

Identifying and Measuring Excessive and Discriminatory Policing

Alex Chohlas-Wood, Marissa Gerchick, Sharad Goel, Aziz Z. Huq, Amy Shoemaker, Ravi Shroff, and Keniel Yao†

We describe and apply three empirical approaches to identify superfluous police activity, unjustified racially disparate impacts, and limits to regulatory interventions. First, using cost-benefit analysis, we show that traffic and pedestrian stops in Nashville and New York City disproportionately impacted communities of color without achieving their stated public-safety goals. Second, we address a long-standing problem in discrimination research by presenting an empirical approach for identifying “similarly situated” individuals and, in so doing, quantify potentially unjustified disparities in stop policies in New York City and Chicago. Finally, taking a holistic view of police contact in Chicago and Philadelphia, we show that settlement agreements curbed pedestrian stops but that a concomitant rise in traffic stops maintained aggregate racial disparities, illustrating the challenges facing regulatory efforts. These case studies highlight the promise and value of viewing legal principles and policy goals through the lens of modern data analysis—both in police reform and in reform efforts more broadly.

INTRODUCTION

Police action is supposed to prevent or suppress private violence. In so doing, however, it often itself involves police coercion of—and, in some cases, violent use of force against—civilians. Even a temporary street stop can be “a serious intrusion upon the sanctity of the person, which may inflict great indignity and arouse strong resentment, and it is not to be undertaken lightly.”¹ The 2020 protests conducted under the banner of “Black Lives Matter” thrust the toll of police violence—which has been the subject of protests going back to at least the 1960s²—once more into the national spotlight. Those protests underscored that

† Respectively: Executive Director, Stanford Computational Policy Lab; Data Scientist, Stanford Computational Policy Lab; Professor of Public Policy, Harvard Kennedy School; Frank and Bernice J. Greenberg Professor of Law, University of Chicago; Data Scientist, Stanford Computational Policy Lab; Assistant Professor of Applied Statistics, New York University; Data Scientist, Stanford Computational Policy Lab.

¹ Terry v. Ohio, 392 U.S. 1, 17 (1968).

² See ELIZABETH HINTON, AMERICA ON FIRE: THE UNTOLD HISTORY OF POLICE VIOLENCE AND BLACK REBELLION SINCE THE 1960s, at 20, 313–38 (2021).

coercive policing tactics are not evenly distributed across different racial and ethnic groups. Minority racial status, poverty, and exposure to both violent crime and coercive policing are all tightly correlated—not just in Chicago but also in many other cities.³ As a result, the costs of coercive policing fall heaviest on the minority communities that are already exposed to the highest rates of crime.⁴

Chicago's experience with conjoined criminal violence, aggressive policing, and racial stratification has parallels in many other jurisdictions.⁵ Nationally, there are pervasive racial disparities in the frequency at which officers kill civilians, the most serious form of police violence.⁶ Other involuntary encounters with police are also unevenly dispersed.⁷ Even if police coercion fell evenly on all racial groups, its costs would not necessarily be felt equally. Among racial minorities, and Black individuals in particular, involuntary police contact is associated with “stigma,” “trauma,” “anxiety,” and “depressive symptoms.”⁸ Negative contact with police also reduces the willingness of individual minority citizens to later proactively seek police aid.⁹ Communities that

³ See generally John Rappaport & Aziz Z. Huq, *Symposium Introduction: This Violent City? Urban Violence in Chicago and Beyond*, 89 U. CHI. L. REV. 303 (2022).

⁴ See *id.* at 307–08.

⁵ For a careful study of the same dynamics in Milwaukee, see generally Jenna M. Loyd & Anne Bonds, *Where Do Black Lives Matter? Race, Stigma, and Place in Milwaukee, Wisconsin*, 66 SOCIO. REV. 898 (2018).

