Sample Learning Agendas and Measurement Guide for Reimagining Police Stops

February 2023

The Lab @ DC
Office of the City Administrator
Government of the District of Columbia

THE LAB @ DC

GOVERNMENT OF THE DISTRICT OF COLUMBIA
MURIEL BOWSER, MAYOR

Contents

Introduction

In September 2019, Washington, DC's Metropolitan Police Department (MPD) first released data on police stops, consistent with the enhanced data collection called for by the Neighborhood Engagement Achieves Results Amendment Act.

MPD defines a stop as:

> A temporary investigative detention of a person for the purpose of determining whether probable cause exists to make an arrest. A stop is a seizure of an individual's person and occurs whenever an officer uses his or her authority to compel a person to halt, remain in a certain place, or to perform an act (such as walking to a nearby location where the member can use a radio or telephone). If a person is under a reasonable impression that he or she is not free to leave the member's presence, a stop has occurred.[1]

MPD's stops data report summarizes comprehensive data on all MPD stops, including traffic stops, investigative stops, and arrests. In their reporting, MPD highlighted disparities in police stops by race and called for further research to understand the role of racial discrimination in producing these disparities. Soon after, DC's Deputy Mayor for Public Safety and Justice and the City Administrator asked The Lab @ DC, DC Government's applied scientific team, to partner with research institutions in reimagining the role of police stops in public safety. This effort aimed to address three key questions:

1. Is there racial bias in stops made by MPD? If so, at what level(s) is this bias operating—among individuals, the police department as a whole, and/or beyond the police department, within or across jurisdictions?
2. What are the effects of police stops? What are their benefits? What are their harms? And for whom?
3. What research and policy efforts can DC and similar jurisdictions undertake to better understand police stops and to reduce harm while promoting equitable public safety?

The first phase of this effort was a virtual six-day *Reimagining Stops* Workshop Series in October 2020 co-hosted by the Georgetown University Law Center's Innovative Policing Program,[2] Howard University, and The Lab @ DC. The workshop was organized with the belief that gaining a more complete understanding of the experienced harms and benefits of police stops in Washington, DC requires an inclusive approach. More than 130 impacted community members, advocates, researchers, public servants, and police practitioners shared their expertise and experience. The series not only offered an opportunity to hear these perspectives, but also for interaction among participants to hash out difficult, complicated issues. Over nearly 24 total hours of workshop conversations, participants described their assessments of the current state

[1] Metropolitan Police Department. (2021). Stop Data Report, January - December 2020. Government of the District of Columbia, p 2.
[2] Now named the Center for Innovations in Community Safety.

of police stops as well as how to reimagine this practice and its role in public safety. These conversations and key takeaways are detailed in the Workshop White Paper.

The next phase of our efforts to reimagine the role of police stops focuses on advancing both policy and research. The Howard University Thurgood Marshall Civil Rights Center has developed Policy Considerations for Harm Reduction, while this document summarizes the existing research on stops and pushes for a solutions-oriented approach to research on police stops. Our goal is to create a clear, feasible, and community-centered suggestions for research on police stops in DC and in other jurisdictions nationally. Through three sample learning agendas, this document offers guidance on measuring racial bias in police stops, examining the broader harms and benefits of stops, and assessing promising opportunities for change. Within each learning agenda, we summarize the available literature, noting key takeaways and limitations, and set forth priority research questions on the role of stops in public safety to continue building evidence on what works.

Throughout this document, we underscore the importance of both distinguishing between racial disparities and racial bias and of discerning the different levels at which bias might operate (individual, department, community/city). In addition, we recommend research that examines a broader range of outcomes related to public safety and wellbeing. This approach permits a more complete accounting of the harms and benefits of stops, thus encouraging deeper consideration of the tradeoffs of different approaches to public safety. Taken together, rich data collection and rigorous analysis can illuminate whether the most effective solutions lie at the individual officer, department, and/or community/system level. As Workshop Series keynote speaker Dr. Phil Goff observed, this approach allows us to hold police accountable for what they can fix, as well as identify what additional actors and policy solutions are needed to not only reduce bias, but also reduce harm and improve public safety overall.

To identify the priority questions on the role of stops in public safety, we drew on the insights shared by workshop series participants as well as a robust literature review.[3] The extant research on police stops spans numerous disciplines, from legal analyses of the constitutionality of the practice, to qualitative research on the lived experiences of those stopped, to quantitative measurement of bias using administrative data on millions of stops, to experimental evaluations of interventions designed to reduce bias. We focused our review on what is currently known about racial bias in stops and how bias is measured, and the broader effects (harms and benefits) of police stops.

In three sample learning agendas that follow, we summarize key findings and gaps in the literature in:

1. racial disparities and bias in stops,
2. broader effects of police stops; and

[3] Research by Workshop attendees is denoted in this document with an asterisk (*) in the footnote where we first cite their work.

3. opportunities for change.

Within each focus area, we highlight priority questions for research, which range from foundational fact finding to program evaluation efforts. These questions (and suggested approaches to start answering them) aim to build on the existing evidence base and address the issues of the greatest public interest in our community.

Finally, we present a comprehensive measurement guide that details available methods as well as the advantages and limitations of each approach. This guide can be employed by any jurisdiction that seeks to take an objective and clear-eyed look at the effects police stops have in their community.

Taken together, the learning agendas and measurement guide aims to highlight rigorous, policy-relevant research opportunities to better understand the dimensions of this issue in the Washington, DC context, inform the design of solutions, and evaluate new interventions that reimagine police stops and deliver better public safety outcomes.

1. Sample Learning Agenda: Racial Disparities and Bias in Stops.

A. Key Findings

Social science and legal researchers have long contemplated the question of racial bias in policing and have applied a range of methods in an attempt to assess and quantify the presence of bias. We provide a brief overview of key findings from both literatures.

Social Science. Leveraging the increasing access to and quality of police-civilian interaction administrative data, social science researchers have taken a variety of approaches to measuring and understanding racial bias, or discrimination, in police stops. Researchers most often adopt Neil and Winship's (2019) definition of discrimination in their analyses:

> [A]n interaction [is] discriminatory if similarly situated individuals of different races are treated differently. Similarly situated refers to similarity in contextual, behavioral, and individual-level features. The strong assumption here...is that similarly situated individuals have equal probabilities of being engaged in criminal behavior and thus should be treated similarly.[4]

These studies apply a variety of different tests to assess biased behavior at the level of the individual officer-civilian encounter. Two types of tests—benchmark and outcome/hit-rate—are most commonly used, though each is associated with varying levels of limitations on how accurately or precisely bias is measured. The Limitations subsection (1B) below and Table 1 in Section 4 detail the usefulness and shortcomings of each method. Notwithstanding the caveats,

[4] Neil, R., & *Winship, C. (2019). Methodological challenges and opportunities in testing for racial discrimination in policing. *Annual Review of Criminology*, *2*, p. 77.

researchers studying data from a variety of jurisdictions have found evidence of persistent racial bias in who is stopped, ticketed, and searched by police:[5]

- In the largest study to date of traffic stops, Pierson et al. (2020) examine data on 100 million stops nationwide and find that Black and Hispanic drivers were stopped and searched more often than white drivers.[6] Notably, the authors also found that Black drivers were less likely to be stopped after sunset, when it is darker, and officers are less able to discern race. Their analyses suggest that racial bias plays a role in deciding who is pulled over.

- Research on stops in North Carolina, Illinois, [7] and California[8] also finds that, of drivers stopped, Black drivers are more likely to be searched compared to their white or Latinx counterparts. In Los Angeles, these disparities persist even after accounting for the race of crime victims and suspects in the locale, and for the probability of finding contraband (i.e., the hit rate) by race of the person stopped.[9]

- In another study examining stops made by the Florida Highway Patrol, researchers found that white drivers who were stopped were more likely to receive a "discount" on their speeding tickets than were drivers of other races,[10] suggesting a pattern of biased behavior by individual officers.

In recent years, researchers have also begun to examine the role of public safety communications—911 call-takers and dispatchers, and the associated protocols and laws

---

[5] Chohlas-Wood, A., *Goel, S., Shoemaker, A., & Shroff, R. (2018). An Analysis of the Metropolitan Nashville Police Department's Traffic Stop Practices. Stanford Computational Policy Lab. https://policylab.stanford.edu/media/nashville-traffic-stops.pdf.
Dunn, R.A. (2009) Measuring Racial Disparities in Traffic Ticketing Within Large Urban Jurisdictions, *Public Performance & Management Review*, 32:4, 537-561.
Farrell, A., McDevitt, J., Bailey, L., Andresen, C., & Pierce, E. (2004). Massachusetts Racial and Gender Profiling Study Final Report. Northeastern University Institute on Race and Justice. https://repository.library.northeastern.edu/files/neu:344627/fulltext.pdf.
Fliss, M.D., *Baumgartner, F., Delamater, P. et al. Re-prioritizing traffic stops to reduce motor vehicle crash outcomes and racial disparities. *Inj. Epidemiol.* 7, 3 (2020). https://doi.org/10.1186/s40621-019-0227-6.
Goel, S., Rao, J.M., & Shroff, R. (2016) Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy. *The Annals of Applied Statistics* 10 (1), 365-394.
Goncalves, F. & Mello, S. (2020). A Few Bad Apples? Racial Bias in Policing. Available at SSRN: https://ssrn.com/abstract=3627809 or http://dx.doi.org/10.2139/ssrn.3627809.
Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*. 4(7), 736-745. doi: 10.1038/s41562-020-0858-1. Epub 2020 May 4. PMID: 32367028.
[6] Pierson et al. 2020.
[7] Shoub, K., Epp, D. A., Baumgartner, F. R., Christiani, L., & Roach, K. (2020). Race, Place, and Context: The Persistence of Race Effects in Traffic Stop Outcomes in the Face of Situational, Demographic, and Political Controls. *Journal of Race, Ethnicity and Politics*, *5*(3), 481-508.
[8] California Policy Lab, RIPA in the Los Angeles Police Department: Summary Report (October 2020) https://www.capolicylab.org/wp-content/uploads/2020/10/RIPA-in-the-LAPD-Summary-Report.pdf
[9] Analyses in the other jurisdictions did not consider these factors. We recognize that the data used in all of these analyses are themselves inherently biased, as detailed in the discussion of the denominator problem, below. See also Knox, D., Lowe, W., & *Mummolo, J. (2020). Administrative Records Mask Racially Biased Policing. American Political Science Review, 114(3), 619-637. doi:10.1017/S0003055420000039 and Bronner, Laura. (25 Jun 2020). "Why Statistics Don't Capture the Full Extent of the Systemic Bias in Policing." *FiveThirtyEight.*, for more information.
[10] Goncalves & Mello (2020).

