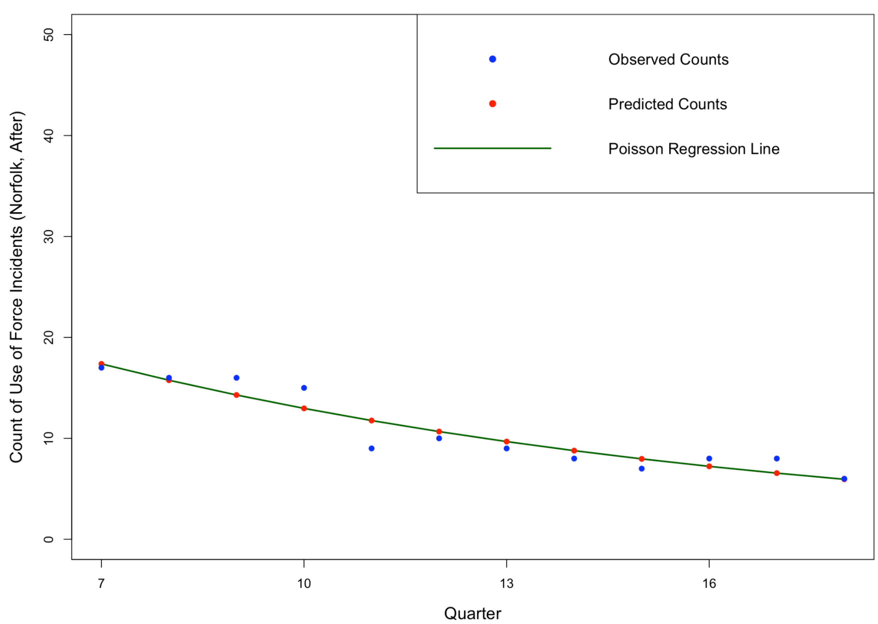


Figure 10b: Predicted counts from Poisson model (for White individuals) compared to observed counts for Norfolk, VA after the defunding time point

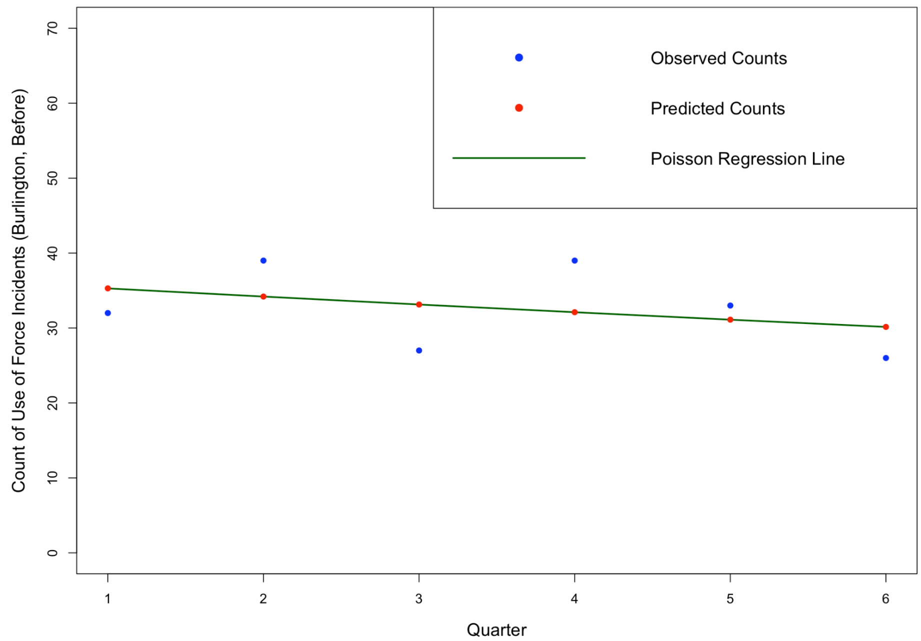


Figure 10c: Predicted counts from Poisson model (for White individuals) compared to observed counts for Burlington, VT before the defunding time point