⁶ For recent studies, see, for example, Frank Edwards, Hedwig Lee & Michael Esposito, *Risk of Being Killed by Police Use of Force in the United States by Age, Race–Ethnicity, and Sex*, 116 PROC. NAT'L ACAD. SCI. 16793, 16793–95 (2019); Cody T. Ross, Bruce Winterhalder & Richard McElreath, *Racial Disparities in Police Use of Deadly Force Against Unarmed Individuals Persist After Appropriately Benchmarking Shooting Data on Violent Crime Rates*, 12 SOC. PSYCH. & PERSONALITY SCI. 323, 324–27 (2021).

⁷ On the uneven distribution of negative police encounters across race groups, see J.E. DeVlyder, H.Y. Oh, B. Nam, T.L. Sharpe, M. Lehmann & B.G. Link, *Prevalence, Demographic Variation and Psychological Correlates of Exposure to Police Victimization in Four US Cities*, 26 EPIDEMIOLOGY & PSYCHIATRIC SCI. 466, 473–74 (2017); and Emma Pierson et al., *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE HUM. BEHAV. 736, 737–39 (2020).

⁸ Amanda Geller, Jeffrey Fagan, Tom Tyler & Bruce G. Link, *Aggressive Policing and the Mental Health of Young Urban Men*, 104 AM. J. PUB. HEALTH 2321, 2321, 2323 (2014); see also Abigail A. Sewell & Kevin A. Jefferson, *Collateral Damage: The Health Effects of Invasive Police Encounters in New York City*, 93 J. URB. HEALTH S42, S48–S49 (Supp. 1 2016) (identifying correlations between police contact and indexes of poor physical health).

⁹ The key pathbreaking studies on this score were Patrick J. Carr, Laura Napolitano & Jessica Keating, *We Never Call the Cops and Here Is Why: A Qualitative Examination of Legal Cynicism in Three Philadelphia Neighborhoods*, 45 CRIMINOLOGY 445, 457–60 (2007); and Rod K. Brunson, *“Police Don't Like Black People”: African-American Young Men's Accumulated Police Experiences*, 6 CRIMINOLOGY & PUB. POL'Y 71, 71–72 (2007).

experience higher rates of such negative contact in the aggregate are less willing to subsequently seek police aid, which might increase their later exposure to violent crime. Qualitative studies further suggest that “legal estrangement” among “many African Americans and in many disadvantaged neighborhoods” is a distinctive—and unevenly experienced—effect of police coercion.¹⁰ The human toll for Black communities is further amplified by other background disparities. In Chicago, for example, because medical facilities capable of handling violent trauma are unequally distributed, the ultimate human toll of violence on Black communities is particularly high.¹¹ An accounting of the circumstances in which police violence may be justified, therefore, must attend not only to absolute costs and gains in crime control. It must also account for society’s distinctive racial structure—and so recognize the effects of racially asymmetrical distributions of aggressive policing and the distinctive effects of police violence upon racial minorities and communities experiencing concentrated, racialized poverty.

As one element of a broader inquiry into the problem of urban violence, we explore in this Essay the idea of “unnecessary police coercion.” We focus on different ways in which unnecessary policing tactics can be conceptualized, identified, and (crucially) measured. The term “unnecessary” is an inevitably evaluative one. It must be defined in relation to some baseline of *necessary* police coercion. But, of course, there is no consensus on how to define what counts as necessary. One approach would be to analyze policing in terms of alternate, nonpolicing strategies with the same impact on public safety in the long term but without the coercion-related costs. This analysis might focus on the possibility that investments in civil society generate reductions in crime on par with (or greater than) policing.¹² Or, more broadly, it could posit a more robust welfare state as an alternative mechanism of crime control and promote the transfer of resources from the carceral to the

¹⁰ Monica C. Bell, *Police Reform and the Dismantling of Legal Estrangement*, 126 *YALE L.J.* 2054, 2067, 2072 (2017).

