

# CRSTAL: A Benchmarking Framework

CENSUS  
RISK-ADJUSTED DISPARITY  
SUSPECT ADJUSTED DISPARITY  
TRAFFIC ANALYTIC LAYOUT



SAN FRANCISCO POLICE DEPARTMENT  
CITY & COUNTY OF SAN FRANCISCO



# Issue Framing: Analysis of Stop Data

---

- Analyzing stop data is an opportunity for SFPD to understand and interpret our officer's work, including community contact beyond crime statistics.
- When analyzing stop data, the most pertinent question is how to frame the context of the stop data. In other words, it is the responsibility of SFPD to determine the most appropriate “full group”(all persons susceptible to police contact) to compare the “subgroup” (those with documented police contact) within the stop data.
- In other words – who does the SFPD come into contact with, and as compared to what?



# Background: Benchmarking

---

## What is a Benchmark?

- Benchmarking is a process of comparison for the purpose of improvement. In the context of stop data, benchmarking refers to the process of comparing the subgroup – stop data- to a ‘full group’ to generate numbers for comparison.
- Identifying the ‘full group’ in academia is referred to as the “denominator problem”.
- Selection of an appropriate benchmark is considered to be the most controversial part of research in racial profiling scholarship (Smith et. al, 2021)



# Background: Benchmarking

---

- All benchmarks have their limitations.
  - In addition to limitations of benchmarks there are limitations with the stop data:
    - it is self-reported
    - demographic characteristics are based on officer perception
    - data intake can be inconsistent
    - information can be incomplete
- There is no universally agreed upon benchmark that can be used to analyze stop data to determine racial bias (Ridgeway & MacDonald, 2010)



# Current Practice: Census Benchmark

---

## Standard use

- Compares the stop data by race/ethnicity to the residential population of that same race/ethnicity

## Pros

- A well-known benchmark and current status quo
- Often used by researchers due to its availability and ease to understand (Ridgeway & MacDonald, 2010, Smith et. al, 2021)

## Assumptions

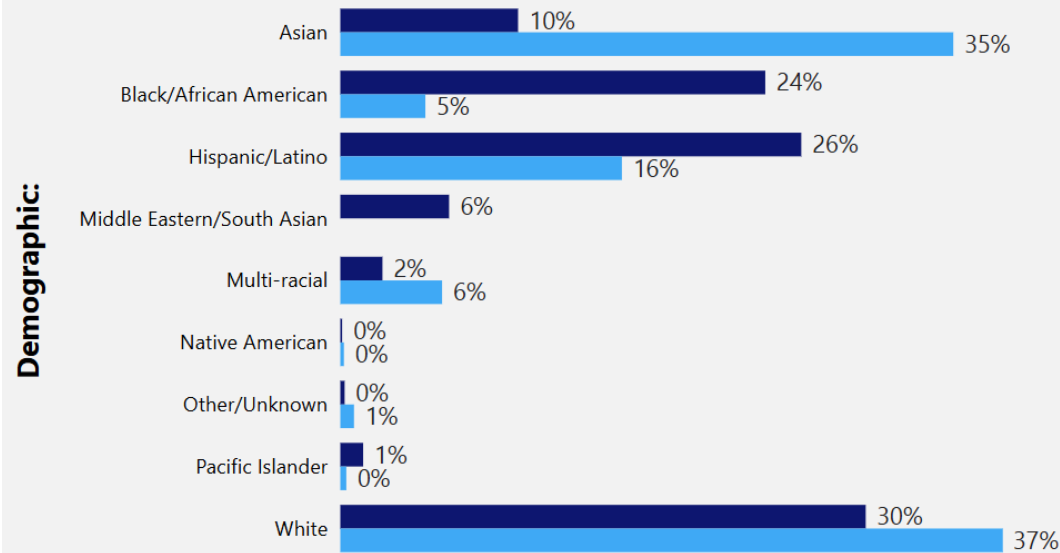
- Stops made by police are random
- All stops are of City residents



# Population Benchmark (New Dashboard)

## Comparing San Francisco's estimated population and relative stops by perceived race or ethnicity for Calendar Year 2023

