Racial Bias in Policing: An In-depth Analysis of Stopping Practices by the Long Beach Police Department

Methodology

September 2023

We relied on data collected by law enforcement agencies under the Racial and Identity Profiling Act (RIPA) of 2015 to analyze trends in the Long Beach Police Department's (LBPD) patrol activities. We combined this data with population data from the U.S. Census Bureau and high-injury network data from Safe Streets Long Beach to identify racial disparities, harms, and inefficiencies in LBPD's practices. Our data analysis focused on the year 2019—the most recent year of data available at the time of analysis that was not impacted by the onset of the COVID-19 pandemic in 2020. While more recent data were available at the time of publication, our initial analysis of this data suggested that the distribution of LBPD's stops, other than the total number of stops, does not differ remarkably from what was observed in 2019.[1]

As with all data, the findings and trends seen in this analysis are dependent on the quality and limitations of the data used. We strongly encourage readers to consider the limitations of RIPA data when interpreting findings. RIPA data are based on officers' reports, meaning the information attached to each stop is solely based on officer disclosure and perceptions of what occurred during the stop. For instance, officers report what they perceive as the race(s) of people stopped, rather than the self-reported race(s) of the people in each incident. Other reports suggest this leads to underreporting, misidentification, or even intentional obstruction of information by officers.[2] To help provide a check on police stop data as reported by officers, it is imperative to consult with the community affected to make sense of trends and check the data for inaccuracies based on everyday community experiences.

Data Sources

Police Stop Data

City of Long Beach, Long Beach Police Department, 2019, Police Stop Data (RIPA). Retrieved from https://www.longbeach.gov/police/about-the-lbpd/lbpd-stop-data/.

Population Estimates by Race

---

[1] In 2021, according to the City of Long Beach's open data portal, LBPD stopped 11,978 people compared to 40,523 stops in 2019. However, in 2021, the majority of stops were still for officer-initiated stops and traffic violations. About 89% of people were still for officer-initiated stops and 89% of these people were stopped for traffic violations.

[2] Office of Inspector General, *The Sheriff's Department's Underreporting of Civilian Stop Data to the California Attorney General*, County of Los Angeles (June 10, 2022). Retrieved from https://assets-us-01.kc-usercontent.com/0234f496-d2b7-00b6-17a4-b43e949b70a2/ee467145-85c7-450c-a73993e1f1d79f78/The%20Sheriff%E2%80%99s%20Department%E2%80%99s%20Underreporting%20of%20Civilian%20Stop%20Data%20 to%20the%20California%20Attorney%20General.pdf.

Barba, M., *Watchdogs want answers on how San Francisco cops misreported race data*, The San Francisco Standard (September 13, 2023). Retrieved from https://sfstandard.com/2023/09/13/san-francisco-cops-misreported-data-meant-to-curb-racial-profiling-now-watchdogs-want-answers/

U.S. Census Bureau, 2016-2020, American Community Survey, 5-Year Estimates. Tables DP05, B04006, B02018. Retrieved from https://data.census.gov/cedsci/.

U.S. Census Bureau, 2020, TIGER/Line Shapefiles, Census Tracts. Retrieved from https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.2020.html#list-tab-790442341.

City of Long Beach, 2022, City Council Districts. Retrieved from https://longbeach.opendatasoft.com/explore/dataset/colb-council-districts/information/

High-Injury Corridors and Intersections
City of Long Beach, Safe Streets Long Beach: A Vision Zero Action Plan (July 2020), 2013-2017, Top 20 High-Injury Pedestrian/Bicycle and Motor Vehicle/Motorcycle Corridors and Intersections. Retrieved from https://www.longbeach.gov/globalassets/go-active-lb/media-library/documents/programs/safe-streets-lb-action-plan---final.

Data Limitations

Officer Reports of Race and Ethnicity

Race and identity fields included in RIPA data are imperfect in their report and subject to officer bias. The race of people stopped in each incident is based on officer perceptions, rather than self-identification by the people involved in each incident. RIPA regulations require officers to report their perception of the race, gender, age, and other identity information without input from the person they stop.[3] Officers' perceptions are the proper lens to use for purposes of understanding racial profiling because this is the information the officer knew (or assumed) before they stopped the person. However, these perceptions are also subject to the officers' own bias or understanding of people's identities. Stop rates for certain groups may be over- or under-reported due to misidentification by officers. In other cases, officers may intentionally obscure the racial identity of the people they stopped.