¹¹ See Elizabeth L. Tung, David A. Hampton, Marynia Kolak, Selwyn O. Rogers, Joyce P. Yang & Monica E. Peek, *Race/Ethnicity and Geographic Access to Urban Trauma Care*, *JAMA NETWORK OPEN* (Mar. 8, 2019), <https://perma.cc/FF8Z-M98T>.

¹² See, e.g., Patrick Sharkey, Gerard Torrats-Espinosa & Delaram Takyar, *Community and the Crime Decline: The Causal Effect of Local Nonprofits on Violent Crime*, 82 *AM. SOCIO. REV.* 1214, 1233–34 (2017).

supportive welfare state.¹³ In effect, this is one way of glossing recent calls for police abolition.

We take a narrower—albeit complementary—policy-specific tack to the problem of identifying unnecessary police coercion. Policing is not a homogeneous activity. It is rather a bundle of different tactics, actions, and activities. Different strands of this bundle can be untangled and evaluated in relative isolation.¹⁴ Some elements of policing, including so-called hot spots policing¹⁵ and stop-and-frisk,¹⁶ have been studied empirically to clarify their impacts on crime. Our aim is to demonstrate the empirical traction that can be gained through a tactic-specific analysis that singles out unnecessary coercion. Given the racial dynamics of policing, we focus also on forms of policing that are not just unnecessary but discriminatory in the sense that they impose greater costs on racial minorities. Legal reform (as distinct from popular mobilization) can gain traction by identifying especially harmful measures and showing concretely the nature of the harm and the absence of any meaningful offsetting justification. While police reform in practice is often incremental and prone to backsliding or evasion,¹⁷ the specification of particular tactics that do more harm than good is, in our experience, a useful approach to achieving substantial changes in policing practice.

A tactic-specific analysis of necessity nevertheless can be developed in different ways. Here, we draw from the law two evaluative frameworks for the identification of unnecessary policing measures. One is cost-benefit analysis (CBA), which is familiar to legal scholars not least from the context of centralized review of agency action by the Office of Information and Regulatory Affairs (OIRA). The second is the framework for disparate-impact

¹³ See Julilly Kohler-Hausmann, *Guns and Butter: The Welfare State, the Carceral State, and the Politics of Exclusion in the Postwar United States*, 102 J. AM. HIST. 87, 88, 91 (2015) (discussing the “negative correlation between welfare spending and imprisonment rates” from a historical perspective). This is one lens through which to understand calls to “abolish” police. See Allegra M. McLeod, *Envisioning Abolition Democracy*, 132 HARV. L. REV. 1613, 1617–19 (2019).

¹⁴ There may also be interactions between different elements of policing strategy; we discuss one way that similar policies can interact in Part IV.

¹⁵ See generally, e.g., Anthony A. Braga & Brenda J. Bond, *Policing Crime and Disorder Hot Spots: A Randomized Controlled Trial*, 46 CRIMINOLOGY 577 (2008).

¹⁶ CHARLES R. EPP, STEVEN MAYNARD-MOODY & DONALD HAIDER-MARKEL, *PULLED OVER: HOW POLICE STOPS DEFINE RACE AND CITIZENSHIP* 31–33 (2014) (summarizing studies).

¹⁷ See Alex S. Vitale, Opinion, *The Answer to Police Violence is Not ‘Reform’. It’s Defunding. Here’s Why*, THE GUARDIAN (May 31, 2020), <https://perma.cc/T5W8-3AMX>.

analysis developed in employment discrimination and fair housing law. Both of these evaluative frameworks, we show, offer a plausible and tractable lens through which to conceptualize unnecessary policing—in particular, unnecessary policing with discriminatory impacts. Further, we recognize and document the risk of circumvention when reform proceeds on a tactic-by-tactic basis. The ensuing risk of “hydraulic”¹⁸ displacement is an important consideration in calibrating effective policy responses to unnecessary and aggressive police tactics.

