Ratcliffe and Hyland *Crime Science* (2025) 14:10
https://doi.org/10.1186/s40163-025-00252-y

Crime Science

RESEARCH Open Access

Police stops and naïve denominators

Jerry H. Ratcliffe[1*] and Shelley S. Hyland[2]

Abstract

A comparison of the racial composition of police stops to the entire population of a city or jurisdiction is frequently cited as evidence of racial bias in proactive policework. This article argues that using base population is naïve to the realities of the distribution of crime and policing. Using the example of Philadelphia, PA (USA), the impact of different benchmarks to estimate racial disparity in stop data is demonstrated. The range of alterative benchmarks include the spatial distribution of calls for service, the locations of violent crimes, and the demographic composition of suspects in crime as reported by the public. The article concludes by arguing that if cities ask police departments to prioritize certain problems and places, benchmarks to which police are held accountable should better reflect those priorities.

Keywords Police stops, Racial disparities, Denominator problem, Philadelphia

Introduction

It is commonly accepted that crime is unequally distributed across places and clustered in small areas (Sherman et al., 1989; Weisburd & Eck, 2004). Hot spots policing, where police activities skew away from mimicking the spatial distribution of the general population in favor of concentrating in the highest crime locations, was established on this evidence (Braga & Weisburd, 2022; Weisburd et al., 2019). It has become a popular technique; 75% of all officers employed by local police departments in the United States were in agencies that used data for hot spot analysis (Goodison & Brooks, 2023). Policing within these hot spots can vary but proactive work, such as preventative patrols through vehicle and pedestrian stops, are popular tactics. In 2022, residents age 16 or older had nearly 83 million contacts with police in the United States, and traffic and pedestrian stops comprised 27% of these contacts (U.S. Bureau of Justice Statistics, 2024). Both types of proactive investigation (vehicle and pedestrian stops) have been under considerable scrutiny for some time and the disparity in police stop rates is frequently cited as evidence of racial and ethnic bias in policing. Issues around bias in proactive policing has become a perpetual topic for academic study, journalism, lawsuits, and legislation (Ratcliffe et al., 2024).

Raw numbers can be contextualized by comparison to a baseline, but finding a suitable baseline metric is vital. Schlaud et al., (1998: 14S) use the metaphor of the 'iceberg phenomenon', where the choice of denominator "determines the validity of results to a high degree" (p. 19S). Contention often arises when comparing the percentage of a race or ethnicity represented in police data to the total population of that race/ethnicity across the jurisdiction (Carvalho et al., 2022; Smith et al., 2021, 2022). However, the use of Census data as a denominator has been criticized in the estimation of racial bias (Hannon, 2020; Ridgeway, 2007). While easy to calculate, it does not reflect the population at risk; that is, the population exposed to policing activities (Ridgeway & MacDonald, 2010; Smith, et al., 2022). It is oblivious to the context of travel patterns of the public, where police concentrate interventions, and the tasks they are asked to undertake.

While considerable effort goes into measurement of criminal justice activities, as with epidemiology and

*Correspondence:
Jerry H. Ratcliffe
jerryr@upenn.edu
[1] Department of Criminology, University of Pennsylvania, Philadelphia, PA, USA
[2] Bureau of Justice Statistics, U.S. Department of Justice, Washington, DC, USA

other disciplines, "comparatively little attention is paid to measuring the size of the population at risk" (Morrison et al., 2020: 1213). This 'denominator problem' or 'iceberg phenomenon' afflicts many fields (Morrison, et al., 2020; Schlaud, et al., 1998; Skopek, 2021; Tregle et al., 2019), and criminal justice is no exception (National Research Council, 2004). While the inappropriateness of a simple population-based count or rate has been long recognized (Walker, 2001), general population in a jurisdiction, absent any policing or crime context, remains a favored metric for critics in the media (Caiola, 2023; Duggan, 2019) and academia (Hinton et al., 2018).

With policing, the denominator represents the group potentially exposed to law enforcement interdiction. The composition of this group can differ depending on the policing activity. For example, while investigations of pedestrians will likely consist of city residents who reside nearby, this does not necessarily hold for business districts, entertainment hot spots, or tourist areas that can attract people from beyond the city limits. Similarly, exposure to traffic stops may involve a more affluent demographic (who can afford a car), commercial delivery workers from other cities, or commuters merely passing through the area with little connection to the neighborhood. Other factors may influence exposure, including where cities choose to prioritize (or reduce) police activity, the specific assignments and priorities given to officers, the distribution of specialized units, the reward mechanisms to which officers respond (such as arrests to generate overtime at court), or the volume of calls that might limit time available for proactive work.

