**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 will attempt to analyze publicly available stop data collected since, as well as attempt to consider newly RIPA (Racial Identity and Profiling Act) compliant data that’s been collected since late 2020, to partially analyze if these racial disparities still exist in more recent interactions with the public. 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. This analysis aims to use available stop data to see if some of these trends are still in effect: specifically arrest rates by race following a stop, search rates by race following a stop, and with the help of the newer information provided by the RIPA-compliant data, if arrest and search odds following perception of race for people of color. Use of force analysis also included data about the distance from the university, crime, income, and racial composition data at the census tract level. We’ve chosen to include these in this analysis as well to see if they are possibly significant outside of use of force analysis.

**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, Nonwhitepop,* 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.

*Nonwhitecomp* and *Poccomp* are a simple calculation of the estimated proportion of nonwhite and poc residents (all categories except white/except white or asian) of total residents per census tract. This is a broad representation but is acceptable for creating simple variables regarding “whiter” neighborhoods where stops would supposedly occur less. *Nonwhite* and *poc* variables are also generated later in analysis during logistic regression.

*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**

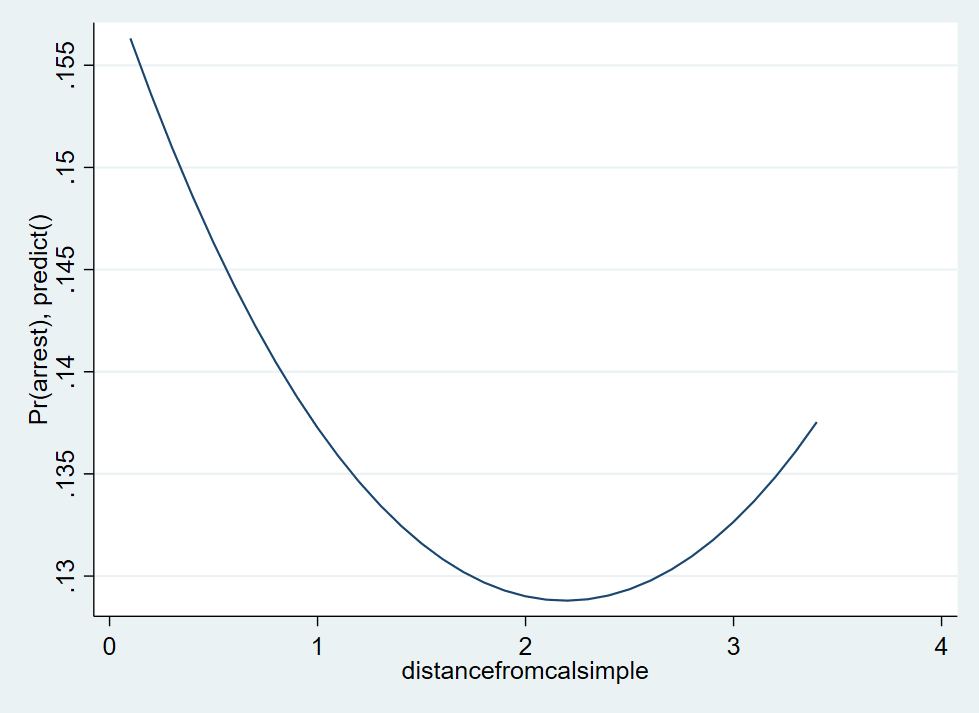
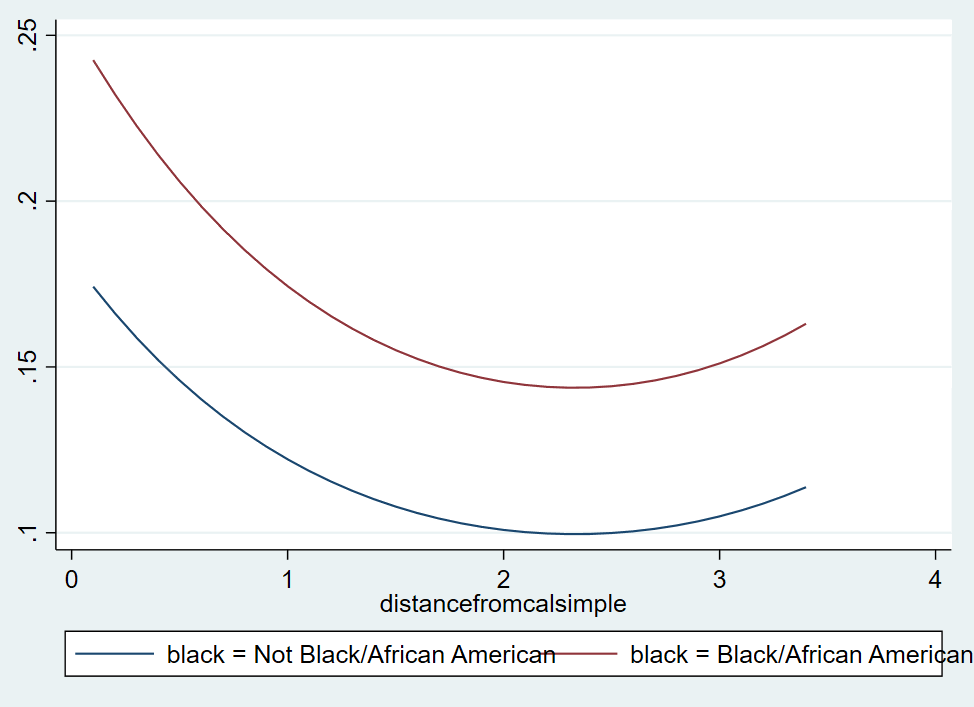
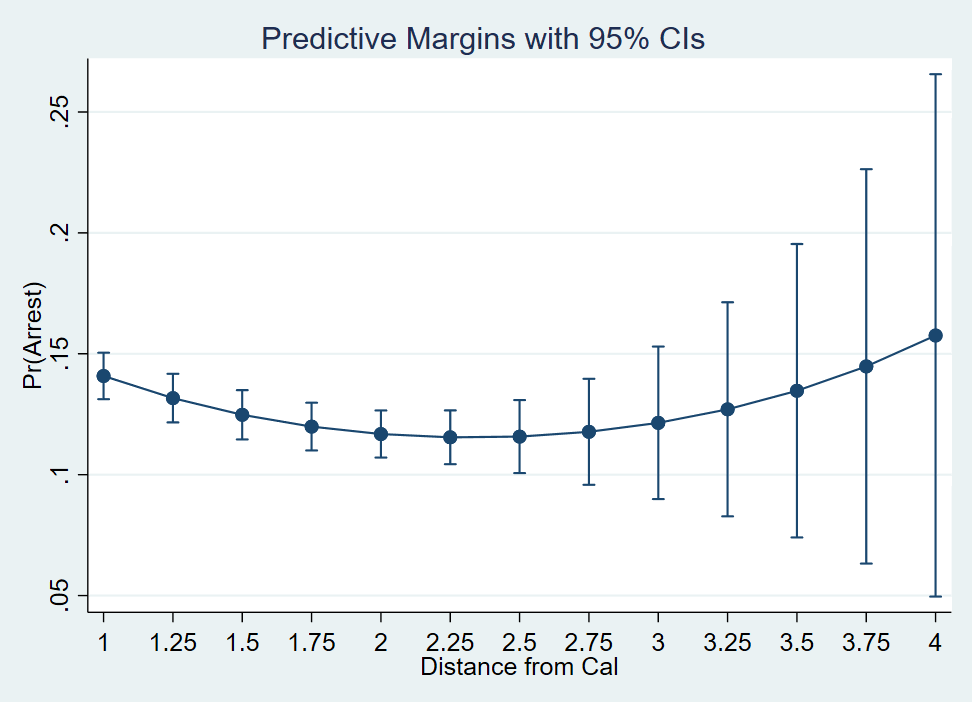
**More results are on the way. I’m building separate models from what data I can put together from the small dataset as well as the large, combined datasets. Sorry will finish over weekend and bring in next week. Very excited about this project and happy with my progress so far.**