This Essay’s core contribution is to introduce three empirical strategies for executing these conceptual frames. We first describe an empirical test of a tactic’s efficacy, in which the identification of little or no crime-suppression benefits suggests that the tactic’s direct costs to citizens are unnecessary on the very terms defined by local police leadership. Second, to implement disparate-impact analysis, we demonstrate how risk-estimation methods can be employed to identify instances in which the objective grounds for coercion, as defined by the police’s own behavior, demonstrate the existence of race-specific excesses of police action.¹⁹ Finally, we describe two instances where, even when a specific coercive policing tactic was suppressed, one can reasonably conclude that the coercion reappeared through substantially similar policing measures.

Part I outlines our two conceptual frameworks—CBA and disparate-impact analysis. Parts II through IV introduce the relevant empirical tests and demonstrate their utility with case studies from several major U.S. cities, including Nashville, New York, Chicago, and Philadelphia. We conclude by summarizing the potential for further empirical and conceptual research.

I. IDENTIFYING UNNECESSARY POLICING TACTICS

Consider first how CBA might apply to policing. Since 1982, any “significant regulatory action” by a federal regulatory agency has been subject to a CBA conducted by OIRA.²⁰ CBA, in general,

¹⁸ The idea of hydraulic effects in law is associated with Samuel Issacharoff & Pamela S. Karlan, *The Hydraulics of Campaign Finance Reform*, 77 TEX. L. REV. 1705 (1999).

¹⁹ See generally, e.g., Jongbin Jung, Sam Corbett-Davies, Ravi Shroff & Sharad Goel, *Omitted and Included Variable Bias in Tests for Disparate Impact*, ARXIV (Aug. 30, 2019), <https://perma.cc/Y59K-J5TA>.

²⁰ See Exec. Order No. 12,291 § 3, 3 C.F.R. 127 (1982), *revoked by* Exec. Order No. 12,866 § 11, 3 C.F.R. 638, 649 (1994), *reprinted as amended in* 5 U.S.C. § 601 app. at 86–91 (2006 & Supp. V 2011).

can be understood as “a welfarist decision procedure” that evaluates whether

a project increases overall well-being, relative to the status quo, if aggregate welfare in the project world is larger than aggregate welfare in the status quo; or, equivalently, if the welfare gains to those whose [sic] are better off in the project world are larger than the welfare losses to those who are worse off.²¹

In this guise, it offers a tool for evaluating policy choices but does not itself embody a decisive normative truth. This, of course, is not necessarily how CBA is always implemented, even in information- and expertise-rich environments such as the federal government.²² But it is a useful way of understanding CBA’s idealized function in relation to public-safety-related policy making.

So conceived, CBA is relatively bound on the range of permissible policing tactics.²³ How often is that bound violated by the police departments in the United States? It is hard to know. Few policing measures are presently subject to evaluation by CBA.²⁴ While there is some effort to evaluate benefits defined in terms of crime suppression, as Professor Rachel Harmon notes, “analysis of the *costs* of criminal justice policy continues to be anemic.”²⁵ Hence, the frequency with which the benefits of policing measures outweigh their costs is unknown. In a recent article, Professor Barry Friedman and Elizabeth Jánosky argue for broad application of CBA as a “natural corrective”²⁶ to the absence of “any sense

²¹ Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165, 194, 196 (1999) (emphasis omitted).

²² See, e.g., Daniel A. Farber, *Regulatory Review in Anti-regulatory Times*, 94 CHI.-KENT L. REV. 383, 400 (2019) (describing how President Donald Trump’s “new restrictions on agency rulemaking . . . have the potential to transform the regulatory process”).

²³ We can think of few if any instances in which a net-negative policing strategy in welfarist terms should be adopted. Some may conclude that distributive concerns can justify policies that detract from rather than add to welfare (especially when welfare is considered without accounting for diminishing marginal effects). We think, however, that few policing tactics fall into this category.

