

# Empirical Approaches to Identify Systemic Discrimination in Policing

Alex Chohlas-Wood  
Stanford University

Marissa Gerchick  
Stanford University

Sharad Goel  
Stanford University

Aziz Huq  
University of Chicago

Amy Shoemaker  
Stanford University

Ravi Shroff  
New York University

Keniel Yao  
Stanford University

On a typical day in the United States, more than 50,000 people are stopped by the police for traffic violations and other minor suspected offenses (Davis et al., 2018). Consequences range from fines, incarceration, and, in some circumstances, even physical injury or death (Baumgartner et al., 2018; Epp et al., 2014; Glaser, 2015). A growing body of statistical evidence indicates that these police stops—as well as post-stop police actions—suffer from persistent racial bias (Antonovics and Knight, 2009; Gelman et al., 2007; Legewie, 2016; Pierson et al., 2020; Simoiu et al., 2017; Voigt et al., 2017). This troubling pattern is often conceptualized as stemming from the decisions of individual, biased officers. Black drivers, for example, are less likely to be stopped at night, when a “veil of darkness” masks their race (Grogger and Ridgeway, 2006; Pierson et al., 2020), ostensibly because a biased officer’s ability to differentially stop Black drivers is limited when information on race is obscured.

Such discrete acts of “individual racism” are widely decried. They are also unlawful under the Equal Protection Clause of the Constitution’s 14th Amendment (*Washington v. Davis*). But individual racism is not the only cause of policing disparities. Both the terms “institutional racism” and “systemic racism” are used to capture the way *policies and practices* can impose disparate tolls on marginalized communities even absent any individual racial intent (Carmichael and Hamilton, 1967). For example, a police department’s choice to focus resources on enforcing minor traffic violations in communities of color may result in disparate impacts on those marginalized groups—even if individual officers make perfectly race-blind decisions when determining whom to stop, and even if no official intends harm or disadvantage toward any group.

This distinction between discrimination defined in terms of individual acts and mental states, and discrimination defined in terms of some pattern of institutional behavior, is echoed in the legal notions of “disparate treatment” and “disparate impact” (Goel et al., 2017). In the employment context, making a disparate impact case under Title VII of the Civil Rights Act requires showing both that racial groups receive benefits or penalties at different rates, and, critically, that there is no policy-relevant justification for the disparities. We apply this framework to examine systemic discrimination in policing. There is growing recognition among policymakers that understanding and addressing institutional racism in policing practices plays an important role in reform. However, despite this newfound appreciation, the development and adoption of empirical tools to identify and ameliorate disparities in policing that harm communities of color is still in its nascent stages.

Here we describe and apply three general approaches to empirically assess disparate impacts in policing—i.e., social costs that fall to an unjustifiable extent on communities of color. First, by drawing on a cost-benefit framework, we show that a policing practice of stopping hundreds of thousands of drivers in Nashville each year for minor traffic violations disproportionately burdened Black individuals, and did so without improv-

ing public safety. Next, by formalizing a concept of “similarity” between suspects, we show that Black and Hispanic individuals detained under New York City’s stop-and-frisk program were frisked more often than comparably risky white individuals, again highlighting an unnecessary cost borne by racial minorities. Finally, by adopting a holistic, system-wide perspective, we show that efforts to curtail Chicago’s use of stop-and-frisk may have resulted in displacing this problematic practice with another, perhaps equally discriminatory practice. In combination, these three examples demonstrate how data analysis can reveal hidden institutional racism. More generally, they illustrate the importance of viewing policing practices through the lens of disparate impact and systemic discrimination.

## Cost-benefit analysis to assess disparate impacts

When evaluating a law enforcement policy, it is important to ask whether the benefits justify the social costs, measured in part by any disproportionate burdens the policy creates for racial minorities. This approach is particularly useful given that many policing practices are touted as having crime-fighting benefits (Weisburd et al., 2015), and are thus frequently concentrated in high-crime neighborhoods, which—as the result of historical discrimination and neglect—are often communities of color. If a policy shows little or no public-safety benefit, and its costs are borne disproportionately by communities of color, the policy can be viewed as a form of systemic discrimination.