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 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 | 87675.09 | 35424.51 | 20579.00 | 206199.00 |
| 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 |
| Nonwhite Person | 61,702 | 0.647 | 0.478 | 0.00 | 1.000 |
| Person of Color | 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 |

|  |  |  |  |
| --- | --- | --- | --- |
| Models (Black) | Model 1 | Model 2 | Model 3 |
|  |  |  |  |
| Subject Arrested | . | . | . |
|  | (.) | (.) | (.) |
| Subject is Black | 1.490\*\*\* | 1.411\*\*\* | 1.408\*\*\* |
|  | (0.098) | (0.096) | (0.097) |
| Subject Age | 1.001 | 0.996 | 0.996 |
|  | (0.003) | (0.003) | (0.003) |
| Subject Perceived Gender | 0.544\*\*\* | 0.579\*\*\* | 0.577\*\*\* |
|  | (0.046) | (0.050) | (0.051) |
| Stop is Traffic Stop |  | 0.319\*\*\* | 0.313\*\*\* |
|  |  | (0.034) | (0.034) |
| Stop Not Based on Previous Info |  | 1.614\*\*\* | 1.605\*\*\* |
|  |  | (0.118) | (0.118) |
| Race Perceived Prior to Stop |  | 2.580\*\*\* | 2.565\*\*\* |
|  |  | (0.182) | (0.186) |
| Distance of Local Census Tract to Cal |  |  | 1.034 |
|  |  |  | (0.089) |
| Local Census Tract Total Population |  |  | 1.000 |
|  |  |  | (0.000) |
| Local Census Tract Median Income |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Local Census Tract Annual Average Stops |  |  | 1.000\* |
|  |  |  | (0.000) |
| Local Census Tract Nonwhite Composition |  |  | 2.818\*\* |
|  |  |  | (0.899) |
| Constant | 0.271\*\*\* | 0.479\*\*\* | 0.150\*\*\* |
|  | (0.039) | (0.087) | (0.066) |
|  |  |  |  |
| Observations | 8,086 | 8,086 | 8,008 |
| seEform in parentheses |  |  |  |
| \*\*\* p<0.001, \*\* p<0.01, \* p<0.05 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Models (White Comparison) | Model 1 | Model 2 | Model 3 |
|  |  |  |  |
| Subject Arrested | . | . | . |
|  | (.) | (.) | (.) |
| Subject is White | 0.977 | 0.851\* | 0.849\* |
|  | (0.067) | (0.061) | (0.061) |
| Subject Age | 1.002 | 0.997 | 0.997 |
|  | (0.003) | (0.003) | (0.003) |
| Subject Perceived Gender | 0.540\*\*\* | 0.580\*\*\* | 0.578\*\*\* |
|  | (0.046) | (0.050) | (0.051) |
| Stop is Traffic Stop |  | 0.322\*\*\* | 0.315\*\*\* |
|  |  | (0.035) | (0.035) |
| Stop Not Based on Previous Info |  | 1.624\*\*\* | 1.616\*\*\* |
|  |  | (0.118) | (0.119) |
| Race Perceived Prior to Stop |  | 2.675\*\*\* | 2.656\*\*\* |
|  |  | (0.189) | (0.192) |
| Distance of Local Census Tract to Cal |  |  | 1.036 |
|  |  |  | (0.089) |
| Local Census Tract Total Population |  |  | 1.000 |
|  |  |  | (0.000) |
| Local Census Tract Median Income |  |  | 1.000\*\*\* |
|  |  |  | (0.000) |
| Local Census Tract Annual Average Stops |  |  | 1.000\*\* |
|  |  |  | (0.000) |
| Local Census Tract Nonwhite Composition |  |  | 2.889\*\*\* |
|  |  |  | (0.921) |
| Constant | 0.314\*\*\* | 0.538\*\*\* | 0.179\*\*\* |
|  | (0.044) | (0.097) | (0.078) |
|  |  |  |  |
| Observations | 8,086 | 8,086 | 8,008 |
| seEform in parentheses |  |  |  |
| \*\*\* p<0.001, \*\* p<0.01, \* p<0.05 |  |  |  |

**Discussion**

While larger analysis is still limited, one advantage here is the detail provided by RIPA. This limitation was noted in the CPE’s report, and highlights one key element they couldn’t examine: perception of race prior to stop.

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)