Various alternatives to a simple population rate have been proposed, such as the 'veil of darkness' (Grogger & Ridgeway, 2006) or the "behavior of other officers working comparable assignments" (Walker, 2001: 89). One group recommended examining racial disparities "among perceived race of persons stopped when controlling for age, gender, offense type, and neighborhood context (e.g., crime, poverty)" (Pryor et al., 2020: 11). While these methods may have internal validity (Ratcliffe, 2023), they are likely too computationally sophisticated for many towns and cities to monitor regularly and effectively.

This article outlines some potential denominators that are less naïve to operational policing reality than citywide population, accessible to most police departments (at least in the United States), take advantage of the growth of the National Incident-Based Reporting System (NIBRS), are applicable to pedestrian and vehicle stops, and not computationally sophisticated. Several benchmarks are based on spatial data at the census tract level, a commonly used spatial unit in criminological literature (Andresen, 2006; MacDonald & Braga, 2019). As noted, crime and policing are differentially concentrated across places so controlling for location provides increased precision compared to citywide data. However, the paper also examines the distribution of calls for service and citywide NIBRS data focusing on suspect and arrestee information (at the time of writing, NIBRS data do not retain intra-jurisdictional spatial information). Both of these have been examined as benchmarks in other studies that have rejected citywide population rates (Ridgeway, 2007; Smith, et al., 2022).

The differences between benchmarks (or denominators) are demonstrated and discussed using a static citywide count of 2022 pedestrian and vehicle stops from the city of Philadelphia, Pennsylvania (PA).

Location

In 2022, Philadelphia, PA, was the sixth largest city in the United States, with the fourth largest police department. The American Community Survey (U.S. Census Bureau, 2022) estimated that in 2022, Philadelphia had 620,615 Black non-Hispanic, 534,291 White non-Hispanic, and 249,723 Hispanic (of either race) residents. These three groups comprised more than 88 percent of the city's 1,593,208 population. For demonstration purposes, this article focuses on just (non-Hispanic) Black and White resident populations, and vehicle and pedestrian stop rates. In 2022, there were 133,448 stops of both types recorded by the police department. Of these, 94,552 (70.9%) were of Black persons and 18,867 (14.1%) White persons.[1] Citywide, the vast majority were traffic stops (just over 90%) with the remainder being pedestrian stops.

Methodology

Like other studies that have examined racial disparities in criminal justice (Tregle, et al., 2019), an odds ratio can indicate the disparity between stop rates and various benchmarks. Here it is based on the Black stop rate divided by the White stop rate, with the respective rate calculated as the number of race-specific stops divided by the race-specific benchmark (in this first example, the base metric of overall city population):

$$\text{Odds ratio} = \frac{\textit{Black stops}/\textit{Black citywide population (benchmark)}}{\textit{White stops}/\textit{White citywide population (benchmark)}}$$

Benchmark 1: population. In 2022, 85,850 Black people and 16,640 White people were stopped by Philadelphia police in vehicle stops, and 8702 and 2227 Black and White people were stopped by police in pedestrian stops respectively. Replacing the equation terms with the

[1] The residential status of persons stopped was unknown so stops may include non-residents.