● Percent of stops ● Percent of population



Demographic Group	Number of Stops	Percent of Stops	Population Estimate	Percent of Population
Asian	2,072	10%	290,116	35%
Black/African American	4,954	24%	40,243	5%
Hispanic/Latino	5,374	26%	133,266	16%
Middle Eastern/South Asian	1,268	6%		
Multi-racial	491	2%	48,160	6%
Native American	21	0%	1,691	0%
Other/Unknown	50	0%	6,461	1%
Pacific Islander				
White				
<b>Total</b>	<b>20,600</b>	<b>100%</b>	<b>836,321</b>	<b>100%</b>



# Additional Benchmarks: CRSTAL

---

- CRSTAL stands for:
  - Census
  - **Risk-adjusted**
  - **Suspect-adjusted**
  - **Traffic Analytic Layout**
- The implementation of CRSTAL adds three benchmarks to the census benchmark.
- The intent of CRSTAL is not to emphasize one benchmark over another but to provide the public with additional ways to interpret the stop data .
- The additional benchmarks: **risk-adjusted, suspect-adjusted, and traffic analytic layout** data add a level of context to the work the SFPD has a duty to perform.



# Risk-Adjusted Disparity

---

## Standard use

- Compares stop data by race/ethnicity to the number of violent crime victimizations of the same race/ethnicity (risk of being a violent crime victim) then compares that ratio to the ratio of stop data of whites and their violent crime victimization.

## Pros

- Uses victimization by race/ethnicity for data completeness
- Creates an easy ratio to compare across racial categories
- Reflects department priorities – addressing violent crime and addressing crime for those most at risk of being victimized.

## Assumptions

- Most violent crimes are intra-racial (victim and suspect are the same race)
- All stops are in furtherance of addressing violent crime

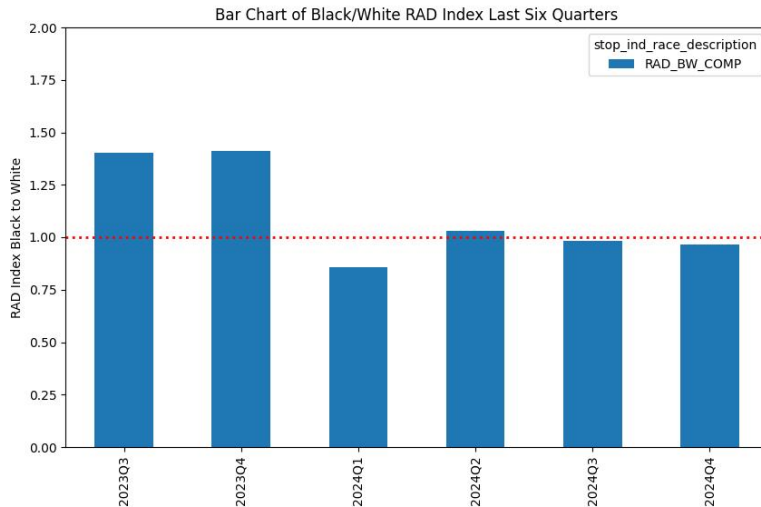
**Citation:** Sherman, L.W., Kumar, S. (2021). Equal Protection by Race with Stop and Frisk: A Risk-Adjusted Disparity (RAD) Index for Balanced Policing. Cambridge Journal of Evidence-Based Policing 5(1), 1-19.