governing their work—in how police respond to calls for service.[11] The information that is relayed by a resident to a call-taker, then from a call-taker to a dispatcher and from a dispatcher to a responding officer, can be critical in determining how a police officer responds to a given scene and the outcome of that interaction. For example, if a person is more likely to call 911 on a person of color than a white person exhibiting similar behavior, then racial disparities in stops can exist and perpetuate without any bias on the part of the individual officer who responds to the call. Moreover, if the caller is more likely to describe a person of color as potentially violent, that bias will also affect how the call-taker, dispatcher, and officer approaches the situation. The potential for racially disparate contact and outcomes may be exacerbated by requirements in many jurisdictions to dispatch police if requested—regardless of the merits of the incident.[12]

Legal Research. Legal researchers have focused on assessing the constitutionality of stops, examining the legal basis on which stops are made and racial disparities in that decision-making process. One landmark decision by the Supreme Court is *Terry v. Ohio* (1968), which set the standard for "reasonable suspicion" as a basis for making a stop. Since then, a number of studies and legal proceedings have highlighted the role of race in the decision to make stops at both the individual and neighborhood levels. Challenges to the practice of "stop and frisk" in New York City in *Daniels et al. v. City of New York* (2003) and *Floyd et al. v. City of New York* (2013) found that many stops conducted by the New York Police Department (NYPD) "did not meet the reasonable suspicion standard and that police targeted individuals based on race."[13] In examining the legal basis for making a stop, Abrams (2014) evaluated a sample of the forms completed by police officers in Philadelphia after performing a stop. He found that nearly half of the stops in the city during the period studied lacked a legal basis.[14] Additionally, Abrams notes that low rates of weapon or contraband recovery in stops involving a search raise questions about the efficiency of this approach to public safety.

Meares (2015) observes that, while *Terry v Ohio* set the precedent for stops to consider the inconvenience of an *individual* interaction relative to the public safety benefits of stops, young Black men—the group most often stopped by the police—experience stops as a program, rather than as a one-off incident.[15] Meares calls attention to the disparate impact of the policy implementation, underscoring that when we evaluate whether stops violate the Fourth Amendment prohibition against unreasonable searches and seizure, "statistics—rather than

[11] See, for example, *Gillooly, J. W. (2020). How 911 callers and call-takers impact police encounter s with the public: The case of the Henry Louis Gates Jr. arrest. *Criminology & Public Policy*, *19*(3), 787-804, and Gillooly, J.W. (2021). Lights and Sirens: Variation in 911 Call-Taker Risk Appraisal and Its Effects on Police Officer Perceptions at the Scene. *Forthcoming in the Journal of Policy Analysis and Management.*

[12] See, for example, Hutchinson, Bill. (10 Oct 2018). "From 'BBQ Becky' to 'Golfcart Gail,' list of unnecessary 911 calls made on Blacks continues to grow." ABC News.

[13] Ridgeway, G. (2017). Stop-and-Frisk Is Essential…And Requires Restraint. *Journal of Policy Analysis and Management*, *36*(3), 683-689.

[14] Abrams, D. (2014). Law and Economics of Stop-and-Frisk. Loyola University Chicago Law Review, 46, 369-379. Social Science Research Network Electronic Paper Collection. http://ssrn.com/abstract=2669515

[15] *Meares, T. L. (2015). Programming Errors: Understanding the Constitutionality of Stop-and-Frisk as a Program, Not an Incident. *The University of Chicago Law Review*, 82(159), 159-179.

stories about stops—are necessary" to understand the broader effects.[16] Her conclusions are reinforced by the empirical research to date that finds evidence of widespread racial disparities and bias in the conduct of stops in jurisdictions across the United States.

B. Limitations

The findings summarized above offer evidence of biased behavior in police stops generally (i.e., in multiple jurisdictions and time periods). There are, however, two notable shortcomings as we seek to understand whether stops are conducted in a racially biased manner in a specific jurisdiction—as opposed to the specific jurisdictions and time periods studied—and if so, the particular source of bias. With few exceptions, these analyses are limited in their ability to establish appropriate benchmarks of similar stop scenarios and tell us about the sources of bias, and therefore, what we can do to address the bias. We discuss each of these limitations in greater detail below.

1. Many of the analyses finding racial disparities in stops may be unreliable because they set inappropriate benchmarks for their comparisons.

Measuring racial bias in stops requires understanding what we would expect the racial distribution of stops would look like without racial bias on the part of police. In Neil and Winship's (2019) definition of discrimination, this "benchmark" or "denominator" defines the situations we are treating as similar, so that we can document whether and how much what we observe deviates. The challenge, however, is that multiple factors drive officers' decisions to make stops in a given encounter with civilians and failing to account for these factors can lead to inaccurate conclusions about the existence and extent of racial bias. These factors may include:

- The context for the encounter. For example, what puts the officer in a position to make a (decision about a) stop? Is it a traffic stop, a response to a 911 call, or a warrant stop? If it is a traffic stop, what laws, policies, and norms govern those stops and where officers are assigned to patrol? If it's a response to a 911 call, what prompted the caller to report and what information did they provide?
- The officer. Factors related to the officer themselves, including their behavior, level of suspicion, and sense of threat during an encounter.
- The person (who might be) stopped. Their race, as well as other characteristics like gender, appearance, behaviors, attitude, and affect during an encounter.

Not only do these factors individually drive decisions about stops, but they may also interact with one another. For example, an officer deployed specifically to look for an armed suspect will likely be more primed to make a stop than an officer on routine patrol.

[16] Ibid, p. 179.

Knox, Lowe, and Mummolo (2020) offer a causal framework for thinking about the "denominator problem" when documenting racial bias in stops.[17] They focus on encounters (or sightings of civilians by police officers) which may or may not lead to a stop in a given situation. They define four categories (or "strata") of that civilian's behavior and race that researchers must consider to accurately measure the level of bias in post-stop outcomes, such as the use of force. To paraphrase the paper, in any police-civilian encounter, officers' decisions about whether to make a stop may fall into one of four categories:

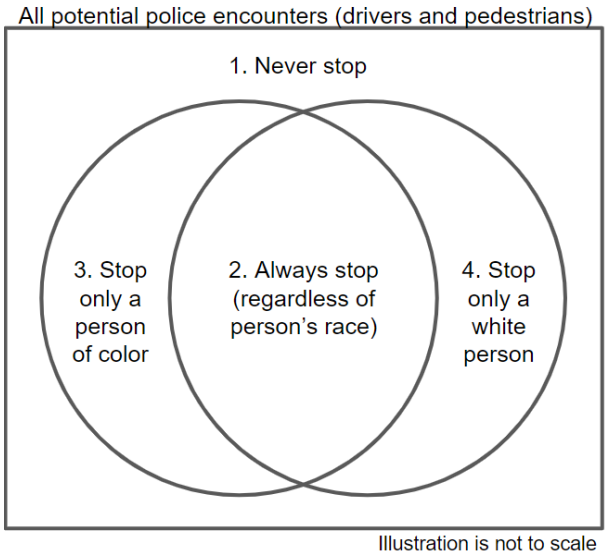
1. Never Stop: For example, a civilian exhibiting behavior that would draw no suspicion or reason to stop regardless of their race (e.g., driving safely at or under the speed limit in a car meeting all legal requirements).

2. Always Stop: For example, a civilian exhibiting a behavior that would always lead to a stop regardless of the person's race (e.g., actively committing robbery or assault).

3. Stop Only if a Person of Color: For example, a civilian exhibiting a behavior that draws suspicion only because the person is (presumed) non-white; the same behavior exhibited by a white person would not be judged as suspicious (e.g., a person of color walking in a predominantly white neighborhood might be inaccurately assumed to be there to buy or sell drugs).

4. Stop Only if White: For example, a civilian exhibiting a behavior that draws suspicion only because they are (presumed) white; the same behavior exhibited by a non-white person would not be judged as suspicious (e.g., a white person walking in a predominantly black neighborhood might be inaccurately assumed to be there to buy or sell drugs).

Theoretical Illustration: The Decision to Make a Stop



Categories 1 and 2 are decisions to (not) stop someone where bias does not play a role. Categories 3 and 4 are purely a result of discrimination—although the authors find little theoretical or empirical evidence that Category 4 (Stop Only if White) is common at all. If there are no discriminatory stops, then all stops fall in Category 2, and neither Categories 3 nor 4 are relevant.

[17] *Knox, D., Lowe, W., & *Mummolo, J. (2020). Administrative Records Mask Racially Biased Policing. *American Political Science Review*, 114(3), 619-637. doi:10.1017/S0003055420000039

While we often have administrative data on everyone who is stopped,[18] it is hard to identify which stops fall into the various categories in the Knox et al. framework, and no administrative data includes information on "Never Stop" encounters. The data often do not allow researchers to disentangle which stops would occur regardless of the person's race (Category 2, unbiased stops) and which would occur for only people of color (Category 3, discriminatory stops). Fully accounting for this selection bias in policing data would require knowing who police choose not to stop (Category 1) or the full accounting of situations in which police could have made a stop but did not. This challenge is particularly acute for pedestrian and traffic stops. Lacking this ideal data, simplistic approaches—like benchmark tests that compare the racial distribution of people stopped to their prevalence in the population—may not only over- or underestimate bias, but they may also even estimate the incorrect direction of bias.[19]

Other analyses, such as hit rate or outcomes tests (described in more detail in Section 4) attempt to work around the denominator problem by focusing on measuring bias in police enforcement actions (e.g., searches, arrests) once a stop has happened. For example, hit rate analyses focus on just those instances in which someone is searched or "frisked" once a stop is made and the rate that contraband—most often weapons, drugs, drug paraphernalia, and stolen property—is seized from those searches. Differences in these rates can suggest that there are different thresholds for decisions to search different racial groups.

While these rates are the foundation for many of the more rigorous techniques (described in Section 4), hit rate analyses do not account for the precipitating factors driving the initial stop decisions, and are applicable only to a subset of stops (i.e., those involving a search). For context, 15% of stops in Washington, DC in 2020 included some form of a pre-arrest search.[20] While hit rate analyses offer firmer conclusions regarding the presence of bias, they are limited in their ability to offer an assessment of bias that encompasses all stops as well as all steps of the process through which a stop can occur.[21] They also do not identify the level of the bias, as discussed in the next section.