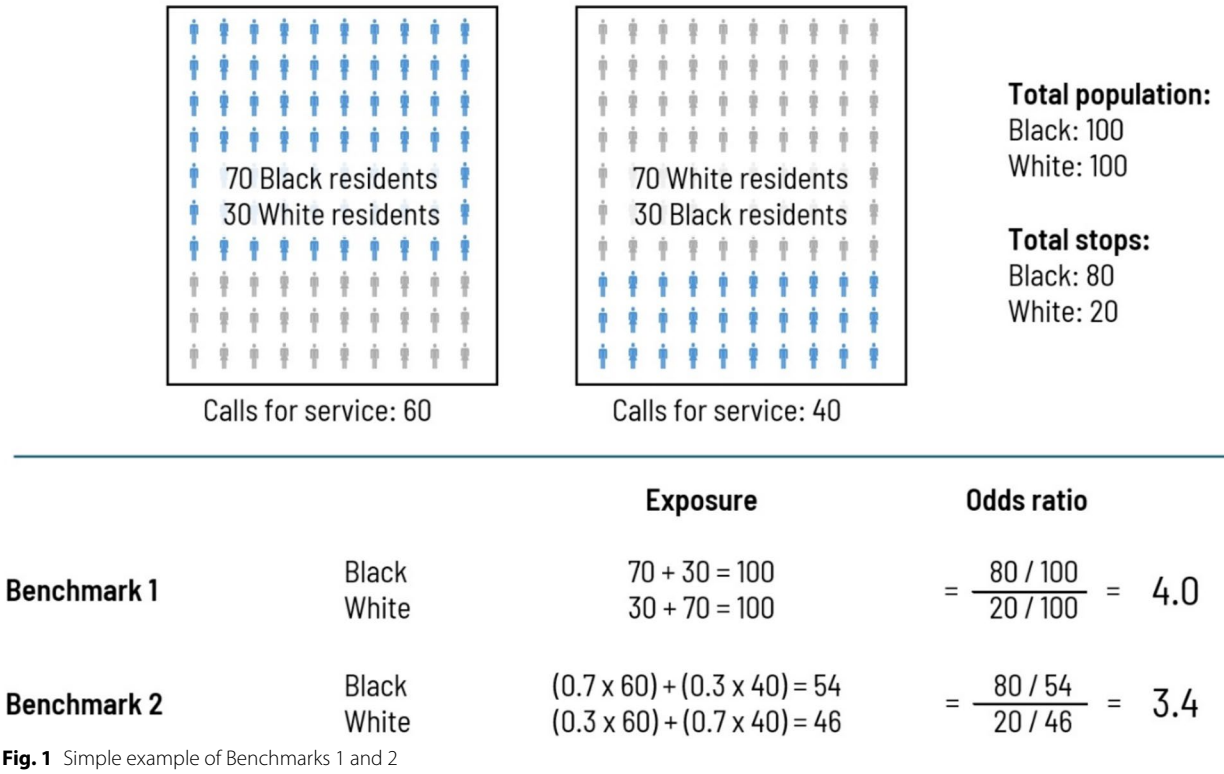


Fig. 1 Simple example of Benchmarks 1 and 2

Philadelphia vehicle stop numbers, we get $\frac{85{,}850/620{,}615}{16{,}640/534{,}291} = 4.442$

We interpret an odds ratio *greater than 1.0 that Black people are more likely to be stopped* relative to the benchmark metric. Using the citywide population in 2022, Black people were 4.4 times more likely to be stopped in vehicles by police than White people. For readers who prefer a percentage change, this can be calculated as:

$$\text{Percent change} = (\textit{Odds ratio} - 1)x100$$

Another way to express the disparity for our first benchmark (Benchmark 1: population) is to say Black people were 342 percent more likely to be stopped in vehicle stops than White people.

But population isn't the only benchmark that could be used. Here we outline eight other possible benchmarks and report them using odds ratios and 95 percent confidence intervals, calculated using the 'epitools' R package (Aragon, 2020).

Benchmark 2: calls for service by census tract. One alternative is to examine where police calls for service (CFS) concentrate. This acts as a proxy measure for where police are required to spend time responding to public requests for assistance. To demonstrate this, 2022 CFS data recorded by the Philadelphia Police Department were filtered to only calls for service from the public (thus excluding officer-initiated) and where officers were dispatched (excluding calls for information only or dealt with in ways that did not require sending an officer). In 2022, there were 1,533,973 such calls and of these 1,517,729 (98.9%) could be mapped to a city census tract. These tract totals were multiplied by the ratio of Black and White residents in each tract (as fractions of overall tract population).

The effect of this difference is shown in Fig. 1 where two hypothetical census tracts have different racial compositions. In this example scenario we only know that across both census tracts there are a total of 100 stops, 80 of which are stops of Black persons. Absent the context provided by the distribution of police services by calls for service, Benchmark 1 has an odds ratio of 4.0, while incorporating the distribution of calls for service as a measure of exposure to police (Benchmark 2) provides an odd ratio of 3.4. This difference occurs because residents of the census tract on the left are exposed to police to a greater extent. Due to officers responding to more calls for service in that census tract, officers spend more time there. It alters the distribution of likely police exposure.

Benchmark 3: CFS multiplied by number of officers and call time by census tract. A previous Philadelphia study

demonstrated both spatial variation and large differences in how long the police attend different types of calls (Ratcliffe, 2021). Anticipating similar variance, the 2022 census tract CFS count (from Benchmark 2) was multiplied by the number of officers attending and the time each call was active (in minutes). This better reflects how many officers and how much time is spent on calls in different tracts. This number was multiplied by the ratio of Black and White residents in each census tract (like Benchmark 2).

Benchmark 4: priority one calls from Benchmark 3. This benchmark uses the same data established for Benchmark 3, filtered to include only high priority calls. This benchmark reflects only the most urgent and serious calls for service that are likely to be of strategic importance. These 'priority one calls' include person with a gun, person screaming, and robbery in progress. This benchmark may reflect where officers perceive the city has a high crime problem, an issue they are often asked to prioritize.