One such example occurred when the Metropolitan Nashville Police Department (MNPDP) adopted a policy of pulling over drivers for minor traffic violations, which resulted in sizable disparate impacts on the city’s Black community without improving public safety. At the time, MNPDP claimed that these stops prevented criminal activity by intercepting individuals driving to and from the scene of a crime. This belief drove the department to heavily rely on this tactic, and, in 2012, MNPDP conducted up to ten times as many traffic stops per capita compared to similar American cities. These stops were concentrated in high-crime Nashville neighborhoods, many of which are majority Black communities. As a result of this geographic concentration, as well as other factors, Black drivers in Nashville were stopped at significantly higher rates than white residents. These disparities were particularly pronounced for stops involving non-moving violations, such as broken tail-lights and expired license plates. For example, in 2012, Black drivers were stopped for non-moving violations 82% more often than white drivers, and thus bore a disproportionate burden of the costs of this policy.

In contrast to the department’s stated rationale, we found this traffic stop policy had no measurable crime-fighting benefit. First—in a direct test of MNPDP’s hypothesis—we examined whether the increased use of traffic stops in small MNPDP geographic units (known as “reporting areas”, or RPAs) were associated with drops in crime within the span of single weeks. In the left panel of Figure 1, each point represents a week in an RPA in 2017, and the axes represent departures from each RPA’s median level of crime and median number of traffic stops. The flat trend line indicates that there is no meaningful relationship between increased stops and crime levels. After adjusting for several factors (including recent local crime trends), the volume of traffic stops appears to be randomly distributed across time and space, allowing us to more precisely measure the causal effect of increased stops on public safety—an analysis that corroborates the visual result in Figure 1 (Chohlas-Wood et al., 2018).

Second, we examined the impacts of this policy over the course of many years, as it is possible that the crime-fighting benefit of traffic stops is not detectable in the span of a week, but rather takes place over longer timespans, or over larger geographies. In the right panel of Figure 1, we compare annual stop rates and annual crime rates for the entire city of Nashville between 2011 and 2017. If traffic stops were an effective tactic for reducing crime, we might expect to see crime rates rise when enforcement of traffic stops is reduced. However, even though stop rates dropped by almost 50% between 2012 and 2017, we see little change in the crime rate.

This cost-benefit analysis suggests MNPDP’s traffic stop policy created unjustified disparate impacts on the

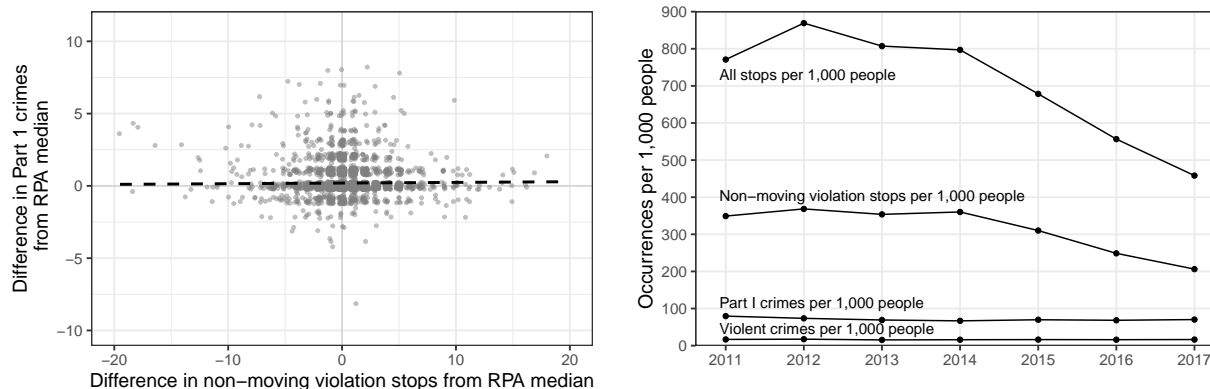


Figure 1: *Traffic stops in Nashville had little short-term or long-term benefit on crime reduction. Left: Weekly changes in stop volume within small, geographic units (RPAs) are uncorrelated with changes in local levels of serious (“Part 1”) crimes, as indicated by the flat slope of the dashed trend line. Right: Large, multi-year changes in traffic stop enforcement were likewise uncorrelated with long-term trends in serious crime. This stop policy lacked any clear public safety benefit, but placed substantial burdens on drivers, and particularly on the city’s Black drivers (not shown in figure).*

city’s Black drivers. In 2018, after being presented with the above findings—together with a sustained years-long campaign from community groups calling on the department to reconsider its policy—MNPd began significantly reducing traffic enforcement. In 2019, the department reported conducting approximately 56,000 stops, a nearly 80% reduction from 2017 levels, and saw no increase in crime rates.<sup>1</sup>

