**Police Stops in the City of Berkeley: Are People of Color Still Being Stopped**

**More Despite Changes in Police Policy?**

by KC Harris

April 26, 2022

**Introduction**

Berkeley, California is known as one of the cities with the most progressive police departments in the US. However, recent reports have said that while the city is better than most, there are still racial disparities found in recent crime data from 2012-2016. This report analyzes city of Berkeley Police stops data since 2015 as well as consider new RIPA (Racial Identity and Profiling Act) compliant features from stops data since 2020 to verify some existing claims, and test new ones specifically regarding disparities in race. Key questions for this analysis include:

* How much does prior perception of race play a role in the result of a stop, particularly if the citizen being stopped is a person of color?
* Can environmental factors like income, residential demographics, and/or amount of previous police activity affect and/or reduce how many stops occur in an area? Do these vary by neighborhood, or are they largely consistent throughout the city?
* Can RIPA-Compliant data reveal more nuanced relationships between features of a stop? Can and/or will more detailed data echo previous discoveries about race and law enforcement?

**Background**

In 2018, the Center for Police Equity released a report on the Berkeley Police Department saying that there were racial disparities in arrest rates between white and nonwhite people of Berkeley. [[1]](#footnote-1) Their study found that people of color were 4.5x-6.5x more likely to be stopped than white citizens, 4.5x-20x more likely to be searched per capita, and 2x as likely to be arrested overall.[[2]](#footnote-2) It’s worth noting that additional analysis found that black and hispanic searches yielded less stops, however this didn’t affect overall arrest rates. Use of force disparities were also found, but due to the complexity of classifying force in stop data we will not be considering this in our report. The report also requested that the city of Berkeley collect more race specific data in their reports, particularly surrounding the perceived race of the subject and if their race had been perceived prior to the stop. Their data, while comprehensive and collected from many arrest, crime, stop, and call sources, simply did not have the features at the time to make more conclusive statements on disparities in police behavior surrounding race.

**Methods**

***Study Design***

We use multiple logistic regression to observe how likelihood (specifically *odds ratios*) of arrest (and other outcomes) vary depending on factors like race, area median income, age, distance from the university, etc. This is done with the understanding that while stop data is not directly equivalent to crime or arrest data, analyses here can at least reveal disparities in stop rates for different populations, and what goes into those disparities. Study variables are based on the previously mentioned project done by the Center for Police Equity and the first portion of their analysis focused on stop rates. While our data is limited and we can’t directly infer differing stop rates just with data recorded after the stop, we do take the same variables and use them to analyze stop outcomes at scale.

***Data***

The data used in this project was collected by the city of Berkeley Police Department from 2015-2022, and downloaded in March of 2022 from its publicly available open access data portal.[[3]](#footnote-3) The data comes in two formats: RIPA[[4]](#footnote-4) and Non-RIPA Compliant data. All data since October of 2020 is RIPA-Compliant and as a result provides increased insights. Non-RIPA Compliant data has been kept to attempt to comprehensively represent the city, but some variables have been changed to match with new RIPA terminology, and certain models may vary in observation size due to lacking shared features between the two datasets. This is unfortunate, but some key assumptions can still be tested regardless of the differences in data.

Crime and arrest data is also not directly available through the city, and will be left out of this analysis. While it is still beneficial to look at stops to analyze police activity, it’s important to clarify the difference between stops and arrests, and admit that while comprehensive, the presently available stops data do not paint the whole picture. Inferences made in previous reports surrounding the likelihood of being stopped, and the full volume of occurring crime cannot be equivalently made with only stops data.

***Measurements***

**Outcomes**

*Arrest* is the main dependent variable examined in this study. In the logistic regression model, it is a simple arrested/not arrested classification variable. Alternative dependent variables include *noactions* and *warning*, although these aren’t given as much focus.

*Longstop* and/or *duration of stop* are also in focus as a dependent variable, but only apply for the data from 2020 and later. RIPA data mandates that the length of the stop be recorded. In the case of this study, and stops over 270 minutes, or 4.5 hourss were excluded. In the context of this study, this classification variable essentially asks “based on certain conditions, what are the odds a person has a longer than average stop?”