Benchmark 5: part 1 violent crime by census tract. Staying with the geography of policing, we can replace CFS (from the three previous benchmarks) with reported crime from the police crime database.[2] We switch to crime incidents since CFS data includes calls that may not result in a crime report or may not reflect what ultimately happened. Serious violent offenses (murder, rape, robbery, and aggravated assault) from the Federal Bureau of Investigation's (FBI) Uniform Crime Reporting (UCR) Summary Reporting System Part 1 offenses were mapped to census tracts. In 2022, there were 15,154 incidents of serious violence, of which 14,915 (98.4%) could be mapped to a census tract. The ratio of Black and White residents in each census tract (compared to the overall tract population) was multiplied by the number of serious violent crimes in the tract in 2022.

The remaining four benchmarks move away from spatial variation to take advantage of NIBRS data to examine individual characteristics of people arrested for or suspected in a crime.

Benchmark 6: NIBRS arrestee data for all crime. The Philadelphia Police Department reports incident-level crime data to the FBI's UCR program through NIBRS. NIBRS is a comprehensive data system, and "the richness and breadth of the NIBRS data are some of its greatest strengths" (Lantz & Wenger, 2020: 406). NIBRS incident data include detailed information on each incident as well as victim, offender, and arrestee characteristics. Most incidents had three or fewer arrestees, and Benchmark 6 sums the number of arrestees (up to three arrestees) reported as Black, and the number reported as White. In 2022, there were 137,881 total NIBRS incidents from Philadelphia with 18,864 reported arrestees across all races and ethnicities (maximum of first three used).

Benchmark 7: NIBRS suspect data for all crime. Benchmark 6 is vulnerable to criticism that arrestees may reflect bias towards arresting Black citizens. An alternative is offender information in NIBRS, drawn from two other sources: the victim or the officer (if they directly witnessed the incident). To reduce concerns regarding bias, offender information associated with a person arrested as either an on-view arrest or summonsed/cited were excluded. The offender information that remains should therefore comprise either suspects described by the victim or arrested on a warrant or previous information from the victim. Of the 137,881 incidents in 2022, 38,735 suspects (28.1%) fulfil this criterion.

Benchmark 8: NIBRS arrestee data for violent crime only. This uses Benchmark 6 data (arrestee information from NIBRS) filtered for only serious violent crime. In NIBRS coding, these are (091) Murder/Nonnegligent Manslaughter, (092) Negligent Manslaughter, (111) Rape, (120) Robbery, and (131) Aggravated Assault. From just over 15,000 serious violent incidents recorded in NIBRS, information was included for 3933 arrestees (all races and ethnicities).

Benchmark 9: NIBRS suspect data for violent crime only. This uses Benchmark 7 data (offender information from NIBRS) filtered for only serious violent crime. Information was available for 7908 suspects (all races and ethnicities).

Results

Table 1 shows how the benchmarks compare across vehicle stops. The racial disparity (odds ratio) varies with the use of different benchmarks and in the case of Benchmarks 8 and 9 flips direction.

For nearly all the benchmarks, the odds ratio remains indicative of a racial disparity in stops. The metric that is primarily generated by the public are Benchmarks 7 and 9. Relative to all crimes reported to NIBRS and based on the racial description of suspects recorded in NIBRS (Benchmark 7; odds ratio = 1.085), Black people are just 8.5 percent more likely to be stopped than White people. This is a sizable reduction compared to the 344 percent greater rate when the citywide population benchmark is applied (Benchmark 1). When the denominator focuses only on persons arrested for or suspected in serious violence (Benchmarks 8 and 9), the disparity flips.

[2] Philadelphia crime incidents are stored in a Premier One Records Management System (P1RMS) (Philadelphia Police Department, 2021). The data provided to the authors from P1RMS did not include any suspect demographic information.

Table 1 Various denominators and odds ratios for vehicle stops in Philadelphia, PA, 2022

	Black persons	White persons	Odds ratio	Lower c.i	Upper c.i
Vehicle stops	85,850	16,640			
Benchmark					
Baseline					
(1) Citywide population	620,615	534,291	4.442	4.367	4.517
Calls for service by census tract					
(2) CFS, by tract	650,020	412,806	3.276	3.221	3.333
(3) CFS x officers and time, by tract[1]	39,196,043	21,486,399	2.828	2.782	2.876
(4) Benchmark 3's priority one calls[2]	16,039,750	7,448,301	2.396	2.356	2.436
Part 1 violent crime by census tract					
(5) Part 1 violent crime locations by tract	7049	3319	2.429	2.323	2.539
NIBRS data (citywide)					
(6) NIBRS arrestee information, all crime	10,591	2756	1.343	1.283	1.404
(7) NIBRS suspect information, all crime	27,402	5761	1.085	1.050	1.121
(8) NIBRS arrestee info, violent crime only	2742	441	0.830	0.749	0.918
(9) NIBRS suspect info, violent crime only	6036	670	0.573	0.528	0.621

c.i. = confidence interval

[1] This number is the count of dispatched CFS from the public, multiplied by the ratio of each racial group's residents as a fraction of census tract population