## Disparate decisions in comparable contexts

A central challenge in assessing disparities in policing involves a baseline question: How does one appropriately compare officer decisions across race groups when the context of those interactions may also vary across groups? For instance, suppose that in some city, police frisk stopped Black individuals more frequently than they frisk white individuals who have been stopped. This may reflect an unjustified racial disparity in officer frisk decisions (whether intentional or not); or it may instead reflect a justified disparity if stopped white individuals in this city generally pose a lower risk to officer safety than stopped Black individuals.

To disentangle these two possibilities, we apply a strategy called risk-adjusted regression (Jung et al., 2018) to measure the extent to which racial disparities among individuals frisked in New York City are justified by differences in individual riskiness (i.e., the chance that a stopped individual is in fact carrying a weapon). This strategy uses a statistical model to first estimate a risk score for each stop—the chances that a frisk, if conducted, would turn up a weapon—and then examines differences in frisk rates between minority individuals and white individuals with the same risk score. Under this approach, differences in frisk rates among similarly risky stopped individuals of different races, or “risk-adjusted disparities,” could be interpreted as unjustified disparities, namely those not explained by differences in otherwise observable threats to officer safety. It thus provides a race-specific measure of excessive police action in the controversial context of street stops.

In Figure 2, we display frisk rates by risk score for a dataset of approximately 1.2 million stops of Black, Hispanic, and white individuals in New York City between 2010 and 2011. Risk scores were calculated using a statistical model trained on the set of all frisks that took place between 2008 and 2009, using information recorded by the officer at the time the frisk occurred. At any given level of risk (on the horizontal axis), Black

<sup>1</sup>See <https://www.nashville.gov/Portals/0/SiteContent/Police/docs/Reports/2019VehicleStopReport.pdf>.

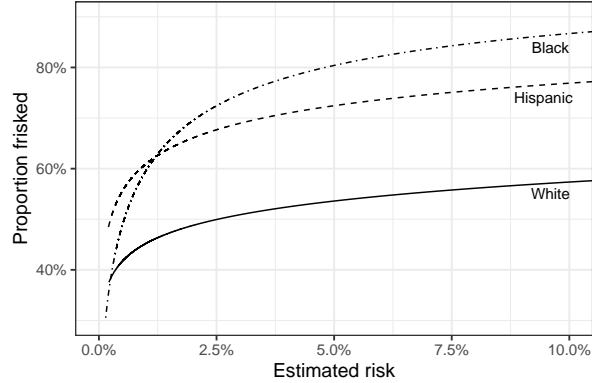


Figure 2: *Among similarly risky individuals detained under New York City’s stop-and-frisk program, Black and Hispanic individuals were frisked more often than white individuals. Here, risk is defined as the estimated likelihood of recovering a weapon from a frisk, based on a statistical model trained on all documented information available to the officer before the frisk.*

and Hispanic individuals were frisked considerably more often than white individuals. This pattern suggests that there exists a racial “surplus” of police frisks that cannot be explained by potentially “legitimate” differences between stopped individuals of different race groups.

Our approach above compares frisk rates among similarly risky individuals. In particular, and in contrast to a traditional disparate treatment analysis, we do not attempt to account for the full set of factors that may potentially provide a race-neutral explanation for the observed differences (Ayres, 2005). Suppose, hypothetically, that: (1) officers have a policy of frisking individuals entering or exiting public-housing complexes, regardless of an individual’s race; (2) that residents of these complexes are disproportionately Black or Hispanic; and (3) after adjusting for other observable factors, public housing residents do not have an elevated risk of carrying a weapon. In this scenario, differences in frisk rates across groups may not be driven by racial animus, but such a policy nonetheless creates a facially unjustified burden on racial minorities, and so is a form of systemic discrimination.