²⁴ See, e.g., Anthony A. Braga, William H. Sousa, James R. Coldren, Jr. & Denise Rodriguez, *The Effects of Body-Worn Cameras on Police Activity and Police-Citizen Encounters: A Randomized Controlled Trial*, 108 J. CRIM. L. & CRIMINOLOGY 511, 513 (2018) (noting how little scientific evidence there is on the effects of police technology).

²⁵ Rachel A. Harmon, *Federal Programs and the Real Costs of Policing*, 90 N.Y.U. L. REV. 870, 895 (2015) (emphasis in original).

²⁶ Barry Friedman & Elizabeth G. Jánosky, *Policing’s Information Problem*, 99 TEX. L. REV. 1, 46 (2020) (quoting Cass R. Sunstein, *Cognition and Cost-Benefit Analysis* 10 (Univ. of Chi. L. Sch. Coase-Sandor Inst. for L. & Econ., Working Paper No. 85, 1999)).

of whether a particular policing tactic, strategy, or technology is worth the cost.”²⁷ However, beyond the statement that CBA should account for “the full range of costs and benefits,”²⁸ Friedman and Jánszky provide little by way of specifics about what such an analysis would entail.

Given the absence of systemic efforts to collect data on the physical, dignitary, and social costs of policing,²⁹ it is difficult to see how full-blown CBA can be effectively implemented across the board. A more limited version of CBA in the absence of across-the-board data about policing costs might nevertheless get off the ground by identifying coercive tactics of little or no benefit in terms of crime suppression. Such tactics are unlikely to pass muster on the aggregative, welfarist terms of CBA, so they can safely be classed as unnecessary. That benchmark for necessity, however, would be underinclusive. It would not include policing tactics that have some small (or even large) benefits where those benefits did not exceed the associated costs.

CBA, applied in the policing context, might also be used as a way to identify practices with discriminatory effects. Many policing practices touted as having crime-fighting benefits are concentrated in high-crime neighborhoods.³⁰ As a result of past and present discrimination and neglect by private and public actors—in Chicago no less than in other U.S. cities—these are often communities of color.³¹ Where a policy shows little or no public-safety benefit and its costs are borne disproportionately by communities of color, that policy can be ranked as both unnecessary and a vector of structural discrimination. We explain here how such policies can be identified.

A second, related theory of unnecessary policing draws on the idea of disparate-impact liability developed first in the federal employment-discrimination context and then exported to fair housing and other policy domains. In 1971, the Supreme Court read the liability provision of Title VII of the Civil Rights Act of

²⁷ *Id.* at 3.

²⁸ *Id.* at 49.

²⁹ For an example of such an effort, see Aziz Z. Huq, *The Consequences of Disparate Policing: Evaluating Stop and Frisk as a Modality of Urban Policing*, 101 MINN. L. REV. 2397, 2429–40 (2017) (cataloging various costs of street policing in particular).

³⁰ See, e.g., David Weisburd, Alese Wooditch, Sarit Weisburd & Sue-Ming Yang, *Do Stop, Question, and Frisk Practices Deter Crime? Evidence at Microunits of Space and Time*, 15 CRIMINOLOGY & PUB. POL. 31, 46 (2016).

³¹ See Michael Friedson & Patrick Sharkey, *Violence and Neighborhood Disadvantage After the Crime Decline*, 660 ANNALS AM. ACAD. POL. & SOC. SCI. 341, 352–54 (2015).

1964 to reach cases where a practice has an adverse disparate impact on Black employees.³² This theory of liability was subsequently extended to the Age Discrimination in Employment Act³³ and the Fair Housing Act,³⁴ among other statutory contexts. For policing, a disparate-impact standard is available both under federal statutes that regulate local police departments as recipients of federal funds³⁵ and under state statutes in California³⁶ and Illinois.³⁷

Across all these contexts, disparate impact has been understood to play one of two functions.³⁸ First, it is an instrument for rooting out improperly motivated actions in the absence of direct evidence of illegal intent. Disparate impact, in this guise, reflects the law's frequent presumption that a person's intentions can be inferred from the expected results of their actions. Second, and separately, the disparate-impact theory of discrimination picks out a legal wrong distinct from intentional discrimination. It encompasses actions that have the effect—regardless of intent—of entrenching the subordinate or inferior status of a group that has historically been subject to discrimination, disadvantage, or

³² See *Griggs v. Duke Power Co.*, 401 U.S. 424, 429–36 (1971).