[2] The total number of officers at each dispatched CFS from the public, multiplied by the total time of the call in minutes and the ratio of each racial group's residents as a fraction of census tract population

Given vehicle stops comprise 90 percent of city stops, Table 1 is also largely reflective of overall citywide stop rates. For completeness, pedestrian stop results are shown in Table 2. Again, for all of the spatial benchmarks (citywide and at the census tract level), there is evidence of racial disparity in stops. Examining the odds ratios, and given the confidence intervals, there is no statistically significant evidence of racial disparity in pedestrians stops when the benchmark is arrestee information for all crime (Benchmark 6). Additionally, White pedestrians are *more likely* to be stopped by police when suspect descriptions or arrestee information for serious violence are used (Benchmarks 8 and 9). The differences between all of these various outcomes are shown graphically in Fig. 2 where Black people are more likely to be stopped if the benchmark odds ratio (red circle or blue diamond) is to the right of the vertical dashed line, and less likely to be stopped if the symbol is to the left of the line.

Discussion

The limitations of administrative data recording by police departments are well known and acknowledged (Huey et al., 2022; Laniyonu & Donahue, 2023). It is also important to stress that the outcome measures in this article are counts of traffic and pedestrian stops, with reference to neither the reasons for the stops nor the result of the police action. Additionally, the residency status of those stopped by police was not known. Therefore, stops can include both residents and non-residents and are being compared to resident populations.

Overall, while the number of stops across the city is constant, the case study from Philadelphia shows that choice of denominator has a huge effect on racial disparity measurement. Benchmark 1 suggests the greatest racial disparities; however, it is also likely to be the least reflective of exposure to proactive police activity. Low crime areas can contribute substantial populations but rarely experience policing, and citywide population is unlikely to reflect equal exposure to traffic enforcement. Like others (MacDonald & Braga, 2019; Ridgeway, 2007; Smith, et al., 2022), our analysis demonstrates how the use of this benchmark can be problematic when estimating bias. Benchmarks 2–5 adjust the spatial exposure of the denominator to better reflect where officers are required to attend calls for service (2 and 3), priority calls for service (4), or violent crime calls (5). Even with these benchmarks, substantial racial disparity in stops exists. Except for pedestrian stops in Benchmarks 4 and 5, Black people are shown to be more than twice as likely to be stopped by police than White persons when census populations are used at the tract level, regardless of benchmark configuration.

If police departments prioritize addressing violent crime, Benchmark 5 is likely more reflective of where people are being exposed to police and is a simple benchmark for departments to track. However, using basic crime counts does not take into consideration

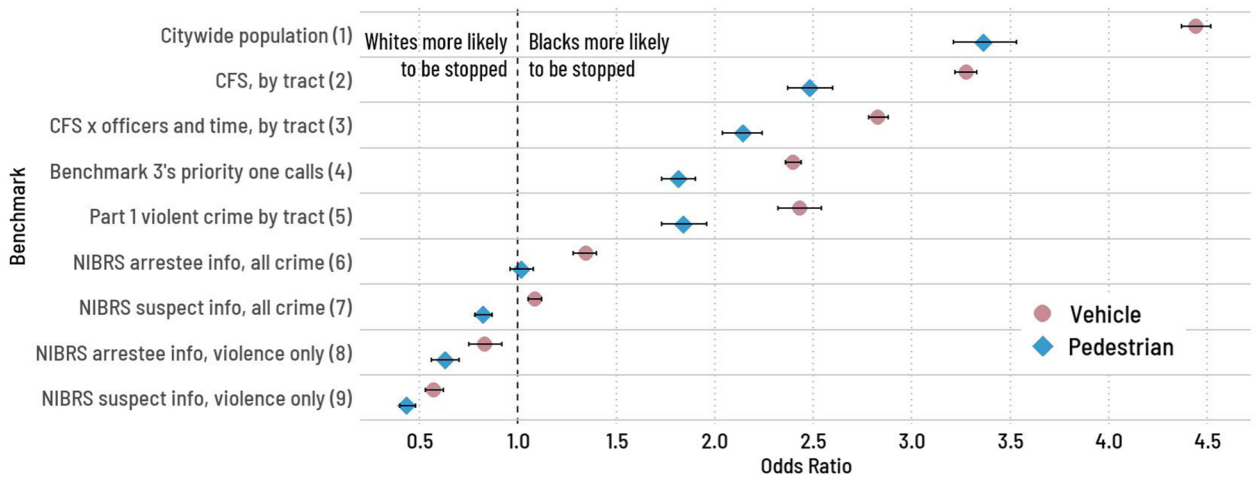
Table 2 Various denominators and odds ratios for pedestrian stops in Philadelphia, PA, 2022