In Long Beach's 2019 data, we found that officers reported people as being five or more races 2.5 percent of the time, or in 1,015 out of 40,523 people stopped in 2019. These numbers far exceed the percentage of multiracial people in Long Beach who are of five or more races based on census data.[4] Recent reports from other jurisdictions suggest that these instances could be intentional obstruction by officers to mask bias in their stops.[5] Other reports suggest that Latinx people are overrepresented in stops that are underreported by agencies.[6]

The disparities we see in stop rates by race are only as good as the quality of the data reported. That is why it is vital to check trends with community and partners on the ground to detect possible discrepancies or inaccuracies. We believe the racial disparities observed in this analysis are a good

Commented [HK1]: Can we elaborate? I'm not sure I understand the underreported part (might just be me though)

---

[3] 11 Cal. Code. Regs. § 999.226. Available at https://oag.ca.gov/sites/all/files/agweb/pdfs/ripa/stop-data-reg-final-text-110717.pdf
[4] Based on American Community Survey 5-year estimates from 2016-2020, less than 0.1% of people in Long Beach who identified as multiracial identified with five or more races. However, in the RIPA data, 1.3% of people officers perceived as multiracial were coded as five or more races by officers.
[5] Barba, M., (September 13, 2023).
[6] Office of Inspector General, (June 10, 2022).

indication of racial bias by LBPD officers, but they may also undercount disparities for certain groups that are either underreported, misidentified, or intentionally withheld by officers.

Stop Duration Times

Stop duration times are manually inputted by police officers. However, data input errors record some stops as being extremely long, over 24 hours, or extremely short, 1 minute, and most likely does not reflect the actual amount of time those stops took. To mediate extreme stop duration times impacting our analysis, we identified certain stop duration times as outliers and capped them to an upper or lower threshold value based on a model of stop characteristics. About 4.8 percent of stops were identified as outliers and capped. This method is imperfect and does not calculate with complete accuracy the total stop duration time. We assume that our total stop duration times are conservative estimates and that it is more likely officers spent more time on stops rather than less on each dimension of stops examined.

Stop Locations

RIPA data regulations provide officers flexibility in how they report the physical location of the stop. Officers may report the block number and street number, closest intersection, highway and closest highway exit, or a road maker, landmark or other description (in that preferred order).[7] Within each type of location, the quality of a given address varies and this impacts the accuracy of our coordinate results. In some cases, street addresses that are only in Long Beach were reported as occurring in a different city. In other cases, a stop that took place at an intersection would include two (or more) streets that did not intersect. In many cases, there were typos or ambiguous street names (e.g., recording a stop at "golden" but not specifying if it was Golden Ave or Golden Shore). This leads to issues with using the data in a few ways. It increases the amount of time and capacity required to thoroughly clean the data. It also reduces the number of observations that we can include in our analysis. Fewer observations can lead to underestimating police stops, and those impacted by them, when analyzing council districts. Lastly, ambiguous data means we cannot guarantee that every coordinate in our analysis is accurate. With these limitations, we were able to successfully geocode, or map, 94.4 percent of stop locations. To estimate accuracy of these points, we took samples of our results and found that in most cases, coordinates were within 0.2 miles of what we assume is the actual stop address.

High-injury Corridor Networks

One of the limitations of this analysis is that the data captured for the top 20 high-injury corridors and intersections within the City of Long Beach was based on five years of available crash data from January 1, 2013 through December 31, 2017. We used the high-injury data from the "Safe Streets Long Beach: A Vision Zero Action Plan" report. This report was released in July 2020, though the corridors and intersections identified are based on data from 2013-2017. On the other hand, we relied on RIPA data from 2019. While the high-injury data does not match the same year as police stop data, the high-injury data is the latest data used by the City of Long Beach. Additionally, we do not expect trends to differ significantly. According to the "Safe Streets Long Beach: A Vision Zero Action Plan" report, "Many of the identified corridors and intersections are concentrated in Downtown, Central, and North Long Beach. These streets are generally principal or minor arterials with high rates of multimodal traffic, situated

---

[7] 11 Cal. Code. Regs. § 999.226. Available at https://oag.ca.gov/sites/all/files/agweb/pdfs/ripa/stop-data-reg-final-text-110717.pdf

adjacent to higher density land uses, and often have posted speed limits in excess of 30 mph".[8] In other words, we expect many of the identified corridors and intersections to remain the same in 2019 due to the street conditions, higher speed limits, and higher traffic. The data also provide an important reference point for showing how low-income communities of color impacted by the trauma of traffic injuries are dually impacted by exposure to police stops.