³³ 29 U.S.C. §§ 621–634; see also *Smith v. City of Jackson*, 544 U.S. 228, 240 (2005).

³⁴ 42 U.S.C. §§ 3601–3619, 3631; see also *Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmty. Project, Inc.*, 576 U.S. 519, 545–46 (2015); Reva B. Siegel, *The Constitutionalization of Disparate Impact—Court-Centered and Popular Pathways: A Comment on Owen Fiss's Brennan Lecture*, 106 CALIF. L. REV. 2001, 2011 (2018) (“Congress has incorporated effects standards and accommodation requirements into a large body of civil rights legislation governing discrimination on the basis of sex, disability, and religion, as well as race.”).

³⁵ Title VI of the Civil Rights Act of 1964 and its implementing regulations apply to police departments that receive federal funds. See 42 U.S.C. § 2000d (“No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.”); see also 28 C.F.R. §§ 42.101–112 (implementing regulations). The Safe Streets Act also prohibits local police action with a racially disparate impact. See 42 U.S.C. § 10228(c)(1):

No person in any State shall on the ground of race, color, religion, national origin, or sex be excluded from participation in, be denied the benefits of, or be subjected to discrimination under or denied employment in connection with any programs or activity funded in whole or in part with funds made available under this chapter.

See also 28 C.F.R. § 42.203 (implementing regulations).

³⁶ See CAL. GOV'T CODE § 11135(a) (West 2017); CAL. CODE REGS. tit. 2, § 11154(c), (i) (2021).

³⁷ See 740 ILL. COMP. STAT. 23/5(a)(1) (2021).

³⁸ See Richard A. Primus, *Equal Protection and Disparate Impact: Round Three*, 117 HARV. L. REV. 493, 522–24 (2003); see also Owen Fiss, *The Accumulation of Disadvantages*, 106 CALIF. L. REV. 1945, 1947 (2018) (explaining how the *Griggs* principle pushes firms to create employment tests that minimize disparate impact on Black people).

deprivation.³⁹ This second reason has particular resonance given the historical pedigree of policing as a site where the ideological and material grounds of Black subordination are produced.⁴⁰ The dual functionality of disparate impact means that it can be used both as a test for malign, and hence unlawful, subjective intent—i.e., an individual-focused understanding of discrimination⁴¹—and also as a way of identifying practices with structural effects on racial formations. This last concept, to be sure, is a complex one; we briefly address in the margin what we mean by saying this without addressing the many complications and difficulties entailed.⁴²

In the employment-litigation context, disparate impact is implemented through a three-part burden-shifting test.⁴³ First, a plaintiff shows that a specific practice denies employment opportunities to the protected group. Second, a defendant can point to a business-related justification for the practice. Finally, the plaintiff can identify an alternative that advances the defendant's goals with a reduced racially disparate impact.⁴⁴

³⁹ See *Washington v. Davis*, 426 U.S. 229, 239, 247 (1976).

⁴⁰ A superlative study is KHALIL GIBRAN MUHAMMAD, *THE CONDEMNATION OF BLACKNESS: RACE, CRIME, AND THE MAKING OF MODERN URBAN AMERICA* (2010).

⁴¹ Constitutional and statutory doctrine respecting discrimination is, in fact, unclear about what makes an intent impermissible. See Aziz Z. Huq, *What Is Discriminatory Intent?*, 103 CORNELL L. REV. 1211, 1240 (2018).