2. **Many of the analyses finding racial disparities in stops do not pinpoint the sources of bias, limiting their ability to inform the development and implementation of effective solutions.**

[18] In Washington, DC, local law and policy requires that all stops are recorded, including the race of the individual(s) stopped. However, such data collection is not required in all jurisdictions. Incomplete data on who is stopped (e.g., Categories 2, 3, and 4) further exacerbates the measurement problem.

[19] See Neil and Winship (2019); Knox, D., Lowe, W., & Mummolo, J. (2020); and Ridgeway, G. & MacDonald, J. (2010) "Methods for Assessing Racially Biased Policing," in *Races, Ethnicity, and Policing: New and Essential Readings* S. Ric and M. White (eds). NYC: NYU press. pp 180-204. https://doi.org/10.7249/RP1427. To address this issue, Knox et al. (2020) proposes a technique to sharply bound racial bias in post-stop outcomes (i.e., estimate best- and worst-case levels) given information available in policing data.

[20] Metropolitan Police Department. (2021). Stop Data Report, January - December 2020. Government of the District of Columbia, pp. 15-16

[21] As discussed below, there is promising work, using new sources of administrative data and hit rate tests to understand bias when specific types of stops happen. For example, using 911 call data to understand racial bias in stops in response to calls for service; such analyses, however, cannot explain racial bias in the calls for service made by residents.

While many of these studies offer evidence of disparate outcomes in police stops, they do not tell us the particular *source* of bias. They do not precisely quantify how much of the disproportionate rate of stops by race is because of individual officers' racial bias, as opposed to other structural factors. For all the harm of pervasive individual racial bias, there may be additional causes of racial inequalities beyond explicit and implicit bias of individual officers. As Phil Goff notes, we must also be able to "imagine a world where racism does not require racist actors."[22] Indeed, understanding those broader sources of bias is critical to developing effective policy solutions that target the root causes of the disparities we observe.

We summarize Goff's framework for conceptualizing racial bias in stops in the inset below.

[22] *Goff, P. A. (2013). A measure of justice: What policing racial bias research reveals. In F. C. Harris & R. C. Lieberman (Eds.), Beyond discrimination: Racial inequality in a postracist era (p. 157–185). Russell Sage Foundation, p. 161

Goff's framework for conceptualizing racial bias in stops to better target root causes[23]

Racial bias may occur within or across three levels:

- **Individual Level**: Individual officers treat people of different racial groups differently. This corresponds closely to the definition of discrimination provided by Neil and Winship, above.
- **Department Level**: Police departments deploy officers differently and apply different tactics/ strategies to different neighborhoods (e.g., enforcement of drug laws on college campuses is often more lax than it is in lower-income neighborhoods).
- **City/ Community Level**: Law enforcement missions, resources, and policies vary across jurisdictions, as do the missions, resources, and policies of non-law enforcement actors (e.g., "upstream" resources promoting economic stability, alternative responses to 911 calls). Often, those variations correspond to differences in the geographic segregation of racial groups in the population (e.g., majority white suburban areas are often policed in a manner that is different from policing in more diverse urban settings).

In addition, two factors have implications for bias across these three levels:

- **Broader social norms and expectations** that inform decision-making at all levels (e.g., the mantra that if you "call 911, police will come") have cascading effects on the sheer number of interactions police have with civilians, as well as the type. The long-standing societal acceptance that law enforcement is the appropriate responder for a wide range of social issues, from mental health to school safety to homelessness to domestic disagreements to violent crime, remains entrenched across large swaths of American society.
- The **historical and structural context** that shapes the variation in poverty and crime that we observe in cities and communities today is relevant to how we understand where and when police make stops. For example, while police deployment is often based most immediately on where crime is occurring, the research suggests that a) this tactic does not explain all disparities in stops; and b) the geographic variation in crime rates can be attributed, in large part, to the historical underinvestment and biased practices (e.g., redlining, eminent domain, school segregation) that have fostered criminogenic conditions in these communities.[24]

[23] Klein, Ezra. (15 Sept 2020). "Race, policing, and the universal yearning for safety." Vox.com. https://www.vox.com/2020/9/15/21437156/police-racial-bias-shootings-the-ezra-klein-show.
[24] Ibid.

Consider a simplified example of a police stop under Goff's framework. When the average resident thinks of a police stop, they likely envision an officer stopping a vehicle (or pedestrian) on the street based solely on an officer's observations and judgment, in that moment. The fact that an officer uses their individual judgment invites bias—either implicitly or explicitly—into the officer's decision to conduct a stop or not. On top of that potential for individual-level bias, policies at the police district or department-level determine whether the officer was ever deployed to that street to make the observation and judgment in the first place. Those policing strategies themselves also have potential to introduce bias, as well as to amplify or mitigate any individual-level biases. Moreover, even if all individual officers and the district or department-level strategies are completely unbiased, the historical and structural context described above plays a significant role in the rate of contact between police and communities of color that have long been subject to systemic inequalities and concentrated disinvestment. These community-level factors can lead to disparate outcomes by race even in the absence of discriminatory acts by police officers or systemically racist policies of the police department. They also can amplify biases at the other levels.

Above, we used a relatively simple example—police observing the behaviors of drivers and pedestrians—to illustrate the potential for bias at multiple levels to impact stops. Yet, many stops occur as the result of an even more complex set of factors. Those stops are influenced by factors within the control of individual officers or police departments, as well as those beyond their direct control. That complexity—and the limits of current social science and legal research to account for it—limit our ability to accurately measure bias in both who is stopped and in the benefits and harms of a stop.

Consider, for example, the 2020 data from Washington, DC, on all MPD stops that involved an arrest, search, or other type of police report (i.e., all "non-ticket stops," which excludes most traffic stops), summarized in the table below.[25] MPD reports a range of reasons for police to make a stop. Each stop can serve multiple constitutional purposes in law enforcement, but each type of stop may also be vulnerable to different sources of bias, at different levels. Individual-level bias is more likely to drive biased stops when individual officers have wide discretion in deciding to make a stop, e.g., when people are stopped due to individual actions (7%), officer suspicion of criminal activity (10%), or demeanor during a field contact (3%). At the department-level, bias may enter the process that results in stops through departmental decisions about how many officers to deploy and where, as well as what tactics to use. These decisions can put officers in a position to make more stops in certain locations and fewer stops in others. Community or city-level bias may play a larger role when an officer or the department has limited discretion in making stops. In Washington, DC, 40% of non-ticket stops are in response to calls for service (i.e., 911 calls for police); 9% are to serve required warrants or court orders. Moreover, community or city-level bias in placing 911 calls or in issuing warrants likely plays a

[25] Adapted from Table 8 of Metropolitan Police Department. (2021). Stop Data Report, January - December 2020. Government of the District of Columbia, p. 20.

larger role in those stops than in a typical traffic stop.

MPD Reported Reason for Stopping Person/Vehicle	% Non-ticket Stops Mentioning Each Reason
Call for service	40%
Individual's actions	22%
Traffic violation	13%
Be on the lookout (BOLO)	13%
Suspicion of criminal activity (self-initiated)	10%
Warrant/court order	9%
Information obtained from law enforcement sources	8%
Individual's characteristics	7%
Information obtained from witnesses or informants	6%
Prior knowledge	4%
Demeanor during a field contact	3%
Observed a weapon	1%
Response to crash	1%

Percentages sum to more than 100% because a single stop may mention multiple reasons.

As Goff stated in his keynote remarks at the *Reimagining Stops* Workshop Series, careful data collection and analyses can help distinguish between racial disparities and racial bias and help us understand which level is responsible for the differences we observe. This will, in turn, allow us to hold police accountable for what they can fix, and also identify what additional actors and policy solutions are needed to reduce the bias we observe in policing outcomes.

C. Potential Research Approaches

In light of the fundamental limitations and complications of stops data, as well as important advances in available methods to more accurately measure racial bias in stops, we present a measurement guide to inform future analyses in DC and beyond in Section 4 of this document. In the guide, we categorize available methods as follows:

- *Common, but limited approaches relying on existing administrative data*
    - Simple Benchmark Analysis
    - Simple Outcome or "Hit Rate" Analysis

- *More accurate analytic approaches that rely on existing administrative data*
    - Veil of Darkness
    - Regression-Adjusted Benchmark Tests
    - Regression-Adjusted Outcome Tests
    - Bounded Estimates of Race Effects
- *Prospective analytical approaches that will require additional data collection*
    - Combining Administrative Data with Video or Photographic Data
    - New Data Collection

In the Measurement Guide, we discuss each category in detail, provide an overview of available methods, and note the advantages and limitations of each approach. We also identify the question(s) that each method can and cannot answer, as well as key takeaways for future efforts.

Priority Research Questions on Bias in Stops:

Many jurisdictions want to accurately detect whether racial bias exists in police stops. The table below identifies core research questions, indicates the type of evidence that such research would contribute, and draws on the measurement guide above to note potential data sources and approaches to be applied for each question.

Priority Question	Evidence Types	Potential Data Sources	Potential Research Approaches
Is there evidence of racial bias in stops?	Foundational fact finding	Administrative data, observational data, video footage, photographic records, new data collection	Draw from Table 2 and Table 3 in Section 4 based on the question to be answered and the data and resources available
If racial bias is observed in stops data, at what level(s) is it occurring? What are the sources of disparities observed in the jurisdiction's stops data?	Foundational fact finding	Administrative data, interviews, focus groups, observational data, archival/ historical data, new data collection	Draw from Table 2 and Table 3 in Section 4 based on the question to be answered and the data and resources available. Analyze stop or hit rates for specific police districts or officers by race, being careful to isolate disparities in rates that cannot be explained by other factors (e.g., crime rates or demographic make-up of a specific district/ beat)[26]
What factors, including race, predict that a stop will lead to an enforcement action (search, arrest, etc.), directly or indirectly?	Foundational fact finding	Police department administrative data, court records	Machine learning
What is the role of 911	Foundational	911 operator	Process evaluation, focus groups/ interviews, analysis of

[26] Analyses such as these—that take into account crime rates or geography—are likely to understate the total amount of bias since those factors, themselves, are prone to systematic racial bias (e.g., due to economic, social, and geographic stratification).

communications in increasing or decreasing racial bias in stops that originate from a call for service?	fact finding	administrative data, police department administrative data, 911 operator and police department standard operating procedures, administrative rules/ protocols, local laws	outcomes (search, arrest, etc.) specifically within calls for service. Analyze whether specific neighborhoods are more or less likely to call 911 than would be expected based on their crime rates and other factors.