While risk-adjusted regression is a powerful strategy for identifying disparate police practices, we note two key limitations of the approach. First, “risk” (e.g., of possessing a weapon) may not be the only legitimate factor informing decisions. For instance, in our example, officers may be justified in applying a lower risk threshold for conducting a frisk on public transit, where the threat posed by a weapon may be more consequential. Accordingly, if minority individuals take public transit more often than white individuals, then observed risk-adjusted disparities may reflect this potentially justifiable difference rather than systemic discrimination. (In our case, we find that risk-adjusted disparities persist even after additionally adjusting for location, including whether a stop was conducted on public transit, but the general concern is an important one.) Second, the estimated risk may not fully account for the information available to decision makers. In theory, it is possible that officers can distinguish between individuals that appear similarly risky based on the recorded data. As above, this possibility may provide a justification for observed risk-adjusted disparities. In practice, one can assess the sensitivity of results to varying forms and degrees of such omitted variables. In our case, the large gaps in risk-adjusted frisk rates across race groups indicates officers would need access to substantial unrecorded information to erase the apparent disparities.

## A system-wide view of disparate impacts

The above examples illustrate the disparate impacts of specific enforcement policies considered in isolation—for example, Nashville’s use of traffic stops for minor infractions, and New York City’s use of stop-and-frisk.

But it is also important to consider the disparate impacts stemming from a collection of related policies that share similar methods and goals. Analyses that focus only on one specific policy, while often useful for targeted reform, may still underestimate the cumulative impacts of policing on marginalized communities.

In 2015, following many years of local activism, the Chicago Police Department (CPD) entered into a settlement agreement with the ACLU of Illinois to monitor its pedestrian stop practices. CPD reported a large and immediate reduction in the number of pedestrian stops per year, dropping from 600,000 stops in 2015 to 100,000 stops in 2016, the year the agreement went into effect. However, during the same time period, CPD steadily increased traffic stops—which were not monitored under the agreement. These rose from fewer than 100,000 reported in 2015 to more than 600,000 in 2019. As in Nashville, the bulk of the increased stops were for equipment and registration violations. Traffic stops, both in quantity and kind, increasingly resembled the pedestrian stops that policymakers and the public explicitly aimed to curb. By 2019, the race-specific stop rates for pedestrian and traffic stops combined were comparable to those prior to the settlement agreement (see Figure A1 in the online appendix).<sup>2</sup> The numbers serve as a cautionary tale: reforming institutional racism requires long-term evaluation of system-wide disparate impacts.

## Summary and conclusion

Discrimination—both in policing and beyond—is commonly viewed as the product of individuals acting with animus. The research community has accordingly focused much of its collective attention on identifying and quantifying such disparate treatment, developing statistical methods to estimate the effect of race on individual decisions (Gaebler et al., 2020). However, while important, that perspective largely fails to recognize the systemic discrimination stemming from policies or practices that may not be motivated by racial hostility per se, but nonetheless place undue burdens on marginalized communities. As Carmichael and Hamilton (1967) write, systemic discrimination “is less overt, far more subtle....But it is no less destructive of human life.” Indeed, the pernicious effects of systemic discrimination can be even more injurious than individual acts of racism.

Despite its significance, there is a relative dearth of empirical work on systemic discrimination. To close the gap, here we have provided three case studies that illustrate the application of statistical methods to identify and assess disparate impacts in policing. The approaches we have taken generalize, extending beyond the policing context, and provide templates for studying systemic discrimination in a broad set of policy domains. Looking forward, we hope this work will spur further interest in the empirical analysis—and remediation—of discrimination in all its forms.

## References

- Antonovics, K. and B. Knight (2009). A new look at racial profiling: Evidence from the Boston Police Department. *The Review of Economics and Statistics* 91(1), 163–177.
- Ayres, I. (2005). Three tests for measuring unjustified disparate impacts in organ transplantation: The problem of “included variable” bias. *Perspectives in Biology and Medicine* 48(1), S68–S87.
- Baumgartner, F. R., D. A. Epp, and K. Shoub (2018). *Suspect citizens: What 20 million traffic stops tell us about policing and race*. Cambridge University Press.
- Carmichael, S. and C. V. Hamilton (1967). *Black power: The politics of liberation in America*. Random House.

---

<sup>2</sup>The settlement agreement took effect on January 1, 2016. Concurrent with this work, Hausman and Kronick (2021) independently argued that the agreement had the effect of displacing pedestrian stops by traffic stops in Chicago.