	Black persons	White persons	Odds ratio	Lower c.i	Upper c.i
Pedestrian stops	8702	2227			
Benchmark					
Baseline					
(1) Citywide population	620,615	534,291	3.364	3.211	3.526
Calls for service by census tract					
(2) CFS, by tract	650,020	412,806	2.481	2.369	2.601
(3) CFS x officers and time, by tract[1]	39,196,043	21,486,399	2.142	2.045	2.244
(4) Benchmark 3's priority one calls[2]	16,039,750	7,448,301	1.814	1.732	1.901
Part 1 violent crime by census tract					
(5) Part 1 violent crime locations by tract	7049	3319	1.840	1.729	1.958
NIBRS data (citywide)					
(6) NIBRS arrestee information, all crime	10,591	2756	1.017	0.955	1.083
(7) NIBRS suspect information, all crime	27,402	5761	0.821	0.778	0.868
(8) NIBRS arrestee info, violent crime only	2742	441	0.629	0.562	0.702
(9) NIBRS suspect info, violent crime only	6036	670	0.434	0.395	0.476

c.i. = confidence interval

[1] This number is the count of dispatched CFS from the public, multiplied by the ratio of each racial group's residents as a fraction of census tract population

[2] The total number of officers at each dispatched CFS from the public, multiplied by the total time of the call in minutes and the ratio of each racial group's residents as a fraction of census tract population



Note: Colored markers indicate odds ratio, and black bars represent confidence intervals.

Fig. 2 Vehicle and pedestrian stop benchmarks with odds ratios and confidence intervals

the racial and ethnic composition of those engaged in crime.

Where Benchmarks 1–5 employ all citizens across the city (albeit in differing ways), Benchmarks 6–9 take a person-centered approach, and in particular, people who come to police attention. These benchmarks reflect small subsets of the city's population. Benchmarks 6 and 8 utilize the population of people arrested by the police. When all arrestees are included, vehicle stops still show a small racial disparity, but pedestrian stops indicate racial parity. When arrestee information for violence crimes (Benchmark 8) is used, the results show that Whites are more likely to be stopped than Black people; however, this benchmark is constructed from only 3183 individuals. This benchmark is vulnerable to the critique that a police department may be more effective at arresting Black suspects than White suspects, thus potentially skewing the denominator.

As noted by Ridgeway (2007: 13) "a valid benchmark requires that suspects, regardless of race, are equally exposed to police officers." Constructing a baseline metric from such a low number of individuals invites potential hidden bias (Ridgeway & MacDonald, 2010). Experienced officers can often recognize known offenders and the arrestee database may contain a number of individuals who are represented in the database multiple times.

Assuming the public report their victimization accurately, and it is recorded as such, Benchmarks 7 and 9 reflect a metric that originates from the community and is not as vulnerable to police-introduced bias. Benchmark 9 has the same significant limitation as 8, comprising only 6706 suspects, and may not be sufficient to capture exposure to police. Benchmark 7, however, is constructed with a greater number of people than the arrestee database, and because it originates with public—not police—information, may be appealing to police departments due to the external nature of the data source.

Which benchmark to use?

It is unlikely that any benchmark will accurately reflect the myriad range of police activities and satisfy both police executives and critics. It may be that benchmarks could be tailored to the police unit and their assigned duties. For example, if patrol officers are asked to focus on violent crime suppression in between responding to a range of public CFS, then Benchmarks 4 and 5 reflect proactive priority locations where officers may encounter violent acts or calls. A dedicated intelligence-led unit specifically tasked with interdicting serious repeat offenders will likely focus on suspect information, and Benchmark 7 may be more appropriate (remembering that 9 has significant limitations).

Traffic officers charged with accident prevention or managing bad driving may find that none of the benchmarks in this study are well suited for traffic stops. Literature supports the use of other benchmarks for traffic stops such as composition of drivers stopped during day vs night and racial composition of at-fault and not-at-fault accidents compared to those stopped (Alpert et al., 2004; Grogger & Ridgeway, 2006; Smith, et al., 2021). Policy makers seeking to construct a reasonable baseline denominator may choose to blend different benchmarks to reflect the diversity of policework reflected in CFS, crime locations, GPS of where officers spend their time, and the priorities handed down to frontline officers. Though how exactly to construct a blended measure will likely involve considerable complexity. One conclusion can, however, be drawn: When considering all these factors, Benchmark 1 (citywide population) is clearly divorced from the realities of where police officers concentrate their time.