2. Sample Learning Agenda: Broader Effects of Police Stops on Communities

Stops are often framed as a balance between the public safety benefits of making stops and the individual rights of those being stopped.[27] A detailed accounting of both the benefits and harms of stops can inform decisions about the policy and programmatic changes that can help increase the benefits and decrease the harms. We consider a wide range of individual and community outcomes in this section, from crime reduction to mental health, educational, political, and economic outcomes.

A. Key Findings

Benefits of Stops. Most studies on the benefits of stops focus on changes in criminal activity that occur when there are more (or fewer) police stops. In one experiment comparing patrol areas with no stops to those areas with a typical rate of stops, researchers found that crime increased in the zones where stop activity ceased, and that crime dropped to pre-treatment levels upon resumption of stops.[28] In more recent work leveraging a natural experiment, researchers found that "stops documented as probable cause stops generated crime reductions while stops documented as suspicious behavior had no effect on crime."[29] These findings correspond to research findings on the most effective crime reduction strategies—they work best when focused on known offenders.[30] In particular, these strategies are most effective when focused on criminally active gangs, individual chronic offenders, and drug markets.

In addition, two studies of NYPD stops found some reductions in crime after the implementation of *more* stops. Weisburd et al. (2016) and MacDonald et al. (2016) both found a statistically significant, but small, crime reduction benefits in areas with aggressive stop practices.[31] MacDonald et al. (2016) looked at the implementation of NYPD's Operation Impact, a strategy that designated areas for aggressive enforcement through stops. After implementation, probable-cause related stops were associated with reductions in various types of crime, particularly burglary and robbery. However, Operation Impact increased investigative stops based on suspicious behavior more than it increased probable-cause related stops, and suspicious behavior stops had no measurable impact on crime. Taken together, this study suggests that Operation Impact may have more effectively reduced crime if it emphasized probable cause stops directly related to observable criminal activity. Weisburd et al. (2016)

---

[27] Abrams (2014); Meares (2015).
[28] Ridgeway (2017)
[29] Ridgeway (2017, p. 684); MacDonald, Fagan, & Geller (2016)
[30] Braga, A.A., Turchan, B., Papachristos, A.V. & Hureau, D.M. (2019) Updated Systematic Review: Hot spot policing of small geographic areas effects on crime. *Campbell Systematic Reviews*; 15: e1046. DOI: 10.1002/cl2.1046
[31] Weisburd, D., Wooditch, A., Weisburd, S., & Yang, S. (2016) Do Stop, Question, and Frisk Practices Deter Crime? Evidence at Microunits of Space and Time. *Criminology & Public Policy* 15(1): 31-56. https://doi.org/10.1111/1745-9133.12172
MacDonald J, Fagan J, Geller A (2016) The Effects of Local Police Surges on Crime and Arrests in New York City. *PLoS ONE* 11(6): e0157223. https://doi.org/10.1371/journal.pone.0157223

looked at smaller, "microgeographic" areas in New York City, and found some support that stop, question, and frisks deterred crime. However, researchers noted that it is not clear whether other policing tactics would deter crime to a similar (or even greater) extent. Further, neither of these studies examined the impact of these practices on the community beyond measures of crime, potentially failing to document adverse impacts on individuals and the community as a whole.

Harms of stops. A significant body of research documents the harms of stops across multiple dimensions, at both the individual and community levels.

For example, Dunn (2009) points out that the fines and fees associated with traffic stops can have a significant financial impact, especially on lower-income individuals and communities, as failure to pay an initial ticket can lead to increased fees, increase in auto insurance rates, and/or loss of a driver's license, which in turn presents a whole host of new issues.[32] The accumulation of citations can "metastasize from simple fines to warrants, from warrants to arrests, and further to more severe penalties."[33] Fines from police interactions can affect a person's ability to obtain employment and housing, and contribute to intergenerational poverty.[34] Discrimination in who is fined and for how much can further exacerbate the unjust treatment experienced by communities of color.[35]

Researchers have also documented harms to youth mental health and academic performance arising from interaction with the police. Research on young men who interacted with police in New York suggests stops are associated with more trauma and poorer grades in school.[36] Geller et al. (2014) examined trauma and anxiety symptoms of 18–26-year-old men, finding that their symptoms were associated with the number of stops these young men encountered, the intrusiveness of the stops, and their perceptions of fairness during the stop.[37]

Recent studies of the relationship between one's perceptions of police and one's engagement with the state underscore the broader stakes of police interactions for police legitimacy. Prowse et al. (2020) found that residents in heavily policed areas across the nation "characterize police as contradictory—everywhere when surveilling people's everyday activity and nowhere if called upon to respond to serious harm."[38] This perception that police have a "distorted

---

[32] Dunn, R. A. (2009). Measuring racial disparities in traffic ticketing within large urban jurisdictions. *Public Performance & Management Review*, *32*(4), 537-561.
[33] Fagan, J., & Ash, E. (2017). New policing, new segregation: from Ferguson to New York. *Geo. LJ Online*, *106*, 33, p.36.
[34] Fagan and Ash (2017); Dunn (2009)
[35] Goncalves & Mello (2020); Civil Rights Division.(2015). Investigation of the Ferguson Police Department. U.S. Department of Justice.
[36] Legewie, J., & Fagan, J. (2019). Aggressive policing and the educational performance of minority youth. *American Sociological Review*, *84*(2), 220-247.
[37] Geller, A., Fagan, J., Tyler, T., & Link, B. G. (2014). Aggressive policing and the mental health of young urban men. *American journal of public health*, *104*(12), 2321-2327.
[38] Prowse, G., *Weaver, V. M., & Meares, T. L. (2020). The state from below: Distorted responsiveness in policed communities. *Urban Affairs Review*, *56*(5), 1423-1471, p. 1421.

responsiveness" was persistent both among those who had negative views of police as well as those who felt police "do the best they can."[39] Further, two nationally representative studies suggest that people who are stopped by police in certain contexts (e.g., traffic stops or contact generally perceived to be unjust) are less likely to contact police for help or to report victimization.[40] Slocum (2018) concludes that race, ethnicity, poverty and recent experiences with police all influence the likelihood of reporting victimization in intricate ways.[41]

Negative interactions with the police not only cause fractures in police-citizen relationships, but also jeopardize perceptions of government legitimacy more broadly, as well as subsequent engagements with the state. In other words, negative interactions with individual police officers can lead to legal cynicism, a "cultural orientation…in which the law and the agents of its enforcement, such as the police and courts, are viewed as illegitimate, unresponsive, and ill-equipped to ensure public safety."[42] Geller and Fagan (2019) found that adolescents who had previous experience with police stops (i.e. being stopped, witnessing a stop, or knowing someone who was stopped) reported greater levels of legal cynicism.[43] The study suggests that even vicarious contact—witnessing or personally knowing someone who has been stopped by the police—contributes to this sense of legal cynicism. This means that for youth in "highly-policed" neighborhoods, the chances of such cynicism may be high.

The consequences of deeply entrenched legal cynicism can be significant, ranging from declining to call police for help, to refusal to cooperate with police to solve crimes, to withdrawal from civic life more generally, including activities such as voting and jury duty. Weaver and Lerman (2010) found that contact with the criminal justice system—including police stops—caused people to engage less in certain political activities, even after considering differences in socioeconomic status and criminality.[44] Those who had more contact with the criminal justice system were less likely to turn out to vote, become involved in a civic group or trust the government. More severe encounters with the criminal justice system led to a larger decline in political participation and trust in the government. Weaver et al. (2020) find that residents in highly-policed communities demonstrated an outlook of "collective autonomy," concluding that

[39] Ibid, p. 1435.
[40] Gibson et al. (2010) looked specifically at traffic stops and found that people who experienced at least one traffic stop were less likely to contact police for assistance or to contact police about a neighborhood problem. See Gibson, C. L., Walker, S., Jennings, W. G., & Mitchell Miller, J. (2010). The impact of traffic stops on calling the police for help. *Criminal justice policy review*, *21*(2), 139-159. Slocum (2018) found that when residents had contact with police they viewed as unjust, this interaction had a negative effect on a person's likelihood of reporting victimization. This finding was especially pronounced among African Americans and lower-income populations.
[41] Slocum, L. A. (2018). The effect of prior police contact on victimization reporting: Results from the police–public contact and national crime victimization surveys. *Journal of Quantitative Criminology*, *34*(2), 535-589.
[42] Kirk, D. S., & Papachristos, A. V. (2011). Cultural mechanisms and the persistence of neighborhood violence. *American journal of sociology*, 116(4), 1190-1233.
[43] Geller, A., & Fagan, J. (2019). Police contact and the legal socialization of urban teens. RSF: The Russell Sage Foundation Journal of the Social Sciences, 5(1), 26-49.
[44] Weaver, V. M., & Lerman, A. E. (2010). Political consequences of the carceral state. *American Political Science Review*, *104*(4), 817-833.

power is best achieved by "distancing from state institutions in the short term while building community power in the long term."[45]

B. Limitations

The research findings to date make the need for further research clear (including greater diversity of research methods) as well as a more broad-ranging consideration of outcomes of interest, both across and within studies. Building on earlier work on the relationship between police stops and crime, researchers may consider examining a wider range of police strategies and conducting multi-site studies to determine how generalizable their findings are across different agencies and jurisdictions. Future research should also specifically address a core concern raised by police regarding the number of guns recovered through police stops. In DC, though a small percentage of stops lead to gun recoveries, the majority of guns recovered (69%) come from stops.[46] Nevertheless, gun recovery is not the expected or intended outcome of all (or most) stops. Strategies and evidence that engage with this reality can address a significant gap in the field—i.e., for stops that are intended or reasonably likely to recover a gun, pursue innovations to increase the hit rate and decrease the overall number of contacts, while also identifying alternative methods beyond stops to more effectively remove guns in criminal possession. [47]

More broadly, evaluations of the benefits and harms of stops will need to consider a variety of factors, including the different reasons that a stop may be conducted. For example, while the primary purpose of some stops may be related to roadway safety (e.g., pulling someone over for speeding in a school zone or for running a red light), others may be conducted in support of an investigation or due to suspicion of or observation of criminal activity. What determines whether different types of stops are beneficial or not? Consider, for example, what result constitutes a "successful" traffic stop. Is it a warning? A ticket? A search? An arrest? A reduction in citywide crash or fatality rates? How do we think about these outcomes relative to potential harms of these interactions, as documented in the literature summarized above? Accurate measurement of the broader effects of stops will depend on a variety of factors, from the broader context in which the stop occurs, to the behavior of the officer and the person stopped both before and during the encounter, to the downstream impacts of a stop at the individual, household, and community levels. Many of these factors may be difficult to capture systematically due to the limitations of administrative data collection, and yet, precise assessment can be of great value to decisionmakers responsible for changing policy and practice.