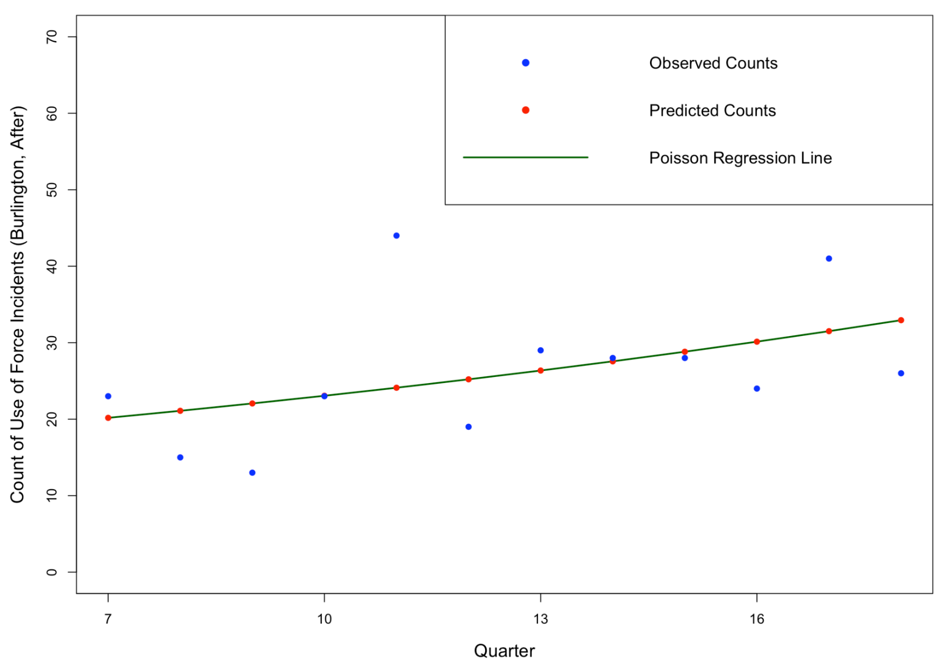


Figure 10d: Predicted counts from Poisson model (for White individuals) compared to observed counts for Burlington, VT after the defunding time point

As seen in Table 10 (see Appendix B), for the Poisson model subsetted to only White individuals, the coefficient for B6 was -0.037975 and was not statistically significant (p=0.590757). Thus, there was no evidence that B6 was not equal to 0 and the parallel trends assumption was satisfied.

For the Poisson regression assumptions, the first two were satisfied as the dependent variable is the count of use of force incidents for White individuals and there are three independent variables, City, Quarter, and BeforeAfter. The third assumption, independence of observations, was again said to be satisfied with the same noted limitations as there were for the overall Poisson model. The fourth assumption was evaluated using Figures 10a-d and appears to be satisfied as the predicted and observed counts are similar with the exception of a couple larger differences. These larger differences include quarters 9, 11, and 16 for Burlington, VT (Figure 10d). Finally, the fifth assumption was satisfied as there was not statistically significant evidence of overdispersion at the alpha=.05 level (p=0.4863).

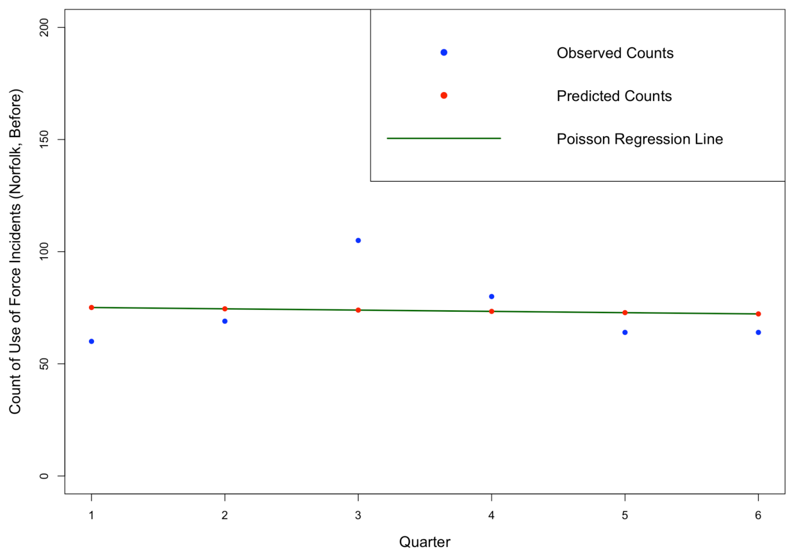


Figure 11a: Predicted counts from Poisson model (for Black individuals) compared to observed counts for Norfolk, VA before the defunding time point