In Philadelphia, on January 2, 2024, the city declared a public safety crisis and directed the police commissioner to "employ any lawful means necessary to abate" the emergency (City of Philadelphia, 2024: 2). The resultant crime plan promised to 'surge' resources to 10 of the 20 police districts in which 78% of shooting victims are struck, and employ "place-based and offender-focused tactics" (Philadelphia Police Department, 2024: 18). Focused deployment on specific people and places changes the population-at-risk of police intervention. Continued use of citywide population rates as a benchmark against which to measure racial bias in police activity would seem naïve at best, and deliberately misleading if deployed by more informed commentators.

Abbreviations

FBI	Federal Bureau of Investigation
UCR	Uniform Crime Reporting
CFS	Calls for service
c.i.	Confidence interval
NIBRS	National Incident-Based Reporting System

Acknowledgements

The authors would like to thank John MacDonald and George Kikuchi for their assistance with earlier versions of this manuscript.

Author contributions

All authors read and approved the final manuscript.

Funding

Funding was not received for this research.

Data availability

Police stop data are available for Philadelphia from OpenDataPhilly at opendataphilly.org. Population estimates for cities are publicly available from the U.S. Census Bureau, and National Incident-Based Reporting System (NIBRS) Series crime reports are available to download from https://www.icpsr.umich.edu/web/ICPSR/series/128. Calls for service records (used for benchmarks 2–4) and the Premier One Record Management System incident data (used for benchmark 5) were drawn from Philadelphia Police Department records accessed through a data sharing agreement and are not publicly available. For readers wishing to explore crime data that are not geocoded, the NIBRS records are recommended. The Originating Agency Identifier for the Philadelphia Police Department is PAPEP0000.

Declarations

Competing Interests

The author declares no financial interest. The views and opinions in this article are those of the authors and do not necessarily represent the official position and policies of the Bureau of Justice Statistics or the U.S. Department of Justice.

Received: 16 September 2024 Accepted: 15 May 2025
Published online: 09 July 2025

References

Alpert, G. P., Smith, M. R., & Dunham, R. (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), 43–69.