Soss and Weaver (2017) also make a compelling case for considering police-citizen contact as a form of political engagement, extending the spectrum of outcomes that further research in this

[45] Weaver, V., Prowse, G., & Piston, S. (2020). Withdrawing and drawing in: Political discourse in policed communities. *Journal of Race, Ethnicity, and Politics*, *5*(3), 604-647.
[46] Metropolitan Police Department. (2021), p. 5.
[47] Goel, et al (2016).

space should consider.[48] Findings across the research cited above also emphasize the importance of applying both quantitative and qualitative methods, as well as looking beyond law enforcement administrative data to measure the efficacy of a particular policing practice. In addition to looking at the effects of stops on crime, research should examine impacts on other outcomes to better understand potential adverse effects in combination with the crime reduction benefits. Another area for further research concerns understanding police and community expectations of policing—what determines if a stop is "worth it?" If it is appropriate and/or necessary? How do police and residents align or vary in their assessments, and how might this inform changes to policy and practice? Ultimately, understanding the implications of police contact for one's relationship with police and the state more broadly—for one's sense of citizenship and social inclusion—is essential to facilitating productive change.[49]

C. Potential Research Approaches

As noted above, a diverse set of research approaches is necessary to fully understand the broader effects of stops, including both quantitative and qualitative approaches. Examining effects of stops across a broader time horizon, as well as at multiple levels of analysis—individual, household, and community/neighborhood—better reflects the ways in which stops are experienced and can yield a more complete picture of their effects. New technologies—such as body-worn cameras—also present additional opportunities for further exploration.

[48] Soss, J., & Weaver, V. (2017). Police are our government: Politics, political science, and the policing of race–class subjugated communities. *Annual Review of Political Science*, *20*, 565-591.

[49] Monica Bell's theory of legal estrangement, which captures both legal cynicism and "the objective structural conditions (including officer behavior and the substantive criminal law) that give birth" to it" offers a helpful theory to guide and orient reform efforts. See Bell, M. C. (2017). Police reform and the dismantling of legal estrangement. *The Yale Law Journal*, 2054-2150.

Priority Research Questions on Broader Effects of Police Stops

In addition to addressing discrimination in stops, many jurisdictions are interested in increasing the benefits of police stops while reducing the harms. The table below highlights priority research questions for study to inform these efforts, indicates the type of evidence that such research would contribute, and notes potential data sources and methods to be applied for each question.

Priority Question	Evidence Types	Potential Data Sources	Potential Research Approaches
What are the *harms* of police stops in a jurisdiction? What are the effects at the individual level? Household level? Neighborhood level?	Foundational fact finding	Police department administrative records, 2020 Census data, American Community Survey data, community surveys, interviews, observations	Statistical analysis of observed costs (fines, fees, jail time, lost wages, police complaints), qualitative research on which stops police and community members find harmful.
What are the *benefits* of police stops in a jurisdiction? What are the effects at the individual level? Household level? Neighborhood level?	Foundational fact finding	Police department administrative records, 2020 Census data, American Community Survey data, community surveys, interviews, observations	Statistical analysis of observed benefits (traffic incidents, crime or victimization, perceptions of safety), qualitative research on which stops police and community members find beneficial.
Are certain types of stops more beneficial or more harmful than others? What stops are most associated	Foundational fact finding and program evaluation	Police department administrative records, National Surveys,[50] community surveys[51]	Statistical analysis of observed benefits and harms of specific types of stops; qualitative research on which stops police and community members find beneficial/ harmful; pilot and evaluate different approaches to stops (e.g.,

[50] See, for example, the Bureau of Justice Statistics Police-Public Contact Survey (PPCS)
[51] See, for example, the Data Foundation's Policing in America Survey.

with improvements to public safety? What stops are less efficient in delivering public safety benefits?			deflection and diversion programs).
What incentives do officers face related to recovering illegal guns? Are certain types of stops more effective at recovering illegal guns? What other methods beyond stops can be tried to recover or reduce the prevalence of illegal guns?	Foundational fact finding and program evaluation	Police department administrative records, surveys, police officer interviews and observations	Qualitative and statistical analysis of observed gun recoveries. Experimental and quasi-experimental evaluations of other methods to recover or reduce the prevalence of illegal guns?
How do police and resident perceptions of the necessity and appropriateness of a stop compare?	Foundational fact finding	Body Worn Camera footage, surveys, community interviews and observations	Survey analysis, qualitative research
What factors determine how the government responds to public safety issues in demographically different neighborhoods (e.g., deploying more officers or patrols, installing automated traffic enforcement cameras, making non-law enforcement investments in the community)? Does the response to similar concerns (e.g., unsafe driving, proportional increase in violent crime) differ in	Foundational fact finding	Police department administrative records, Body Worn Camera footage, surveys, police and community interviews and observations	Statistical analysis can determine when demographically different neighborhoods experience proportionally similar changes in crime. A qualitative comparison of the policing and non-policing response can help identify similarities and differences.

majority white vs. majority non-white neighborhoods?			

3. Sample Learning Agenda: Opportunities for Change

Policymakers, community advocates, and researchers have identified a number of avenues for changes in policy and practice to both reduce racial bias in stops and decrease the harms of stops while still preserving public safety benefits.[52] We highlight several of these proposals in this section, and also make note of those practices that have been less successful. We also include policy changes that have been introduced recently and have not yet been evaluated.

A. Key Findings

Changes to Supervisor Oversight. Several studies suggest that frontline supervisors—those officials who directly oversee patrol officers—can significantly influence the behavior of officers on the street. Field research conducted in the Indianapolis, IN and St. Petersburg, FL police departments found that the "style or quality of field supervision can significantly influence patrol officer behavior…active supervisors appear to be crucial to the implementation of organizational goals."[53] A more recent study of a supervisory change in the NYPD appears to affirm this finding. In 2013, NYPD issued a memo requiring that patrol officers write narrative memos of why they made each stop, and then submit these descriptions to their supervisors. Mummolo (2018) found that this change decreased the number of times police stopped pedestrians and dramatically increased the rate at which police seized weapons from the pedestrians that they did stop. Based on interviews with officers, Mummolo hypothesizes that the memo made officers feel like they were under heightened scrutiny, or that supervisors were "really watching [them] now" (Mummolo 2018, p. 8). However, there was no evidence that racial disparities in stops that existed before the intervention were eliminated. Additionally, while patrol made dramatically fewer stops overall after the memos were instituted, patrol officers also made fewer stops that yielded weapons. Mummolo found that this decrease in fewer weapon-yielding stops did not lead to any detectable increase in homicides or robberies.

Changes to stop policies. In 2012, Fayetteville, North Carolina, mandated that police officers obtain written permission from motorists before conducting searches absent any probable cause. Using traffic stops data, Epp and Erhardt (2021) find that the mandate led to a significant reduction in the use of consent searches. Though officers made fewer overall searches, "contraband continued to be recovered at pre-reform levels, indicating a reduction in low-quality searches with minimal substantive impact." Notably, they also find that "homicide rates are statistically indistinguishable between the pre- and post-reform periods."[54] Another set of researchers examining NYPD's stops observed the low rate of weapon recovery in stops, and found that by focusing officers instead on only the small percentage of stops that are

[52] The Thurgood Marshall's Center's policy recommendations focus on community perspectives and harm reduction and are available at: https://thurgoodmarshallcenter.howard.edu/Reimagining-traffic-stops.
[53] Engel, R. S. (2003). How police supervisory styles influence patrol officer behavior. *Critical issues in policing: Contemporary readings, p. 219.*
[54] Epp, D. A., & Erhardt, M. (2021). The use and effectiveness of investigative police stops. *Politics, Groups, and Identities*, *9*(5), 1016-1029, p. 1016.

"statistically most likely to result in weapons seizure, one can both recover the majority of weapons and mitigate racial disparities in who is stopped."[55] Whether this finding holds, and whether it is generalizable when fully implemented in a real-world context, remains to be seen.

Targeting stops through an intensive approach. Though focused deterrence—which involves targeting both enforcement and social supports to high-risk people in high-risk places—is a strategy designed primarily to reduce violence, stops are often a key component of how focused deterrence is implemented. How the concept of "focus" is operationalized is important, as a geographically-based focus can leave entire neighborhoods or communities feeling overpoliced or unfairly targeted,[56] whereas a person-based focus can address this risk *if* implemented in a fair and informed manner. A 2019 meta-analysis of studies of focused deterrence found such strategies were associated with moderate reductions in crime.[57] In particular, these strategies were most effective when focused on criminally active gangs, individual chronic offenders, and drug markets. This sort of intensive, more narrowly focused approach should be considered in light of the harms that an increased police presence can cause in a community. Relatedly, a small study gave residents a survey before and after a police department targeted a nearby block with a combined strategy of focused deterrence and police-community partnerships that emphasized racial reconciliation.[58] The survey found Black residents in particular perceived less non-violent crimes and less racial profiling after the intervention. However, researchers did not find any change in other measures of police legitimacy, such as perceptions of police effectiveness and fairness.