**Explanatory Variables**

*Perceived Race or Ethnicity* represents the race of the person stopped. This variable is a combination of “race” from the non-ripa dataset and “perceived race” from the ripa dataset.

*Race Perceived Prior to Stop* is a binary variable representing (1) the subject’s race was perceived by the officer before the stop and (0) the subject’s race was not perceived before the stop. This data, while insightful, is only available for the models based on the RIPA-compliant data. For the purpose of this analysis, single and double race categories have been simplified into “White”, “Black/African American”, “Hispanic/Latino”, “Asian”, and “Other”. Any perceived racial categories with more than 3 mentioned races were classified as “Other”. The 2017 report on Berkeley PD specifically cited concerns with enforcement towards black and hispanic populations, so these categories are focused on most here.

*Whitepop, aapop, na\_aipop, hawaiian, and mixed2* all represent specific estimated counts of racial demographics per Berkeley census tract in 2020 based on data from the US Census.[[5]](#footnote-5) These are useful in the multilinear regression as well as in the mapping portion of the analysis.

*Totalpop, Bipocpop,* and *Pocpop* are the estimated total racial population numbers per census tract from the same census data and city information from geolocation maps.[[6]](#footnote-6) Nonwhite includes all other racial categories outside of white, and POC includes Black, Hispanic, Hawaiian, and Mixed races.

*Bipoccomp* and *Focus Group Comp* are a simple calculation of the estimated proportion of BIPOC residents and Focus Group Residents (The Black and Hispanic populations that were the focus of the CPE report) per census tract. This is a broad representation but is acceptable for creating simple variables regarding “whiter” neighborhoods where stops would supposedly occur less.

*Distancefromcal, Far,* and *Close* represent another variable mentioned in the earlier study, distance from the university. Although we later find that this is at best curvilinear, it’s still worth including to see if it is still statistically significant.

**Results**

We provide the descriptive statistics in Table 1. We did not find multicollinearity issues with any variables. Notably, certain variables had less observations as they could only be generated from information in the RIPA-Compliant dataset. Because of this multiple models were run both for the larger and smaller datasets based on what variable information was available. This will be discussed more in later sections of the paper.

Table 1. Descriptive statistics (RIPA and Non-RIPA)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| DESC STATS | Obs | Mean | Std. Dev. | Min | Max |
| Subject Arrested | 61,702 | 0.018 | 0.133 | 0 | 1 |
| Perceived Age of Subject | 61,702 | 37.342 | 12.995 | 5 | 99 |
| Stop Distance from Cal | 61,702 | 1.435 | 0.704 | 0.25 | 2.5 |
| Census Tract Total # of Stops | 61,702 | 546.984 | 469.688 | 13.00 | 1454.000 |
| Census Tract Median Income | 61,702 | 87675.090 | 35424.510 | 20579.00 | 206199.000 |
| Census Tract Annual Average # of Stops | 61,702 | 68.377 | 58.837 | 2.00 | 182.000 |
| Census Tract Racial Composition (NW/W) | 61,702 | 0.388 | 0.184 | 0.00 | 0.768 |
| Perceived Race of Subject | 61,702 | 2.765 | 1.007 | 1.00 | 5.000 |
| Reason for Stop | 61,702 | 1.549 | 0.710 | 1.00 | 6.000 |
| BIPOC Person | 61,702 | 0.647 | 0.478 | 0.00 | 1.000 |
| Focus Group Person (Black/Hispanic) | 61,702 | 0.583 | 0.493 | 0.00 | 1.000 |
| Far From University | 61,702 | 0.236 | 0.424 | 0.00 | 1.000 |
| Close to University | 61,702 | 0.578 | 0.494 | 0.00 | 1.000 |
| Reasonable Suspicion Based Stop | 61,702 | 0.435 | 0.496 | 0.00 | 1.000 |
| Traffic Stop | 61,702 | 0.944 | 0.229 | 0.00 | 1.000 |
| Subject Let Off with Warning | 61,702 | 0.354 | 0.478 | 0.00 | 1.000 |
| Result of Stop | 61,702 | 6.798 | 4.263 | 1 | 11 |
| Duration of Stop | 8,088 | 16.613 | 17.631 | 1.00 | 270.000 |
| Information Based Stop | 8,088 | 0.227 | 0.419 | 0.00 | 1.000 |
| No Actions Resulting from Stop | 8,088 | 0.518 | 0.500 | 0.00 | 1.000 |
| Race Perceived Prior to Stop | 8,088 | 0.425 | 0.494 | 0.00 | 1.000 |
| Stop Duration Longer Than Average | 8,088 | 0.297 | 0.457 | 0.00 | 1.000 |
| Perceived Gender of Subject | 8,086 | 1.245 | 0.436 | 0.00 | 2.000 |