Andresen, M. A. (2006). Crime measures and the spatial analysis of criminal activity. *British Journal of Criminology, 46*(2), 258–285.
Aragon, T. (2020). Epitools: Epidemiology Tools. R Package, version 0.5–10.1.
Braga, A. A., & Weisburd, D. L. (2022). Does hot spots policing have meaningful impacts on crime? Findings from an alternative approach to estimating effect sizes from place-based program evaluations. *Journal of Quantitative Criminology, 38*(1), 1–22.
Caiola, S. (2023). As Philly mayoral candidates debate public safety strategies, activists are raising the alarm on stop and frisk. WHYY. Retrieved from https://whyy.org/articles/philadelphia-mayoral-elections-public-safety-strategies-debate-stop-and-frisk/. Accessed 13 June 2025.
Carvalho, A. A. S., Mizael, T. M., & Sampaio, A. A. S. (2022). Racial prejudice and police stops: A systematic review of the empirical literature. *Behavior Analysis in Practice, 15*(4), 1213–1220.
City of Philadelphia. (2024). Executive Order No. 1-24: Declaring a Public Safety Emergency. 1-4.
Duggan, P. (2019). A disproportionate number of D.C. police stops involved African Americans. Washington Post. Retrieved from https://wapo.st/3yT1dUg. Accessed 13 June 2025.
Goodison, S. E., & Brooks, C. (2023). *Local Police Departments, Procedures, Policies, and Technology, 2020 – Statistical Tables*. Bureau of Justice Statistics.
Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association, 101*(475), 878–887.
Hannon, L. (2020). An exploratory multilevel analysis of pedestrian frisks in Philadelphia. *Race and Justice, 10*(1), 87–113.
Hinton, E., Henderson, L., & Reed, C. (2018). *An unjust burden: The disparate treatment of Black Americans in the criminal justice system*. Vera Institute of Justice.
Huey, L., Ferguson, L., & Koziarski, J. (2022). The irrationalities of rationality in police data processes. *Policing and Society, 32*(8), 947–962.
Laniyonu, A., & Donahue, S. T. (2023). Effect of racial misclassification in police data on estimates of racial disparities. *Criminology, 61*(2), 295–315.
Lantz, B., & Wenger, M. R. (2020). Analyzing the NIBRS data: The impact of the number of records used per segment. *American Journal of Criminal Justice, 45*(3), 379–409.
MacDonald, J., & Braga, A. A. (2019). Did post-Floyd et al. reforms reduce racial disparities in NYPD stop, question, and frisk practices? An exploratory analysis using external and internal benchmarks. *Justice Quarterly, 36*(5), 954–983.
Morrison, C. N., Rundle, A. G., Branas, C. C., Chihuri, S., Mehranbod, C., & Li, G. (2020). The unknown denominator problem in population studies of disease frequency. *Spatial and Spatio-Temporal Epidemiology, 35*(100361), 1213–1220.
National Research Council. (2004). Fairness and effectiveness in policing: the evidence (No. Editors: Wesley Skogan and Kathleen Frydl). Washington, DC: Committee to Law and Justice, Division of Behavioral and Social Sciences and Education.
Philadelphia Police Department. (2021). Directive 12.6 (Abandoned Vehicles). Philadelphia, PA: City of Philadelphia.
Philadelphia Police Department. (2024). 100 Day Report: Prepared in response to Mayor Cherelle L. Parker's Executive Order No. 1-24. 53.
Pryor, M., Goff, P. A., Heydari, F., & Friedman, B. (2020). *A guidebook: Law enforcement intelligence, a guide for state, local, and tribal law enforcement agencies (Final Release Version)*. Office of Community Oriented Policing Services.
Ratcliffe, J. H. (2021). Policing and public health calls for service in Philadelphia. *Crime Science, 10*(5), 1–6.
Ratcliffe, J. H. (2023). *Evidence-Based Policing: The Basics*. Routledge.
Ratcliffe, J. H., Braga, A. A., Nagin, D. S., Webster, D. W., & Weisburd, D. (2024). When is a benefit a harm? The paradox of stop, question, and frisk. *Translational Criminology, Fall, 2024*, 8–10.
Ridgeway, G. (2007). *Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices*. Rand Corporation.
Ridgeway, G., & MacDonald, J. (2010). Methods for assessing racially biased policing. In S. K. Rice & M. D. White (Eds.), *Race, Ethnicity, and Policing: New and Essential Readings* (pp. 180–204). NYU Press.
Schlaud, M., Brenner, M. H., Hoopmann, M., & Schwartz, F. W. (1998). Approaches to the denominator in practice-based epidemiology: A critical overview. *Journal of Epidemiology and Community Health, 52*(1), 13S-19S.
Sherman, L. W., Gartin, P., & Buerger, M. E. (1989). Hot spots of predatory crime: Routine activities and the criminology of place. *Criminology, 27*(1), 27–55.
Skopek, J. (2021). Grandparent status and multigenerational families. In M. Kreyenfeld & N. F. Schneider (Eds.), *Research Handbook on the Sociology of the Family* (pp. 278–299). Edward Elgar.
Smith, M. R., Tillyer, R., Lloyd, C., & Petrocelli, M. (2021). Benchmarking disparities in police stops: A comparative application of 2nd and 3rd generation techniques. *Justice Quarterly, 38*(3), 513–536.
Smith, M. R., Tillyer, R., Smith, M., & Lloyd, C. D. (2022). Assessing police stops of pedestrians: Toward a new generation of benchmarks. *Urban Affairs Review, 58*(4), 1152–1181.
Tregle, B., Nix, J., & Alpert, G. P. (2019). Disparity does not mean bias: Making sense of observed racial disparities in fatal officer-involved shootings with multiple benchmarks. *Journal of Crime and Justice, 42*(1), 18–31.
U.S. Census Bureau. (2022). American Community Survey 5-Year Data (2009–2023).
U.S. Bureau of Justice Statistics. (2024). Police Public Contact Survey, 2022. Inter-university Consortium for Political and Social Research [distributor], 2024-10-03. https://doi.org/10.3886/ICPSR38872.v1
Walker, S. (2001). Searching for the denominator: Problems with police traffic stop data and an early warning system solution. *Justice Research and Policy, 3*(1), 63–95.
Weisburd, D., & Eck, J. E. (2004). What can police do to reduce crime, disorder, and fear? *The Annals of the American Academy of Political and Social Science, 593*(1), 43–65.
Weisburd, D., Majmundar, M. K., Aden, H., Braga, A. A., Bueermann, J., Cook, P. J., et al. (2019). Proactive policing: A summary of the report of the National Academies of Sciences, Engineering, and Medicine. *Asian Journal of Criminology, 14*(2), 145–177.

Publisher's Note