Automated traffic enforcement. Traffic safety is a core goal of traffic enforcement. As reported by the National Conference of State Legislatures, "more than two people in the U.S. are killed each day by drivers running red lights, with 28% of crash deaths occurring at signalized intersections as a result of drivers running red lights…the National Highway Traffic Safety Administration (NHTSA) reported nearly 9,400 fatalities from speed-related crashes in 2018."[59] Early research on automated enforcement through red-light and speeding cameras indicates that this technology improves safety, reducing speeding as well as the number of crashes on roadways with camera devices, though more rigorous tests of cameras are needed to confirm these findings.[60] Moreover, as some jurisdictions consider increasing the use of automated traffic

[55] Goel, S., Rao, J. M., & Shroff, R. (2016). Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy. *The Annals of Applied Statistics*, *10*(1), 365-394, p. 365.
[56] ibid.
[57] Braga et al., (2019)
[58] Saunders, J. & Kilmer, B. (2021) Changing the Narrative: Police-Community Partnerships and Racial Reconciliation. *Justice Quarterly* 38(1): 47-71. https://doi.org/10.1080/07418825.2019.1568520.
[59] Bloch, S. Shinkle, D., & Bates, J. (2 July 2021). Traffic Safety Trends | State Legislative Action 2020. National Conference of State Legislatures, citing the AAA Foundation for Traffic Safety and the National Highway Traffic Safety Administration.
[60] Hu, W. & Cicchino, J. B. (2017). "Effects of turning on and off red light cameras on fatal crashes in large U.S. cities." Insurance Institute for Highway Safety (IIHS); Hu, W. & McCartt, A.T. (2016). "Effects of automated speed enforcement in Montgomery County, Maryland, on vehicle speeds, public opinion, and crashes." Insurance Institute for Highway Safety (IIHS); Retting, R.A., Farmer, C.M. & McCartt, A.T. (2008). "Evaluation of automated speed enforcement in Montgomery County, Maryland." Insurance Institute for Highway Safety (IIHS).

enforcement as a means of decreasing police contact, the broader effects of this tradeoff remain unknown. Concerns about where devices are placed and whether there may be a potentially undue focus on revenue generation, rather than traffic safety, continue to be core issues. In DC, more than half of MPD's stops (55%) concluded with the issuing of one or more traffic tickets to a driver, bicyclist, or pedestrian and no further law enforcement action, suggesting a large volume of stops that could potentially be handled through automated enforcement.[61]

Implicit Bias/ Diversity/ Cultural Sensitivity Training. Numerous police agencies (and government and non-government organizations more broadly) have delivered some form of implicit bias/ diversity/ cultural sensitivity training to their personnel over the past few years with the goal of reducing racial disparities. While these trainings have not been standardized, those versions evaluated to date have not had the hoped-for effects.[62]

B. Limitations

Our review of the literature emphasizes the opportunity for innovation and creativity in reimagining police stops. Further investments in foundational research and evaluations of new policy proposals are necessary to advance our knowledge, and thus, more effectively reimagine stops. For example, a better understanding of how organizational factors within a police department can influence stop practices (e.g., through changes to supervision) can directly inform department-level policy change. Similarly, further research on the role of 911 operators and information flow can highlight additional opportunities to reduce racial disparities and other harms while strengthening public safety.[63] As jurisdictions around the country grapple with these issues and pursue reforms, these efforts should be rigorously evaluated so that those jurisdictions and the broader community have an opportunity to learn from their experiences. For example:

- The Virginia state legislature passed a bill to reduce the number of "pretexts" officers can use to conduct stops (e.g., banning vehicle searches based solely on an officer smelling marijuana, converting primary vehicle offenses to secondary offenses).[64]
- Berkeley, California authorized the shift of traffic enforcement duty from police to a civilian office.
- Minneapolis enacted a prohibition on traffic stops for minor offenses (e.g., air freshener on inside mirror).
- Law enforcement leaders in Ramsey County, Minnesota directed police officers to minimize non-public safety-related stops, and the local prosecutor indicated his office

[61] Metropolitan Police Department (2021), p. 17.
[62] Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice reduction: Progress and challenges. *Annual review of psychology*, *72*, 533-560.
[63] In DC, a call for service was at least one of the reasons listed for making 40% of non-ticket stops and 6% of ticket stops citywide. See Metropolitan Police Department. (2021).
[64] Justice Forward Virginia, "Success Story: Many Policing 'Pretexts' Eliminated in Virginia." Retrieved on December 3, 2021, https://justiceforwardva.com/pretextual-policing

would "no longer prosecute cases based solely on a non-public-safety traffic stop or those that are the result of a vehicle search based solely on consent, without any other articulable suspicion."[65]

- Philadelphia, Pennsylvania recently passed legislation to eliminate police authority to make stops for minor traffic violations that are often used as a pretext to conduct a search.[66]

Though it is too soon to tell how these policies will affect racial disparities, other effects (positive or negative) of stops, or public safety more broadly, rigorous evaluation of these efforts, as well as of any other new programs, can help guide decision-makers to more optimal policies.

C. Potential Research Approaches

Randomized evaluations offer the most rigorous assessment of a new policy or program, allowing us to understand the *causal* effects of the intervention, and thus they are recommended for assessing the effects of new programs and policies that are being implemented. Wherever possible, we recommend starting a new approach as a pilot that is paired with a randomized evaluation to isolate the effect of the new approach from the other factors influencing crime rates, stops, and the outcome of stops overtime. Many quasi-experimental methods will often struggle to account for these factors within a jurisdiction—consider, for example, trying to evaluate the impact of a city-level program that started in January 2020 and then having to account for the effect of the pandemic on crime and policing, much less the additional reforms and effects following the murder of George Floyd. These methods should be thought of as a secondary option when randomization is not feasible. Pairing randomized evaluations with surveys as well as qualitative research methods can illuminate the mechanisms of change and help better understand why a program worked (or did not work). Process evaluations documenting how a program is designed and implemented can also be essential to understanding how to replicate and scale a program, both within a city (e.g., from a pilot program to full implementation) and across jurisdictions. Finally, cost-benefit analyses—that take a broader set of outcomes discussed in Section 2 and a longer time horizon to see those outcomes bear out—can offer a useful comparison as decision makers contemplate public safety budgeting.

[65] News release, "Ramsey County Attorney, police leaders announce plans to reduce non-public-safety traffic stops," September 8, 2021.
[66] The Editorial Board, "Curtailing improper traffic stops is a path to more effective and equitable policing," *The Philadelphia Inquirer*, September 3, 2021.

Priority Research Questions on Opportunities for Change

Jurisdictions can and should explore opportunities for change that fit their individual challenges, resources, and local context. The table below provides a sample of research questions based on existing research and ideas being considered nationally. The table indicates the type of evidence that such research would contribute and notes potential data sources and methods to be applied for each question.

Priority Question	Evidence Types	Potential Data Sources	Potential Research Approaches
Does an increase in supervisor oversight decrease the number of stops made? If so, which types of stops decrease? What are the effects on crime?	Policy analysis and program evaluation	Police department administrative records	Experimental and quasi-experimental evaluations
Does a focused deterrence approach reduce the number of stops made? Does it improve public safety outcomes while minimizing harm?	Program evaluation	Police department administrative records, interviews, court data	Experimental and quasi-experimental evaluations
What stops (and at what locations) are most effective in producing public safety benefits (e.g., reducing traffic crashes and fatalities)? Is it more effective to target stops to certain populations (e.g., individuals with repeat justice system involvement)?	Foundational fact finding/ policy analysis	Administrative data	Statistical analysis
Can automated traffic enforcement (ATE) reduce racial	Program evaluation	Police department and department of	Experimental and quasi-experimental evaluations, qualitative research

disparities? What are other consequences of expanding ATE?		transportation administrative records, observations	
What calls for service result in an enforcement action (seizure of contraband, arrest, etc.)? What calls do not?	Foundational fact finding/ policy analysis	Administrative data	Statistical analysis
How can 911 protocols be changed to minimize harm and optimize public safety (e.g., non-police responses to behavioral health calls or minor traffic crashes)? What policy changes are most effective?	Foundational fact finding	Administrative data, observation, surveys	Statistical analysis, observational research, survey research, experimental and quasi-experimental evaluations
What is the effect of non-law enforcement responses to specific 911 calls, quality of life violations, and traffic violations?	Program evaluation	Administrative data	Experimental and quasi-experimental evaluations Note: The research to date suggests that implementing place-based strategies that are intended to improve community relations/ engagement can lead to an increase in calls for service and reports of crime, potentially due to an increase in trust and reporting. As a result, program impacts need to be put in the appropriate context and leverage emerging measurement techniques.[67]
What is the effect of non-law enforcement strategies on crime reduction and the volume/ type of contacts residents have with	Program evaluation	Administrative data	Experimental and quasi-experimental evaluations

[67] See, for example, Weisburd, D. and C. Gill. (2020). "Rethinking the Conclusion that Community Policing Does Not Reduce Crime: Experimental Evidence of Crime Reporting Inflation." *Translational Criminology.*

police?			
What are the benefits and harms of policy changes to how legal standards of consent for search, reasonable suspicion, and probable cause are documented? (e.g., request written permission for a consent search, as in Fayetteville, NC; greater oversight on reasons given for reasonable suspicion, as in NYC)	Program evaluation	Administrative data	Experimental and quasi-experimental evaluation; simulations in training, observational research

4. Measurement Guide: Accurately Measuring Racial Bias in Stops

Section 1 provides an overview of existing research on racial bias in stops as well as the limitations of many commonly used approaches. In this section, we present a guide to the different approaches for documenting racial bias in stops, including what each method can and cannot tell us in light of the issues of identifying the appropriate denominator and level of bias described in previous sections. While commonly used approaches to measuring racial bias in stops are prone to error or a lack of precision for the reasons described above, advances in statistical methods and available technology offer paths forward to addressing this critical issue and informing decisionmakers. We categorize the methods as follows, and discuss each category in detail below:

- Common, but limited approaches relying on existing administrative data;
- More accurate analytic approaches that rely on existing administrative data; and
- Prospective analytical approaches that require additional data collection.

We conclude with a discussion of recommended approaches and key takeaways for measuring racial bias in policing. We view the use and further application of these methods as foundational to reducing disparities in police stops—without careful identification of the source(s) of bias, jurisdictions will lack necessary information to identify the appropriate solutions.

A. Common, but limited approaches that rely on existing administrative data [Not Recommended]

Simple benchmark and outcome tests are relatively easy to perform and require only administrative data, making them popular among those seeking to measure racial bias in stops. These analyses constitute an important first step to assessing the impact of police stops on a community, however, they often yield highly imprecise, or even inaccurate, results. As a result, they are insufficient in diagnosing the cause of the disparities in police stops and provide little insight as to what solutions will best fit the needs of the community. More pointedly, if the root cause of bias is structural and socio-economic inequality, then changes that focus on individual-level police officers or even department-level policy decisions would likely have limited effect and may even be counterproductive). Generally, to estimate racial bias in stops, researchers caution against using these simplistic, or unconditional, benchmark analyses; this is particularly true for those using population distributions (e.g. the racial distribution of a jurisdiction as a whole) as the benchmark.[68] At minimum, researchers using benchmark tests should employ some form of matching or a weighting approach to define groups of similarly situated individuals for comparison (see Table 2 for specific methods).

The outcomes or hit-rate test offers a slight improvement over the simple benchmark test, though many of the same limitations remain. This test measures, for similar situations (e.g., a stop with a search), the likelihood that a particular outcome (e.g., recovery of contraband) occurs.

[68] Neil, R. & Winship, C. (2019); Ridgeway, G. & MacDonald, J. (2010), Knox, D., Lowe, W., & Mummolo, J. (2020)

However, this test cannot pinpoint the source of the observed disparities, limiting our ability to make informed policy or programmatic decisions based on the results of the analysis.