The results of the from the larger logistic regression models are displayed in Tables 2 & 3. Results show that that the subject race variable was significant and increased likelihood of arrest, and that in models where the subject was white this lowered the likelihood of arrest.

Table 2. Logistic Regression Model (RIPA & Non-RIPA Compliant) (Subject is Black)

|  |  |  |  |
| --- | --- | --- | --- |
| VARIABLES (Large) | Model 1 | Model 2 | Model 3 |
| Subject is Black | 1.574\*\*\* | 1.371\*\*\* | 1.359\*\*\* |
|  | (0.096) | (0.086) | (0.087) |
| Age of Subject | 1.009\*\*\* | 1.006\* | 1.006\* |
|  | (0.003) | (0.002) | (0.002) |
| Perceived Gender of Subject | 0.471\*\*\* | 0.676\*\*\* | 0.673\*\*\* |
|  | (0.039) | (0.057) | (0.057) |
| Traffic Stop |  | 0.076\*\*\* | 0.069\*\*\* |
|  |  | (0.005) | (0.005) |
| Tract Distance from University |  |  | 1.154 |
|  |  |  | (0.095) |
| Tract Total Population |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Tract Median Income |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Tract Average Annual Stops |  |  | 0.999\*\*\* |
|  |  |  | (0.000) |
| Tract BIPOC Composition |  |  | 1.644 |
|  |  |  | (0.465) |
| Constant | 0.028\*\*\* | 0.084\*\*\* | 0.028\*\*\* |
|  | (0.004) | (0.012) | (0.011) |
|  |  |  |  |
| Observations | 61,702 | 61,702 | 60,897 |
| seEform in parentheses |  |  |  |
| \*\*\* p<0.001, \*\* p<0.01, \* p<0.05 |  |  |  |

Table 3. Logistic Regression Model (RIPA & Non-RIPA Compliant) (Subject is White)

|  |  |  |  |
| --- | --- | --- | --- |
| VARIABLES (Small) | Model 1 | Model 2 | Model 3 |
| Arrest | . | . | . |
|  | (.) | (.) | (.) |
| Subject is White | 0.966 | 0.820\*\* | 0.820\*\* |
|  | (0.062) | (0.054) | (0.055) |
| Age of Subject | 1.009\*\*\* | 1.008\*\* | 1.008\*\* |
|  | (0.003) | (0.002) | (0.002) |
| Perceived Gender of Subject | 0.465\*\*\* | 0.678\*\*\* | 0.674\*\*\* |
|  | (0.038) | (0.057) | (0.057) |
| Traffic Stop |  | 0.074\*\*\* | 0.067\*\*\* |
|  |  | (0.005) | (0.005) |
| Tract Distance from University |  |  | 1.168 |
|  |  |  | (0.096) |
| Tract Total Population |  |  | 1.000\*\* |
|  |  |  | (0.000) |
| Tract Median Income |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Tract Average Annual Stops |  |  | 0.999\*\*\* |
|  |  |  | (0.000) |
| Tract BIPOC Composition |  |  | 1.707 |
|  |  |  | (0.481) |
| Constant | 0.034\*\*\* | 0.095\*\*\* | 0.033\*\*\* |
|  | (0.005) | (0.013) | (0.012) |
|  |  |  |  |
| Observations | 61,702 | 61,702 | 60,897 |
| seEform in parentheses |  |  |  |
| \*\*\* p<0.001, \*\* p<0.01, \* p<0.05 | |  |  |

The results from the smaller logistic regression models are displayed in Tables 4 and 5. These more nuanced models, while smaller, echo the earlier significant race variables and also add that when available, the race perceived prior to stop variable was very significant.