How We Use RIPA Race Categories

The racial categories available to officers according to RIPA regulations are: Asian, Black/African American, Hispanic/Latino/a, Middle Eastern or South Asian, Native American, Pacific Islander, and White. We adjusted these categories and labels as needed to be more reflective of the identities of each community and the language used by organizations representing these communities. We also added a category for Multiracial, or Two or More Races, to account for people officers perceived as being more than one race. We use the following categories in our analysis as compared to their original label in RIPA data. Not all categories are mutually exclusive.

- American Indian or Alaska Native (AIAN): People perceived as AIAN (or Native American) alone or in combination with another race or Latinx.
- Asian: People perceived only as Asian and non-Latinx.
- Black: People perceived only as Black (or African American) and non-Latinx.
- Latinx: People perceived as Latinx (Hispanic/Latino/a), whether alone or in combination with another race.
- Two or More Races/Multiracial: People perceived as more than one race and non-Latinx.
- Native Hawaiian or Pacific Islander (NHPI): People perceived as NHPI (or Pacific Islander) alone or in combination with another race or Latinx.
- South Asian, Southwest Asian, or North African (SSWANA): People perceived as SSWANA (or Middle Eastern or South Asian) alone or in combination with another race or Latinx.
- White: People perceived as only White and non-Latinx.

In the 2.5 percent of cases where officers reported people as being five or more races, we assume the officer made an error or intentionally obscured the race of the person they stopped, and we do not include these cases in our analyses by race. These stops are included in all other analyses.

How We Create Population Estimates by Race

To scale stop counts by race and measure racial disparities, we created population estimates for the City of Long Beach and each council district based on data from the U.S. Census Bureau's American Community Survey (ACS). Our population estimates by race correspond to the racial groups we use from RIPA but are based on individuals' self-report in the ACS. All groups other than Latinx, NHPI, AIAN, and SSWANA are mutually exclusive and do not include Latinx individuals.

- American Indian or Alaska Native (AIAN): AIAN alone or in combination with another race or Latinx.
- Asian: Asian alone, non-Latinx.
- Black: Black or African American alone, non-Latinx.

---

[8] City of Long Beach, *Safe Streets Long Beach: A Vision Zero Action Plan,* (July 2020). Retrieved from https://www.longbeach.gov/globalassets/go-active-lb/programs/safe-streets-lb-action-plan---07-07-20v2.

- Latinx: Latinx (Hispanic/Latino/a), alone or in combination with another race.
- Two or More Races/Multiracial: People of two or more races, other than Latinx
- Native Hawaiian or Pacific Islander (NHPI): NHPI alone or in combination with another race or Latinx.
- South Asian, Southwest Asian, or North African (SSWANA): Includes people who identified with Southwest Asian or North African ancestry and/or as South Asian origin under their race, alone or in combination with another race or Latinx.
- White: White alone, non-Latinx.

Calculating Proxy Estimates for SSWANA

RIPA data includes a category for Middle Eastern or South Asian, but this label is not commonly used in other datasets. We opt for the term South Asian, Southwest Asian, or North African (SSWANA), which includes people of Southwest Asian or North African (SWANA), or Middle Eastern or North African (MENA), descent and South Asian people. Southwest Asian or North African is an alternative term to MENA that decolonizes and moves away from the term Middle Eastern. Census data have no equivalent category to help scale stop counts by population for these communities. We have created our own proxy population estimates for these communities based on a combination of ancestry and race tables from the ACS. We estimate the number of people who identify as SSWANA by summing the number of people who reported any SWANA ancestry and/or South Asian racial origin. We define SWANA ancestry based on these categories: Afghan, Algerian, Arab, Assyrian, Bahraini, Berber, Chaldean, Egyptian, Emirati, Iranian, Iraqi, Israeli, Jordanian, Kurdish, Kuwaiti, Lebanese, Libyan, Middle Eastern, Moroccan, North African, Omani, Palestinian, Qatari, Saudi, Sudanese, Syriac, Syrian, Tunisian, Turkish, Yazidi, and Yemeni. We define South Asian origin based on these categories: Asian Indian, Bangladeshi, Bhutanese, Maldivian, Nepalese, Pakistani, Sikh, Sindhi, and Sri Lankan.