Bottom Line: The tests detailed in Table 1 below can only demonstrate that racial disparities exist in a jurisdiction. While these tests can help identify a problem at the 30,000-foot level, further analyses are necessary for accurate measurement of racial bias in police stops (or policing more broadly).

Table 1: Common, but Limited Methods

Method	Simple Benchmark Analysis	Simple Outcome or "Hit Rate" Analysis
What is it?	• Compare the frequency with which members of different racial groups are stopped (and/or searched, arrested, etc.) to some race-specific risk set, the "benchmark" or "denominator." • This is the most commonly applied test. • Example: "Black drivers account for X% of stops even though they represent only Y% of the population in city Z."	• Conditional on a search being made, what is the rate of arrest or discovery (or "hit rate") of contraband? When the hit rate is lower for one racial group, this is interpreted as evidence of a lower threshold for initiating a search, which in turn suggests discriminatory behavior. • Example: "During a search, police are X times less likely to find contraband with Black drivers than when they search white drivers."
Questions this method can answer	• *Are people of a certain race stopped more or less than their prevalence in the population of a specific area or jurisdiction would suggest?*	• *Is contraband recovered at different rates by the race of the individual searched?*
Questions this method cannot answer	• *Is the observed disparity due to bias? Is it due to differences in "stop-able" behavior?* • *What is the source of the disparity? What is the level at which bias is operating?* • *What policies might be targeted to reduce disparities?*	• *What is driving this disparity? Are there differences in who is searched and what drives those differences?*
Key Takeaways	• The administrative data we rely on for these tests are generated through a complex, multi-step process—the data do not capture all the relevant steps and decisions. **This means we might be severely under- or overestimating bias in our analyses.** • Proper application of this test requires 1) selecting the appropriate denominator (e.g., using population benchmarks vs. the rate at which police encounter civilians of different backgrounds vs. the rate of illegal behavior) and 2) correctly estimating that denominator—which can be extremely challenging without detailed stop data and more advanced techniques	• This test relies only on observations of outcomes, not on the threshold for searching a person. The hit rate depends on the distribution of signals; and because those signals that lead to a search (e.g., suspicious behavior) are rarely evenly distributed in a population, especially at the margins, **it is possible to obtain results using this method that suggest bias where there is none, or vice versa.**
Sources	Ridgeway and MacDonald (2010); Neil and Winship (2019); Knox,	Neil and Winship (2019)

	Lowe, & Mummolo (2020).	

B. More accurate analytic approaches that rely on existing administrative data

This set of methods also relies on administrative data but incorporates additional statistical techniques to increase the precision and accuracy of estimates of racial bias. These methods also offer insights on whether individual police officers engage in biased behavior. As with the methods summarized in Table 1, however, the reliance on administrative data continues to present a significant limitation to these analyses. As noted earlier, administrative data—on stops, arrests, crime, etc.—are generated through a complex, multi-step process (e.g., reason for stop, where officers are patrolling, officer training, officer decision-making etc.). The data do not capture significant parts of that process, thus posing some limitations in how we interpret and use these analyses. As a result, these methods are best suited to understand bias in post-stop outcomes, as their reliance on administrative data limits their ability to capture bias that occurs earlier in the process (e.g., in the decision to deploy an officer to a particular neighborhood, to make that particular stop, etc.). The bias these tests can capture will most often be bias at the individual officer level, but they also may capture differences in how non-white communities are policed (e.g., a department-level emphasis—either explicit or implicit—on conducting stops in non-white communities would lead to disparate outcomes of stops, even if individual officers were unbiased).

Importantly, these analyses can also shed light on whether disparities are driven by racial bias at the City/Community level in Goff's framework. Where these analyses show no bias—or substantially less bias than a simple benchmark or outcome test indicates—the absence of bias demonstrates that the source of disparities in police stops is being driven by societal problems beyond policing. In these scenarios, reforms that focus on police tactics or specific officers will do little to reduce racial disparities in stops and may prove counterproductive.

Bottom Line: The tests detailed in Table 2 represent a substantial and vital improvement over the more common methods summarized in Table 1, and can deliver more accurate, precise estimates of racial bias.

Table 2: More Accurate Methods Using Administrative Data

Method	Veil of Darkness	Regression-Adjusted Benchmark Tests	Regression-Adjusted Outcome or Regression-Adjusted Hit-rate Analysis	Bounded Estimates of Race Effects
What is it?	• Takes advantage of changes in the amount of daylight to identify evidence of racial profiling in traffic stops. • Underlying hypothesis is that it is more difficult to determine the race of a driver after dark than during the day. • If a smaller proportion of Black drivers are stopped at night vs. during the day, this would be evidence suggestive of biased discriminatory behavior.	• Uses a hierarchical multilevel regression to statistically adjust stop rates for the rates at which individuals of each racial group are arrested in a police precinct or neighborhood. • Attempts to uncover whether individuals of a specific race are being stopped at disproportionately higher rates than they are being arrested. • Attempts to account for the factors that contribute to the number of justified stops in a specific geographic area, like underlying rates of criminal activity or 911 calls for people of different races. • Example: "We find that persons of African and Hispanic descent were	• Builds on Gelman et al. (2007). • Uses a logistic regression model to estimate stop-level hit rates (vs. aggregate race-level hit rates) within a specific suspected crime (e.g., suspicion of criminal possession of a weapon) to reduce other factors that may contribute to a stop and are correlated with race. • The model estimates the probability of a hit (e.g., finding a weapon, drugs etc. on a stopped suspect), and includes variables that reflect what the officer would have observed before the stop was made (e.g., suspect demographics, location of stop, date and time of stop, reason given for stop, if stop was result of	• Improves on simple outcome or hit-rate tests by explicitly accounting for bias in the decision to make the stop (correcting for the problem that "... when there is any racial discrimination in detainment, selection on stops introduces unavoidable statistical bias"[70]). Failure to account for that bias will result in underestimating biased stop outcomes, such as use of force, seizure of contraband, or arrest. • Method involves adjusting estimates of outcomes to account for potential racial discrimination in who is stopped, resulting in a more conclusive range ("bounds") of outcomes.

[70] Knox et al. (2020). We note that "statistical bias" is a technical term referring to "anything that leads to a systematic difference in between the true parameters of a population and the statistics used to estimate those parameters. In other words, [statistical] bias refers to a flaw in the experiment design or data collection process, which generates results that don't accurately represent the population." See here for more information. This is distinct from the general use of "bias" (e.g., unfairly favoring one group over others), throughout this learning agenda.

		stopped more frequently than whites, even after controlling for precinct variability and race-specific estimates of crime participation."[69]	call for service). The model also adjusts for crime and hit rates in different geographic areas to account for the likelihood that police tactics change in high crime areas. • A high proportion of stops that have a low probability of a hit (e.g., less than 1% probability) are used as an indication of unnecessary stops. • The method measures the presence of potential racial bias by comparing the rates of low-probability stops between racial groups, where more low-probability stops for a specific race shows the possibility of bias. • The method also allows comparisons of stops and hit rates for people of color and "similarly situated" white individuals. • The method also offers potential opportunities to achieve higher hit rates with fewer, more targeted	

[69] Gelman, A., *Fagan, J., & Kiss, A. (2007). An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias. Journal of the American Statistical Association, 813-823. http://www.stat.columbia.edu/~gelman/research/published/frisk9.pdf

			stops.	
Questions this method can answer	• *Is there evidence of racial bias in traffic stops?*	• *To what degree are racial differences in stops explained by higher crime rates or other factors in specific geographic areas?* • *After accounting for those differences, are there still significant differences in stops by race?*	• *What threshold of "reasonable suspicion" is being applied for stops on average? Is there evidence of different standards by race?* • *How much more likely is a member of a nonwhite racial group to be stopped than "similarly situated" white individuals?* • *Are there criteria for making stops that can be used to increase the hit rates of stops and lower the overall number of stops?*	• *If racial discrimination in who police stop exists, what is a reasonable estimate of racial disparities in the outcomes of stops?*
Questions this method cannot answer	• *Do officers use other methods of discerning driver race?* Do officers use streetlights, neighborhood characteristics, or the vehicle make or model to make guesses about the race of a driver? If they make these guesses about a driver's race often and guess accurately, **then this method will underestimate the amount of racial bias in**	• *If bias exists, at what level(s), per Goff's framework above, is that bias operating?* Gelman et al. note, "The summary statistics that we study here cannot directly address questions of harassment or discrimination, but rather reveal statistical patterns that are relevant to these questions." [71]	• *Is there evidence of bias in stops that do not have a clear outcome of interest or "hit"?* Goel et al (2016) focus primarily on stops for suspicion of criminal possession of a weapon, where the intended outcome—recovery of a weapon—is clear. For other stops, like stops for investigatory purposes, the outcomes are less clear (i.e., weapon	• *Whether racial discrimination in who police stop exists.* Method allows an analyst to input a measure of the bias in who is stopped. In the paper, they combine estimates of the bias in who is stopped (e.g., from Gelman et al. 2007) with the outcomes of those stopped for the same jurisdiction (New York City).

[71] Gelman, A., Fagan, J., & Kiss, A. (2007).

	stops. If the officers often make these guesses, but guess inaccurately, then the method will do little to uncover racial biases. • *Are there other factors that explain differences in who is stopped?* Amount of daylight fluctuates during the year, and driver behavior may change during the year. If those changes are correlated with race, this test may lead to incorrect conclusions. • *Are other types of stops—like pedestrian stops—driven by race?* Lack of daylight likely has less of an impact on an officer's ability to discern the race of an individual on the street than in a vehicle.		recovery may not be the appropriate metric). • *What chance of a hit (5%? 1%? 0.5%?) constitutes a "reasonable suspicion" and should that standard change by crime type and location?* This method only attempts to measure the rates that seem to constitute reasonable suspicion with the data available. What that threshold should be is a question of policy and values, with tradeoffs. • *What factors in the data collection process are affecting the ability to measure hit rates?* If police officers record stops and their justifications inconsistently or inaccurately it will limit the applicability of this method.	
Key Takeaways	• This method offers suggestive evidence of racial bias in police traffic stops. • The precision of the estimate is subject to the limitations described above, and generally this method is more narrowly	• This method offers a significant improvement in the precision of estimates of racial bias, as compared to simple benchmark tests. • However, since many	• This method offers a significant improvement in the accuracy of estimates of racial bias compared to simple outcomes tests by controlling for the locations of and reasons for a stop. • However, this method	• This method offers a significant improvement in the precision of estimates of biased outcomes of stops. • Important caveat: To apply this method, one must have a reasonable estimate of bias in who is

	applicable to traffic stops than to pedestrian stops. • This approach can offer a valuable starting point for further inquiry.	crimes go unreported[72] and many that are reported never result in an arrest,[73] arrest rates provide an incomplete proxy for the behaviors that might justify a police stop. • Moreover, if racial disparities exist in where police are deployed and who they choose to arrest, then arrest rates provide an unreliable proxy for behavior that would justify a stop. • To our knowledge, this method has not been applied to traffic stops, and may be more challenging because the crime rates in a specific area may have little meaningful correlation with driver behavior.	depends on having consistent and detailed documentation of stops and the reasons for those stops. A jurisdiction must have detailed and faithful reporting policies and procedures to utilize this method.	stopped for the jurisdiction one is analyzing. This may be developed through the application or extension of Gelman et al. (2007) or Goel et al. (2016) methods for Regression-Adjusted Benchmark Tests.
Sources	Grogger and Ridgeway (2006)[74]	Gelman et al. (2007)	Goel et al. (2016)	Knox et al. (2020)

[72] Morgan, R.E. & Truman, J.L. (2018). Criminal Victimization, 2017. Bureau of Justice Statistics Bulletin NCJ 252472; Gramlich, J. (1 March 2017). "Most violent and property crimes in the U.S. go unsolved." Pew Research Center.