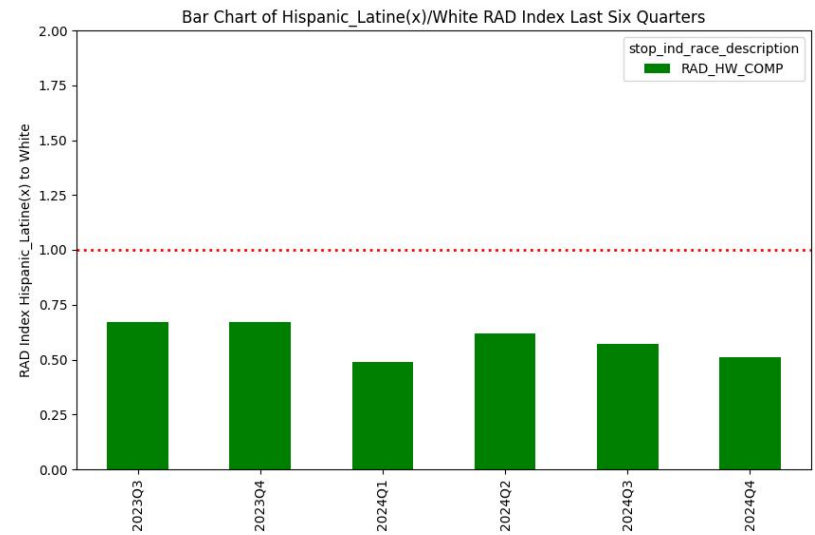


# RAD Index: Black:White & Hispanic:White

**Table 1: RAD Index  
Black:White**



**Table 2: RAD Index  
Hispanic:White**



A RAD index of 1.0 (marked by the red line) indicates no difference between stops:victim ratio of the minority racial group compared to the stops:victim ratio of the majority racial group. In Table 1 the stopping of Black individuals varies in comparison to white individuals on a quarterly basis using RAD. In Table 2 the stopping of Hispanic/Latine(x) individuals is consistently under the stopping of white individuals on a quarterly basis using RAD.



# Suspect-Adjusted Disparity

---

## Standard use

- Compares stop data by race/ethnicity to the number of reported violent crimes where suspect descriptions are of the same race/ethnicity then compares that ratio to the ratio of stop data of whites and their number of reported violent crimes where the suspect description race/ethnicity is white

## Pros

- More directly approaches RAD index intent while avoiding homogeneity challenges
- Relies on suspect information which narrows population of those contacted by police
- Reflects department priorities to focus on violent crime

## Assumptions

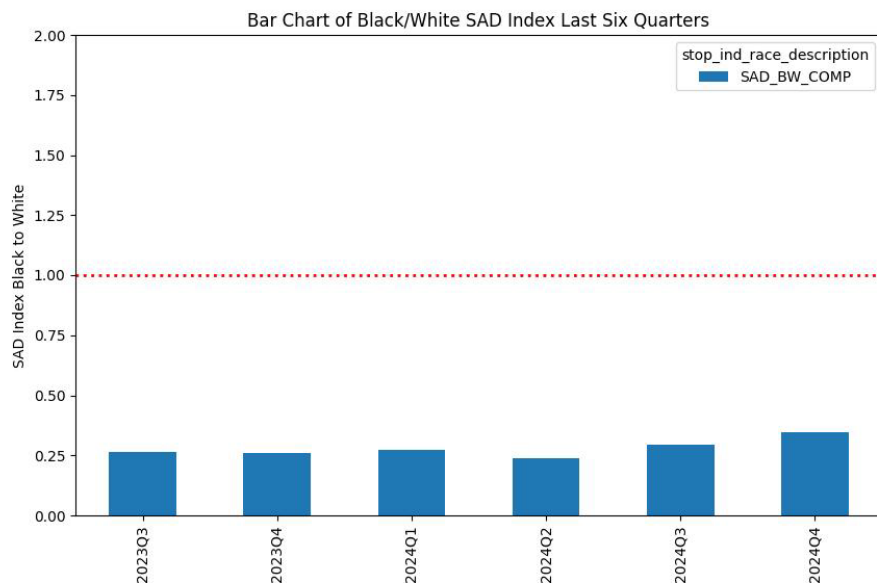
- All stops are in furtherance of addressing violent crime
- All stops are of the correct suspect
- Does not account for one suspect who committed repeated violent crimes

**Citation:** Exploration of an alternative approach to calculating stop and search rates in the Metropolitan Police Force Area – Experimental Statistics, 2021. England & Wales, United Kingdom.