Table 4. Logistic Regression Model (RIPA-Compliant Only)

|  |  |  |  |
| --- | --- | --- | --- |
| VARIABLES (Small) | Model 1 | Model 2 | Model 3 |
| Subject is Black | 1.490\*\*\* | 1.430\*\*\* | 1.429\*\*\* |
|  | (0.098) | (0.097) | (0.098) |
| Age of Subject | 1.001 | 0.997 | 0.996 |
|  | (0.003) | (0.003) | (0.003) |
| Perceived Gender of Subject | 0.544\*\*\* | 0.572\*\*\* | 0.571\*\*\* |
|  | (0.046) | (0.050) | (0.050) |
| Traffic Stop |  | 0.306\*\*\* | 0.301\*\*\* |
|  |  | (0.033) | (0.033) |
| Race Was Perceived Prior |  | 2.769\*\*\* | 2.730\*\*\* |
|  |  | (0.193) | (0.195) |
| Tract Distance from University |  |  | 1.001 |
|  |  |  | (0.086) |
| Tract Total Population |  |  | 1.000 |
|  |  |  | (0.000) |
| Tract Median Income |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Tract Average Annual Stops |  |  | 1.000\* |
|  |  |  | (0.000) |
| Tract BIPOC Composition |  |  | 2.782\*\* |
|  |  |  | (0.885) |
| Constant | 0.271\*\*\* | 0.538\*\*\* | 0.173\*\*\* |
|  | (0.039) | (0.097) | (0.076) |
|  |  |  |  |
| Observations | 8,086 | 8,086 | 8,008 |
| seEform in parentheses |  |  |  |
| \*\*\* p<0.001, \*\* p<0.01, \* p<0.05 |  |  |  |

Table 5. Logistic Regression Model (RIPA-Compliant Only)

|  |  |  |  |
| --- | --- | --- | --- |
| VARIABLES (Small) | Model 1 | Model 2 | Model 3 |
| Subject is White | 0.977 | 0.840\* | 0.836\* |
|  | (0.067) | (0.060) | (0.060) |
| Age of Subject | 1.002 | 0.998 | 0.998 |
|  | (0.003) | (0.003) | (0.003) |
| Perceived Gender of Subject | 0.540\*\*\* | 0.572\*\*\* | 0.571\*\*\* |
|  | (0.046) | (0.050) | (0.050) |
| Traffic Stop |  | 0.309\*\*\* | 0.302\*\*\* |
|  |  | (0.033) | (0.033) |
| Race Was Perceived Prior |  | 2.877\*\*\* | 2.834\*\*\* |
|  |  | (0.201) | (0.203) |
| Tract Distance from University |  |  | 1.002 |
|  |  |  | (0.086) |
| Tract Total Population |  |  | 1.000 |
|  |  |  | (0.000) |
| Tract Median Income |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Tract Average Annual Stops |  |  | 1.000\*\* |
|  |  |  | (0.000) |
| Tract BIPOC Composition |  |  | 2.849\*\*\* |
|  |  |  | (0.906) |
| Constant | 0.314\*\*\* | 0.611\*\* | 0.210\*\*\* |
|  | (0.044) | (0.109) | (0.091) |
|  |  |  |  |
| Observations | 8,086 | 8,086 | 8,008 |
| seEform in parentheses |  |  |  |
| \*\*\* p<0.001, \*\* p<0.01, \* p<0.05 | |  |  |

**Discussion**

The models above indicate continued disparities surrounding race and likelihood of arrest in the city of Berkeley. There are both direct differences in likelihoods of arrest for Black and White subjects in our models, but also other indicators that race plays a role exist in the significance of the race perception and BIPOC composition variables. These suggest not only that there are differing odds of arrest for people of color, but more interestingly that there are general perceptions around race held by Berkeley Police that affect how stops are conducted.

While the larger models lacked more specific racial data, we were still able to find broad differences in Black and White arrest odds. In the model where Subjects were white, initially the White variable did not have a significant relationship, and by the third stage had a slightly significant relationship with arrest but was still ultimately less likely to be arrested. In the model where the Subjects were black, there was always a very significant relationship (p < .001), and by the third stage the higher likelihood of arrest had very little change. Black subjects in these large models were not 2x as likely to be arrested, but approximately 1.6x as likely instead. It’s also worth noting that in the larger models, the tract BIPOC composition variable was not significant and therefore did not play any role in affecting the likelihood of arrest. This is in contrast with the smaller models, where it was a significant or very significant variable. This could suggest that like the other tract features, bipoc composition does not matter, or it could necessitate further examination of the interaction of the race perception and tract bipoc composition variables.