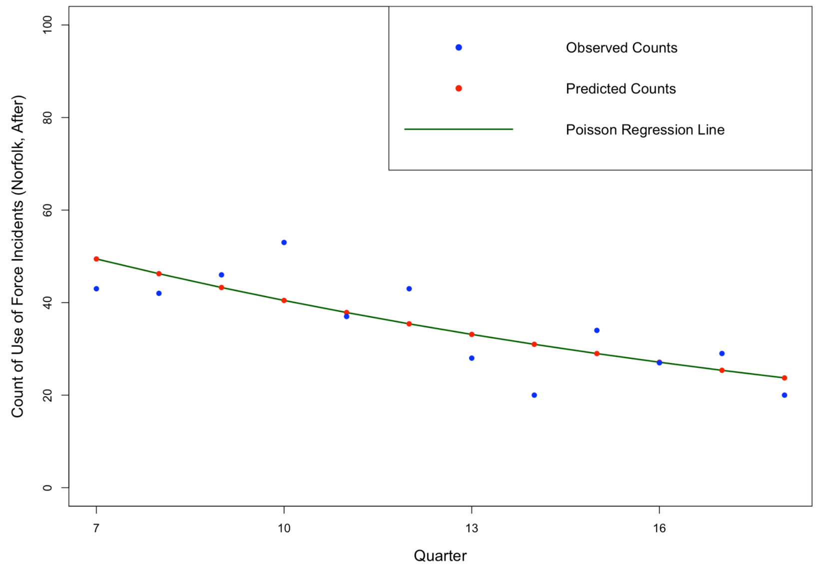


Figure 11b: Predicted counts from Poisson model (for Black individuals) compared to observed counts for Norfolk, VA after the defunding time point

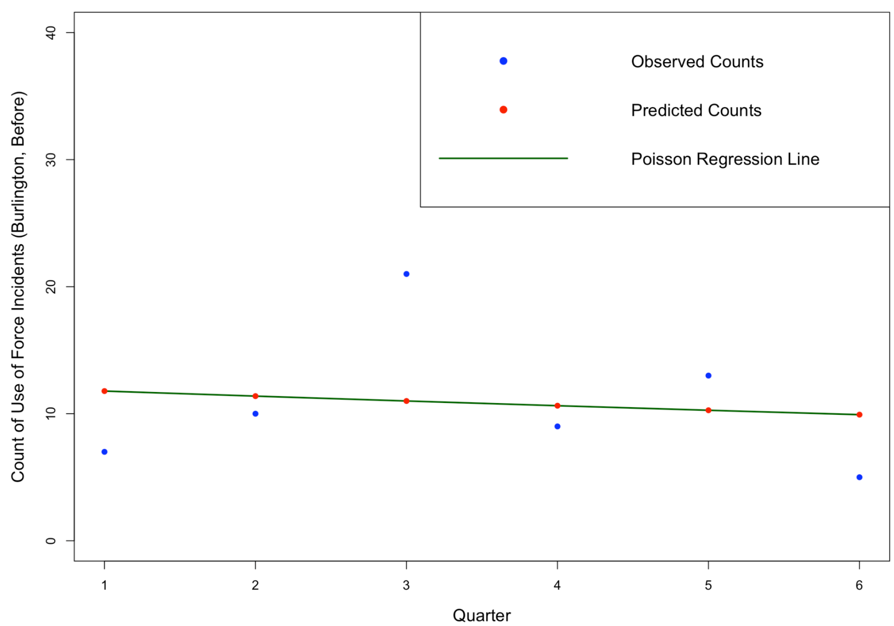


Figure 11c: Predicted counts from Poisson model (for Black individuals) compared to observed counts for Burlington, VT before the defunding time point

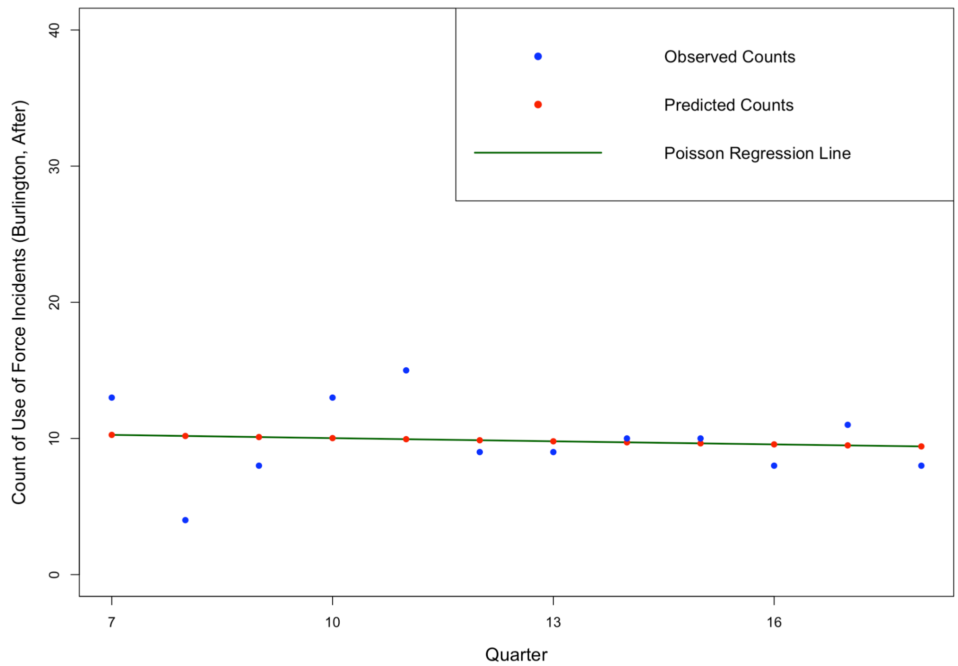


Figure 11d: Predicted counts from Poisson model (for Black individuals) compared to observed counts for Burlington, VT after the defunding time point