- Chohlas-Wood, A., S. Goel, A. Shoemaker, and R. Shroff (2018). An analysis of the Metropolitan Nashville Police Department’s traffic stop practices. Technical report, Stanford Computational Policy Lab.
- Davis, E., A. Whyde, and L. Langton (2018). Contacts between police and the public, 2015. Technical report, Bureau of Justice Statistics.
- Epp, C., S. Maynard-Moody, and D. Haider-Markel (2014). *Pulled over: How police stops define race and citizenship*. University of Chicago Press.
- Gaebler, J., W. Cai, G. Basse, R. Shroff, S. Goel, and J. Hill (2020). A causal framework for observational studies of discrimination. Available online at: <https://arxiv.org/abs/2006.12460>.
- Gelman, A., J. Fagan, and A. Kiss (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* 102(479), 813–823.
- Glaser, J. (2015). *Suspect race: Causes and consequences of racial profiling*. Oxford University Press, USA.
- Goel, S., M. Perelman, R. Shroff, and D. A. Sklansky (2017). Combatting police discrimination in the age of big data. *New Criminal Law Review* 20(2), 181–232.
- Grogger, J. and G. Ridgeway (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association* 101(475), 878–887.
- Hausman, D. and D. Kronick (2021). When police sabotage reform by switching tactics. Available online at: <https://ssrn.com/abstract=3192908>.
- Jung, J., S. Corbett-Davies, R. Shroff, and S. Goel (2018). Omitted and included variable bias in tests for disparate impact. Available online at: <https://arxiv.org/abs/1809.05651>.
- Legewie, J. (2016). Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination. *American Journal of Sociology* 122(2), 379–424.
- Pierson, E., C. Simoiu, J. Overgoor, S. Corbett-Davies, D. Jenson, A. Shoemaker, V. Ramachandran, P. Barghouty, C. Phillips, R. Shroff, and S. Goel (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour* 4(7), 736–745.
- Simoiu, C., S. Corbett-Davies, and S. Goel (2017). The problem of infra-marginality in outcome tests for discrimination. *The Annals of Applied Statistics* 11(3), 1193–1216.
- Voigt, R., N. P. Camp, V. Prabhakaran, W. L. Hamilton, R. C. Hetey, C. M. Griffiths, D. Jurgens, D. Jurafsky, and J. L. Eberhardt (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences* 114, 6521–6526.
- Weisburd, D., A. Wooditch, S. Weisburd, and S.-M. Yang (2015, 11). Do stop, question, and frisk practices deter crime? Evidence at microunits of space and time. *Criminology & Public Policy* 15, 31–56.

# A Appendix

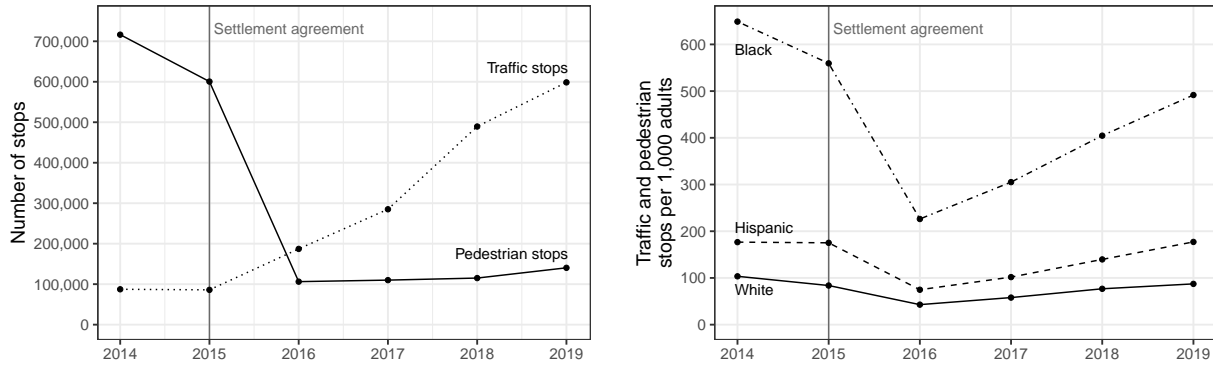


Figure A1: As pedestrian stops fell in Chicago after the settlement agreement, there was a rise in traffic stops, which were not monitored under the agreement (left panel). Due to this displacement of pedestrian stops with traffic stops, there was little change in the combined disparate impacts of police stops on Black and Hispanic residents, undermining the broader goals of the agreement (right panel).