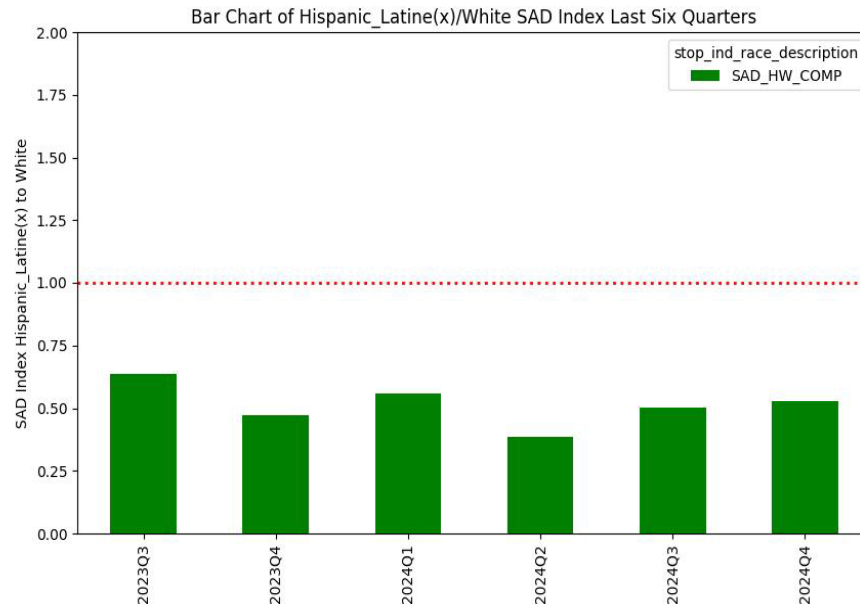


# SAD Index: Black:White & Hispanic:White

**Table 3: SAD Index  
Black:White**



**Table 4: SAD Index  
Hispanic:White**



A SAD index of 1.0 indicates no disparity between stops:suspect ratio of the minority racial group compared to the stops:suspect ratio of the majority racial group. In Table 3, Black individuals are stopped less than white individuals on a suspect adjusted basis. In Table 4, Hispanic/Latine(x) individuals are stopped less compared to white individuals on a suspect adjusted basis.



# Traffic Analytic Layout

---

## Standard Use

- To understand if a group's representation in the stop data is over or under their representation in the random driver population by using 'not at fault' collision data.

## Pros

- Compares the race/ethnicity of traffic stop data to the not at-fault collision data
- Not at-fault data mirrors a random population sample

## Assumptions

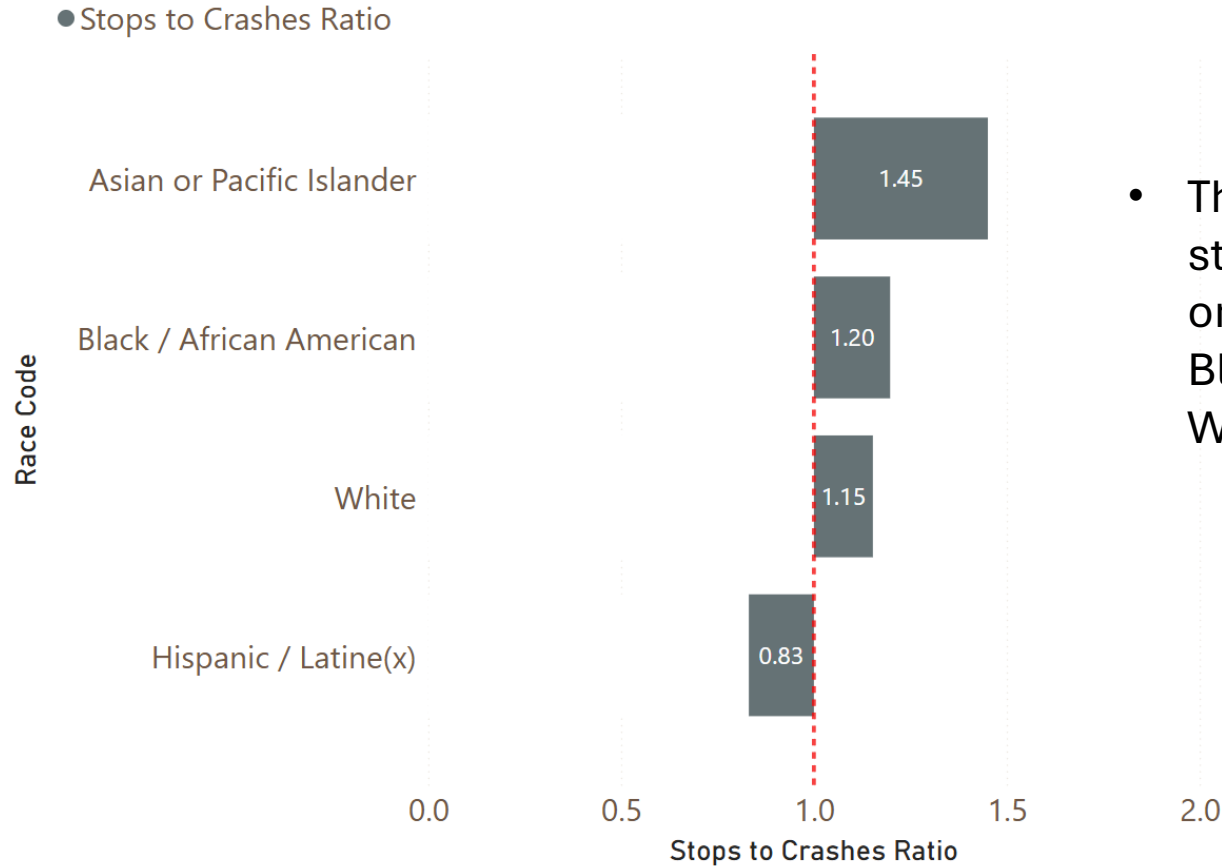
- Only analyzes traffic stop data
- Collision data only includes injury collisions
- Samples are drawn from a normally distributed population