As seen in Table 11 (see Appendix B), for the Poisson model subsetted to only Black individuals, the coefficient for B6 was -0.026553 and was not statistically significant (p=0.7331). Thus, there was no evidence that B6 was not equal to 0 and the parallel trends assumption was satisfied.

For the Poisson regression assumptions, the first two were satisfied as the dependent variable is the count of use of force incidents for White individuals and there are three independent variables, City, Quarter, and BeforeAfter. The third assumption, independence of observations, was again said to be satisfied with the same noted limitations as there were for the overall and White Poisson models. The fourth assumption was evaluated using Figures 11a-d and appears to be satisfied as the predicted and observed counts are generally similar with the exception of a couple larger differences. These larger differences include quarters 3, 10, and 14 for Norfolk, VA (Figures 11a-b) and quarters 3 and 8 for Burlington, VT (Figure 11c-d). Finally, the fifth assumption was satisfied as there was not statistically significant evidence of overdispersion at the alpha=.05 level (p=0.08881).

# Appendix B

**Coefficient Estimates for Segmented Poisson Models**

Table 9: Coefficient estimates, std. error, and p-values for overall segmented Poisson model (** indicates statistical significance at 0.05 level)

| Term | Coef. Estimate | Std. Error | P-Value |
|---|---|---|---|
| $B_0$: Intercept | 4.526993 | 0.075431 | < 2e-16 ** |
| $B_1$: City | -0.675232 | 0.130924 | 2.5e-07 ** |
| $B_2$: Quarter | -0.005005 | 0.025014 | 0.8414 |
| $B_3$: BeforeAfter | 0.116605 | 0.159389 | 0.4644 |
| $B_4$: City * BeforeAfter | -0.739267 | 0.258334 | 0.0042 ** |
| $B_5$: Quarter * BeforeAfter | -0.068824 | 0.028001 | 0.0140 ** |
| $B_6$: City * Quarter | -0.027199 | 0.044066 | 0.5371 |
| $B_7$: City * Quarter * BeforeAfter | 0.131176 | 0.047919 | 0.0062 ** |

Table 10: Coefficient estimates, std. error, and p-values for Poisson model looking at **White** individuals only (** indicates statistical significance at 0.05 level)

| Term | Coef. Estimate | Std. Error | P-Value |
|---|---|---|---|
| $B_0$: Intercept | 2.855446 | 0.173082 | < 2e-16 ** |
| $B_1$: City | 0.708228 | 0.212656 | 0.0009 ** |
| $B_2$: Quarter | 0.006469 | 0.056875 | 0.9094 |
| $B_3$: BeforeAfter | 0.584366 | 0.335365 | 0.0814 |
| $B_4$: City * BeforeAfter | -1.411270 | 0.412653 | 0.0006 ** |
| $B_5$: Quarter * BeforeAfter | -0.103936 | 0.062695 | 0.0974 |
| $B_6$: City * Quarter | -0.037975 | 0.070620 | 0.5908 |
| $B_7$: City * Quarter * BeforeAfter | 0.180026 | 0.077170 | 0.0197 ** |

Table 11: Coefficient estimates, std. error, and p-values for Poisson model looking at **Black** individuals only (** indicates statistical significance at 0.05 level)

| Term | Coef. Estimate | Std. Error | P-Value |
|---|---|---|---|
| $B_0$: Intercept | 4.318856 | 0.083808 | < 2e-16 ** |
| $B_1$: City | -1.852168 | 0.229905 | 7.87e-16 ** |
| $B_2$: Quarter | -0.007757 | 0.027853 | 0.7806 |
| $B_3$: BeforeAfter | -0.017790 | 0.181544 | 0.9219 |
| $B_4$: City * BeforeAfter | -0.073510 | 0.424073 | 0.8624 |
| $B_5$: Quarter * BeforeAfter | -0.058969 | 0.031322 | 0.0597 |
| $B_6$: City * Quarter | -0.026553 | 0.077858 | 0.7331 |
| $B_7$: City * Quarter * BeforeAfter | 0.085456 | 0.083539 | 0.3063 |