The smaller models also had differences in white and black arrest odds, and added that tract bipoc composition and prior perception of race were also significant. Being white had a slightly significant relationship with arrest and made subjects less likely to be arrested, while being black consistently had a very significant relationship with arrest and made subjects more likely to be arrested. Black subjects were approximately 1.7x more likely to be arrested in the same scenario. Prior perception of race was consistently very significant in both models, and had a major increase in the likelihood of arrest. And in the smaller models where race had been perceived, tract bipoc composition was consistently very significant and also had a major increase in likelihood of arrest.

These confirm the existing likelihoods that were produced by the previous report, and strengthen the broader disparities surrounding race that were not available for analysis with older data. While there are less observations in the RIPA dataset, the high significance levels of the perception of race and bipoc composition variables suggest that if the officer perceives the race of the subject prior to stop, the race of the subject and the racial composition of the surrounding area play a large role in the larger result of the stop, particularly if the subject is not white or if the percent of BIPOC residents in that area is high. This suggests that officers behave differently based on race, and that there are disparities in police practices in the city of Berkeley.

In both models, the majority of the local census tract features proved to either be insignificant, or to not increase or decrease likelihood of arrest. Besides racial composition of the area following perception of subject race, other factors like distance from the university, median income, or even historical stop rates did not change likelihood of arrest.

It is important to note however that these are still fairly new – the insights provided by RIPA-compliant data are based on data that has only been collected this way for less than two years. It is possible that certain variables will be less significant, or may change in scale when tested again with more data. Additionally, the < two year period in which this data was collected was also during the COVID-19 pandemic. While it is difficult to say just how different the RIPA data would be if it was collected in a non-quarantine environment, it is important to recognize that the data cannot be made to be the same. It is possible that further research into what residents of berkeley were out more during the pandemic, and what the racial composition of those residents was, could reveal further social disparities in what demographics may have been over or under policed. We also suggest that a similar analysis to this one be performed approximately in 2025 when a larger amount of data has been collected.

1. <https://www.berkeleyside.org/wp-content/uploads/2018/05/Berkeley-Report-May-2018.pdf> [↑](#footnote-ref-1)
2. ^ pages 7-8 [↑](#footnote-ref-2)
3. [Berkeley PD - Stop Data (Jan 26, 2015 to Sep 30, 2020)](https://data.cityofberkeley.info/Public-Safety/Berkeley-PD-Stop-Data-Jan-26-2015-to-Sep-30-2020-/4tbf-3yt8)

   ^ [Berkeley PD - Stop Data (October 1, 2020 - Present)](https://data.cityofberkeley.info/Public-Safety/Berkeley-PD-Stop-Data-October-1-2020-Present-/ysvs-bcge) [↑](#footnote-ref-3)
4. [Racial and Identity Profiling Act (RIPA)](https://post.ca.gov/Racial-and-Identity-Profiling-Act) [↑](#footnote-ref-4)
5. [American Community Survey, B02001 RACE](https://data.census.gov/cedsci/table?t=Race%20and%20Ethnicity&g=1400000US060014,06001421100,06001421200,06001421300,06001421400,06001421500,06001421600,06001421700,06001421800,06001421900,06001422000,06001422100,06001422200,06001422300,06001422400,06001422500,06001422600,06001422700,06001422800,06001422900,06001422901,06001422902,06001423000,06001423100,06001423200,06001423300,06001423400,06001423500,06001423601,06001423602,06001423700,06001423800,06001423901,06001423902,06001424001&tid=ACSDT5Y2020.B02001) [↑](#footnote-ref-5)
6. [(Census Tract Polygons 2010) Census tract polygons built from US Census Bureau 2010 decennial data for the City’s redistricting process](https://data.cityofberkeley.info/Demographics/Census-Tract-Polygons-2010/peq3-2arw) [↑](#footnote-ref-6)