It is important to note that consensus on the identity groups included in the terms SWANA or MENA is still being built. These labels are attempts to distinguish the distinct experiences of these communities from the White experience in the United States, the group they have traditionally been folded into when it comes to data. These groups encompass a diversity of origins and intersections between national, geopolitical, religious, and ethnic identities. For years, advocacy organizations representing these communities have advocated for the inclusion of a MENA category in the U.S Census.[9] While the U.S. Census Bureau still delays in including MENA in data collection, our proxy estimates are the best available to help bring to light the discrimination experienced by these communities when it comes to racially biased policing.

Calculating Population Estimates by Council District

We used census tract level data to calculate population estimates by race for each council district. Using a method called aerial apportionment, we matched census tracts to Long Beach city council districts based on how much each census tract overlaps with each district. We intersected, or joined, census tracts to council district and calculated the percentage overlap of each census tract with each council district. We kept intersects, or census tract-council district matches, where at least five percent of the census tract area overlaps with the council district. We estimated the population of each council district

---

[9] The Leadership Conference on Civil and Human Rights, *Will you count? Middle Eastern and North African Americans (MENA) in the 2020 census,* (April 17, 2018). Retrieved from https://civilrights.org/resource/will-you-count-middle-eastern-and-north-african-americans-in-the-2020-census/

by multiplying the percentage overlap of each census tract by the census tract population for each race. We then summed these products by council district. This method assumes that populations are evenly distributed in a census tract, which may not always be true. However, this is a common method for aggregating smaller geographic estimates to larger geographies.

How We Geocode RIPA Stop Data and High-Injury Networks

To compare LBPD stop data against Long Beach's high-injury networks, we used an API available through the Google Maps Platform to geocode each stop address and retrieve latitude and longitude coordinates. These coordinates allow us to take street addresses or cross-streets and plot their locations on a map.

We found the original stop addresses had a number of data quality issues (e.g., typos, nonexistent addresses, addresses outside of Long Beach, etc.). We tried to minimize these issues with data cleaning methods and by manually geocoding common addresses that returned inaccurate or no coordinates. Three percent of stops were not specific enough to use (i.e., they generalized to the middle of Long Beach, or the middle of street in the case of partial address matches). Additionally, there were 75 stops that the API could not find. We also found that 1,395 stops occurred outside of Long Beach and removed those stops from our analysis. In all, 94.4 percent of all LBPD stops that occurred in 2019 were geocoded in the City of Long Beach and included in the analysis.

We used the same geocoding process to identify the coordinates of high-injury intersections and corridors for pedestrians/bicycles and motor vehicles/motorcycles. We mapped the top 20 highest injury intersections and corridors for both groups. This resulted in 60 coordinate points for each group (20 for intersections, 20 for each corridor start, and 20 for each corridor end). All high-injury intersections and corridors included in Safe Streets Long Beach's report were geocoded and included in our analysis.[10]

How We Create Our Estimates

Calculating How Officers Spend their Time

Identifying and capping outliers

Police officers are required to input when a stop begins and ends to capture the total amount of time a stop took. However, there are some stops in the data that show up as having taken over 24 hours to complete or stops that took 1 minute. These outliers in stop duration times are likely due to data input errors, such as officers forgetting to clock the ending of the stop until after the fact or forgetting to clock the stop at all. To mitigate these outliers affecting our analysis, we created a regression model and conducted an outlier analysis to identify, and cap, stops with extremely high or low stop times.

The first step in our outlier analysis was to create a robust model that accounts for the various factors that impact stop duration times. These factors, also known as independent variables, were used to create a regression model with stop duration time acting as the dependent variable (or the variable influenced by the independent variables).