[73] Gramlich, J. (1 March 2017). "Most violent and property crimes in the U.S. go unsolved." Pew Research Center; Federal Bureau of Investigation. (2019). Percent of Offenses Cleared by Arrest of Exceptional Means, 2019. Crime in the United States. U.S. Department of Justice.

[74] Grogger, J. & Ridgeway, G. (2006) Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness. Journal of the American Statistical Association, September 2006, Vol. 101, No. 475, pp. 878-887.

C. Prospective analytical approaches that will require additional data collection

Our third category covers methodological approaches that yield the most conclusive results regarding bias in stops but are more resource-intensive. They address the "denominator problem" that methods in Table 2 handle only partially and the methods in Table 1 ignore. These methods address the core issues related to relying on administrative data alone, supplementing that data with additional data that documents the denominator—the population at risk of official police contact (e.g., all drivers who are speeding or committing other violations) as well as those who do actually have police contact and the outcomes of those interactions. In these approaches, administrative data are combined with data collected from existing camera systems (e.g., video or photographic data from ATE devices) and/or systematic observation to understand the context in which decision-making about stops happens and to better control for factors which may lead to imprecise estimates (e.g., better estimates of suspicious behavior).

While these methods are the most conclusive, we recognize that they are also more resource-intensive. As such, a practical approach might be to target these methods to stops that have the greatest potential for being impacted by racial bias or stops that could lead to the significant disparate outcomes. Like the methods in Table 2, the bias these methods capture is most often at the individual officer level or district level. Similarly, the absence of evidence of individual bias would also suggest that disparities are driven by city-level biases. Unlike Table 2, however, these methods are useful in finding bias in who is stopped rather than simply measuring bias in outcomes after the stop. By measuring further "upstream" in the process of a stop, these methods can not only improve our understanding of who is stopped, but with a sufficient number of observations, will also yield strong analysis of bias in post-stop outcomes or "hits" as well.

We note that MPD had previously partnered with Lamberth Consulting in 2006 to conduct a study of stops in DC that applied the New Data Collection approach described in Table 3 to a small sample of intersections.[75] Future efforts could substantially improve on the quality of the original study conducted in DC and overcome the shortcomings of other analyses, as recommended by Knox et al. (2020), by leveraging the significant improvement in the quality and completeness of present-day administrative data on stops, a more representative and randomized selection of evaluation sites, and the application of more rigorous statistical tests for discrimination.

[75] Lamberth, J. (2006). Final Report for the Metropolitan Police Department in the District of Columbia. Data Collection and Benchmarking of the Bias Policing Project. Lamberth Consulting.

Table 3: Prospective analytical approaches that will require additional data collection

Method	Combining Administrative Data with Video or Photographic Data	New Data Collection
What is it?	• Jurisdictions are increasingly using cameras as part of Automated Traffic Enforcement (ATE) systems. • Method compares the rates at which minority drivers are ticketed by redlight and speed cameras to the rates they are stopped by police officers to measure racial discrimination in vehicular stops. • As noted in Howard University's policy considerations, ATE cameras capture traffic violations objectively without any consideration of a driver's race, but where these cameras are placed is an important part of the decision-making process. • If a jurisdiction has enough ATE devices to cover a large proportion of its roads, police officers will patrol the same areas that are covered by ATE cameras allowing comparison of outcomes between the two approaches. • By comparing tickets issued by ATE cameras to vehicular police stops that occur at a similar time and location, an analyst can compare whether stops of minority drivers are more likely than their white counterparts for similar infractions. • This method requires merging information from license plates with administrative records on the owner's race if it is collected by the jurisdiction's department of motor vehicles. In DC, like many other jurisdictions, a vehicle owner's race is not collected at registration. In these cases, the owner's address would be connected with geographic data on race and when aggregated across many stops will identify racial disparities in stops and other outcomes.	• Directly captures the behavior prior to a stop to assess racial bias in vehicular and street stops. • Researchers observe a random, representative sample of streets in-person or over video and record instances of behavior that could warrant a police stop and the apparent race of the individual exhibiting that behavior. • The rates at which individuals of different races display behavior that would justify a stop is then compared to the races of individuals actually stopped to assess racial discrimination exists in who is stopped and in the outcomes of stops. • Observations could be taken in real time or by observing footage from closed circuit TV, traffic cameras, body-worn cameras, or from cameras set up specifically for research purposes. • Importantly, these observations should be taken from a large, random sample of locations so that they can generalize to the jurisdiction as a whole.
Questions this	• *Are drivers of a specific race more likely to be stopped for*	• *Whether racial discrimination exists in who is stopped for*

method can answer	*moving violations in an area where both ATE cameras and police officers are present?*	*behaviors that would justify a street or vehicular stop.*
Questions this method cannot answer	• *Are drivers of a specific race more likely to be stopped in areas that are covered by ATE or police, but not both?* This method relies on the overlap of observations from ATE cameras and police officers. Where this overlap does not exist, it cannot detect racial bias in stops. In practice, it is likely that police officers are deployed to areas that do not have ATE cameras to spread resources more evenly.[76] • *Are drivers of a specific race more likely to be stopped for minor noncompliance in an area where both ATE cameras and police officers are present?* For example, broken taillights or failure to signal lane change would not be captured as violations by an ATE camera. Assessing racial discrimination in these types of stops would require individual analysis of the video and photos captured by ATE cameras.	• *How do rates of bias in who is stopped change over time?* Because this method requires detailed data collection it is resource intensive. As a result, it is unlikely that a jurisdiction will be able to conduct these studies regularly. So, they are unlikely to capture positive or negative changes over time without substantial regular investments in data collection.
Key Takeaways	• If the requisite data are available, this method provides an objective measure of the rate at which drivers of any race commit moving violations. • This method relies on a wide use of ATE cameras. • This method will not be able to measure bias in non-vehicular stops. • This is a proposed method, and not one that we are aware has been employed to date.	• This method is likely the only way to conclusively measure bias in all types of police stops, because it captures a wide variety of behavior that could warrant a police stop against which to compare the race of those stopped. • This method still requires careful attention to the process through which stops are initiated, like open warrants, 911 calls, and Be On the Lookout (BOLO) calls that are meant to identify specific individuals to be stopped. • This method **is time and resource intensive because it requires direct observation and data collection.**
Sources	Knox et al. (2020)	Knox et al. (2020), Lamberth (2006)

[76] An experimental design, in which a representative sample of locations are randomly assigned to ATE or police traffic enforcement on alternate dates, may address this question.

While these study designs do facilitate more accurate measurement of racial bias in stops, they do not address the question of the *source* of the disparity. To examine this question, further analyses, both qualitative and quantitative, of the geographic distribution of stops, the individuals stopped, the departmental policies governing stop behaviors, and historical data, are necessary.

Conclusion

Through our efforts to reimagine police stops in DC, we have endeavored to better understand the current state of stops—how and why stops are made, how they are experienced by police and those who are stopped, the public safety benefits and harms of stops—and to develop policy and research guidance aimed at promoting equitable public safety. As underscored in discussions throughout the *Reimagining Stops* workshop series, the work of reimagining stops is challenging, but necessary. This agenda reflects the input of a broad group of stakeholders, including the community members, advocates, researchers, practitioners, and policymakers who participated in the workshop series, and findings from a deep review of the available literature. Four key takeaways emerged from this effort:

1. **Accurate measurement of racial bias in police stops (and policing broadly) is challenging due to limitations in available administrative data. However, recent advancements offer far more accurate and actionable measures than are typically used.** Attention to detail and care in applying analytical methods is essential to generating accurate estimates of bias.

2. **Understanding the source(s) of bias is critical to developing effective policy solutions that target the root causes of the disparities we observe.** Much of the extant research considers individual-level officer bias, though bias in decisions at the department or community/city level can also play a significant role in spurring racially disparate outcomes. As Goff noted, we must be able to "imagine a world where racism does not require racist actors."[77]

3. **Stops can have far-reaching effects, beyond the immediate interaction between a police officer and civilian**. Multi-method research that considers a broader range of outcomes of interest and a wider set of police strategies involving stops, and that takes a multi-site approach to gauge the generalizability of findings, is needed to better understand the harms and benefits of stops and inform policy and practice.

4. **There is significant opportunity for innovation and creativity in reimagining police stops in ways that preserve the public safety benefits while reducing harms.** As communities

[77] *Goff, P. A. (2013). A measure of justice: What policing racial bias research reveals. In F. C. Harris & R. C. Lieberman (Eds.), Beyond discrimination: Racial inequality in a postracist era (p. 157–185). Russell Sage Foundation, p. 161.

around the country develop promising alternatives, rigorous documentation and evaluation of these efforts can help build the evidence base and inform policymakers and communities alike on what works to achieve the desired outcomes.

This document builds on these key takeaways, providing guidance on next steps to measure racial bias in police stops, identify the harms and benefits of stops, and assess potential policy and programmatic changes to stops to determine if they have the desired effect on improving public safety outcomes. The priority research questions set forth in the three sample learning agendas highlight both the gaps in our knowledge at present as well as the tremendous opportunity to learn—to inform our understanding of the optimal role of stops in public safety and community wellbeing, and continue building evidence on what works in DC and beyond.