**Citation:** Alpert, G.P., Smith, M.R., Dunham, R.G. (2004) Toward a better benchmark: Assessing the utility of Not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy* 6(1), 2-27.



# TAL Crash Comparison Results

## SFPD Stops to Crash Ratio by Race/Ethnicity



- There are proportionally more stops than crashes for Asian or Pacific Islander, Black/African American and White groups.



# RAD, SAD and TAL: The Math

Risk Adjusted Disparity & Suspect Adjusted Disparity are analyses of a ratio of ratios using different denominators.

## Risk Adjusted

### Step 1

$$X1 = \frac{Stops_{Black}}{Victims_{Black}}$$

### Step 2

$$X2 = \frac{Stops_{White}}{Victims_{White}}$$

### Step 3

$$RAD = \frac{X1_{Black}}{X2_{White}}$$

## Suspect Adjusted

$$X1 = \frac{Stops_{Black}}{Suspect_{Black}}$$

$$X2 = \frac{Stops_{White}}{Suspect_{White}}$$

$$SAD = \frac{X1_{Black}}{X2_{White}}$$

## Not At Fault Traffic Analytic Layout

### Step 1

Traffic Analytic Layout is a statistical analysis of crash data and stops data using a z-score mean comparison of groups.

### Step 2

Calculate the standard error of proportions to determine expected variance between groups.

Calculate the z-scores to compare the differences between proportions versus expectations.

### Step 3

Calculate the p-value to determine the probability of seeing this value based on chance.



# Conclusions

---

The growth of the Department's disparity measurements via the CRSTAL set of benchmarks seeks to enrich and add context to the ongoing conversation around who the SFPD interacts with.

Things to take away:

- There is no one 'best' benchmark – each comes with its own specific strengths and weaknesses. Therefore, read the benchmarks together and not in isolation.
- Each benchmark takes a different approach to observing disparity. Each is correct, and each observes something different to varying degrees.



# Data Sources and Additional Information

---

- The 4<sup>th</sup> Quarter 2024 Quarterly Activity and Data Report has 20 pages of analysis and explanation. [Quarterly Activity & Data Report \(QADR\) | San Francisco Police Department](#)
- Stop Data Dashboards are live and published here: [SFPD Stop Data Dashboards | San Francisco Police Department](#)
- CRSTL Code and summary data tables are posted on the SFPD GitHub, here: [GitHub - sfpd-public/crstal\\_analysis](#)
- SFPD stop data thru Q4 2024 is posted here: [Police Department Stop Data | DataSF](#)