Dependent Variable: Stop duration time

Independent Variables:

---

[10] We manually collected the coordinates for 6 motor vehicle corridor points and 4 pedestrian corridor points using Google Maps that were not accurately geocoded using the API.

Commented [HK2]: This comes up in later comments so maybe we should squeeze in a blurb after this header that briefly explains calls for service and officer-initiated stops? We use both terms later on but don't explain that "not calls for service" is equivalent to "officer-initiated stops". We could also include what the total number of officer-initiated stops were (and calls for service)

Commented [HK3R2]: Also how we handle stop analysis by race when there are multiple people of multiple perceived identities, etc. could go here. There are several instances later on where we say "our analysis controls for stops with multiple people" but it might be a little vague

- Race of persons stopped
- Number of persons stopped
- Number of actions taken
- Stop reasons
- Stop results, e.g., citation, warning, arrest
- Was the stop a call for service
- Did the stop involve use of force
- Was contraband or evidence found
- Was a person handcuffed
- Was a person detained
- Was a person or their property searched
- Was a person removed from a vehicle

After the final model was determined, we used it to estimate the amount of time each stop should have taken given the combination of independent variables observed in each stop, along with a 95 percent confidence interval. This amount of time that a stop should have taken according to the model is called the predicted value. The 95 percent confidence interval tells us the lowest and highest amount of time a stop with those characteristics would have taken 95 percent of the time.

Using the confidence intervals of the predicted values, a stop was determined to be an outlier if the officer-reported duration time was either below or above the confidence interval. If a stop had a reported duration time within the confidence interval, then that duration time remained unchanged. If the reported stop duration time was below the confidence interval, then it was replaced with the lower confidence interval value. Alternatively, if the reported stop duration time was above the confidence interval, then it was replaced with the upper confidence interval value. All subsequent stop duration analyses were done using a combination of officer-reported and replaced ("capped") stop duration times. The replaced ("capped") stop duration times applied only to outlier stops.

Analyzing time spent by stop reason
The rate of time spent on stops by stop reason was calculated by taking the total number of hours spent on each stop reason divided by the total number of hours spent on all stops that were not calls for service. For this analysis, all stops that were calls for service were excluded. Stops with two or more reasons were recategorized as 'Two or more reasons' for the stop reason. Stop reasons that occurred less than 1% of the time were recategorized as 'Other' to limit the total number of stop reason categories. This analysis controlled for stops where more than one person was stopped.

Analyzing time spent by stop result
This analysis breaks down time spent on stops made for traffic violations based on the result of the stop (e.g., citation, warning, no action, etc.). Traffic stops were filtered to only those that were not made in response to calls for service (i.e., officer-initiated stops). Traffic stops that had more than one result were recategorized as 'Two or more results.' The rate of hours spent on each result type was calculated by taking the total number of hours spent on a traffic stop with a specific result, divided by the total number of hours spent on all traffic stops that were not calls for service. Stop results that occurred in less than 1% of total stops were categorized as 'Other.' This analysis controlled for stops where more than one person was stopped.

Commented [HK4]: Should we include the total number of capped/outlier stops?

Commented [HK5]: Should we say officer-initiated stops? Otherwise maybe we should add a line explaining calls for service since we haven't covered that yet

Commented [HK6]: should we add in the total number of categories? should we add the total number of officer-initiated stops included in this calculation?

Commented [HK7]: It might be too complicated but if there is an easy example to include that might help clarify what we mean by this

Commented [HK8]: Same as above - if there's a simple example it might make sense to add for clarity. Otherwise folks might not understand how this shapes our analysis

Commented [HK9R8]: Also - since some stops can get excluded (when filtering for traffic violations and to control of multiple persons) should we include the total number of stops used in this calculation?

Calculating Racial Disparities in Stop Rates

Calculating stop rates by race

To analyze officer-initiated stop rates by race, we first filtered the data to include only officer-initiated stops. We then summarized the total number of officer-initiated stops by race. Here, observations by race may overlap. For example, a person can be perceived by officers as multiple races, like Latinx and SSWANA. In this case, they are counted twice in the analysis. After summarizing the total number of stops by race, we then divided the total stops by each race's population count in Long Beach and multiplied the result by 1,000 to get stop rates per 1,000 people of the same race.

Calculating traffic stop rates that result in no action by race

As an indicator of pretext, this analysis looked at the rate of traffic stops that resulted in no action by the race of persons stopped. Traffic stops were filtered to only include officer-initiated stops for traffic violations, as opposed to calls for service or stops made for other reasons. The rates were calculated by taking the total number of traffic stops that resulted in no action for each racial group, divided by the total number of officer-initiated stops done for traffic violations for that racial group. The result tells us what percentage of traffic stops for each race resulted in no action or no enforcement needed.

Calculating rates of no contraband or evidence found by race

To calculate rates of officers finding no contraband by race, we first filtered the data to include only officer-initiated stops that were made for traffic violations. Then, we filtered the data to include stops where a person or their property was searched. We summarized whether or not contraband or evidence was found by race. The rate at which officers find contraband or evidence from searches is often referred to as a 'hit rate' or the success rate of searches. Hit rates are often used as an indicator for racial bias – the argument being if officers are not discriminating by race, then the hit rate should be similar across all races.[11] We calculated hit rates by dividing the total number of observations where a search was conducted, and contraband was found, by the total number of traffic stops where a search was conducted for each race. We visualized the inverse of the hit rate to show how officers are not effective at determining when to stop and search an individual given the high rate of searches that yield no contraband across races and the disparity for certain people of color.

Calculating Geographic Disparities in Stop Rates

Analyzing stops by council district and race

We created a map to visualize officer-initiated stops by council district and race. First, we summarized the total number of officer-initiated stops by race and council district. Then, using a point density tool available in the programming language, R, we mapped the distribution of stops by race randomly across each council district with different colored points. Each point represents 10 people stopped, with the color representing one of eight racial groups as indicated in the map legend. Because the points are randomly distributed throughout the respective council district, we use them to visualize the concentration of stops for people of different races in that district. We cannot use them to determine information about the specific locations (e.g., addresses, streets, blocks, etc.) of those stops. For the stop rates included in the map pop-up, we summarized the total number of officer-initiated stops by

---

[11] Open Policing, *Findings: The results of our nationwide analysis of traffic stops and searches,* Stanford University. Retrieved from https://openpolicing.stanford.edu/findings/.

Commented [HK10]: Same as other comments - we should define "officer-initiated stops" somewhere first (either one time at the top of the "How We Create Our Estimates" section or wherever we reference it first. Otherwise, this filtering statement is a little redundant and doesn't clarify what has been excluded or why.

Commented [HK11]: Could add a line to explain why pretext is useful to understand

Commented [HK12]: do we need to mention leaflet as well?

race and council district, divided by the estimated population of each race in each council district, and multiplied the result by 1,000 to get a stop rate per 1,000 people of the same race.

Analyzing hours spent on officer-initiated stops that result in no action by council district
We created a map to visualize the hours spent on officer-initiated stops that resulted in no action, or no enforcement, by council district. For our analysis, we calculated both the total hours and the percentage of hours spent on officer-initiated stops that resulted in no action (e.g., no citation, no warning, etc.) out of all officer-initiated stops. Our calculations control for stops where more than one person was stopped. For example, if one person was arrested in the stop and one person had no action, the time spent on that stop is not included in the 'no action' hours. In other words, stops with multiple people are only included in this calculation if all persons stopped resulted in no action.

Mapping Traffic Safety

Visualizing high-injury networks
For this section, we scraped and geocoded the high-injury intersections and corridors from the "Safe Streets Long Beach: A Vision Zero Action Plan" report. Safe Streets Long Beach is an initiative to reduce traffic-related fatalities and serious injuries to zero by 2026. We pulled data from the report which identified the top 20 high-injury intersections and corridors for pedestrians and bicyclists, and the top 20 high-injury corridors and intersections for motorists and motorcycle riders from 2013-2017. The City of Long Beach identified these high-injury intersections and corridors based on the frequency and severity of reported collisions as reported to the Statewide Integrated Traffic Records System (SWITRS). Collision severity focuses on incidents where people were killed or seriously injured. We visualized the start and end points of each corridor and the exact address location for each intersection.

Commented [HK13]: Should we add a sentence on our choice to combine the mapping of high injury networks with traffic stop rates by CD?