⁴² The term “structural racism” is often used in the legal scholarship either without a clear definition or with only a tautological definition. See, e.g., Michael Siegel, *Racial Disparities in Fatal Police Shootings: An Empirical Analysis Informed by Critical Race Theory*, 100 B.U. L. REV. 1069, 1075 (2020) (“[S]tructural racism is not merely the processes that have led to disadvantaged conditions for people of color, but the *current conditions* that resulted from structural racism, even if the discriminatory processes occurred in the past.” (emphasis in original)). The problem is not limited to law. A recent medical science paper suggests that while structural racism has no single definition, it connotes the view that “racism is not simply the result of private prejudices held by individuals, but is also produced and reproduced by laws, rules, and practices, sanctioned and even implemented by various levels of government, and embedded in the economic system as well as in cultural and societal norms.” Zinzi D. Bailey, Justin M. Feldman & Mary T. Bassett, *How Structural Racism Works — Racist Policies as a Root Cause of U.S. Racial Health Inequities*, 384 NEW ENG. J. MED. 768, 768 (2021). Notice that, even here, it is not clear how “racism” is being defined.

We define here structural racism as reaching institutional, legal, and social processes that preserve and transmit forward in time historically durable patterns of disadvantage organized about racial categories. This definition is meant to capture the array of forces that promote and reproduce a perceived or actual correlation between racial identity and social, economic, cultural, or legal disadvantage.

⁴³ See *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 425 (1975).

⁴⁴ The framework applied to disparate-impact claims under the Fair Housing Act uses the terms “substantial, legitimate, nondiscriminatory interests” and requires a showing that “substantial, legitimate, nondiscriminatory interests supporting the challenged

This is not a test of instrumental rationality as such. It is instead a way of sorting (assuming sufficient evidence is available on all relevant points) between (1) policies that impose racially asymmetrical burdens in the absence of an approach that is less burdensome *for the protected racial group* and (2) policies that impose racially asymmetrical burdens where there is a less racially burdensome approach for that group. That is, disparate-impact liability should not be understood as a comprehensive mechanism for identifying all “structural” forms of “racial stratification that ha[ve] survived the abolition of slavery and the dismantling of Jim Crow.”⁴⁵ Like CBA, the standard doctrinal framework for disparate impact is instead best understood as one element of a larger inquiry that flags some (but not all) pathways by which historical patterns of disadvantage are maintained, transmitted forward in time, and perhaps exacerbated.⁴⁶ In other words, disparate impact is a way of isolating some (but not all) measures that are both unnecessary and discriminatory in effect. In this way, disparate-impact analysis is distinct from CBA both in its focus on racial subordination and in its explicit consideration of alternative policies for achieving one’s stated goals.

As we shall see, it is not always obvious how to implement the disparate-impact framework in the policing context, just as it is not always clear how CBA can be implemented. In our experience, police departments do not typically define the policy objectives of a specific tactic with sufficient precision to permit an evaluation of benefits. Hence, it is often difficult or impossible to ascertain whether those goals are, in fact, being advanced. At the third step of the burden-shifting test, yet another difficulty in application is created by an absence of information about the marginal effects on racial disparities of shifting from one tactic to another. That is, even in its most straightforward form, disparate impact is akin to CBA in being too empirically demanding for most contemporary policing contexts.

A simple—if yet more incomplete—version of disparate-impact liability might be framed around a comparative analysis: If departments can be observed policing two similarly situated

practice could be served by another practice that has a less discriminatory effect.” *Inclusive Cmty. Project, Inc.*, 576 U.S. at 527 (first quoting 24 C.F.R. § 100.500(c)(2) (2014); and then quoting 24 C.F.R. § 100.500(c)(3) (2014)).

⁴⁵ Fiss, *supra* note 38, at 1949–50.

⁴⁶ See *id.* at 1948–50 (noting the incompleteness of disparate impact as a lens onto structural disadvantage).

populations in terms of crime-related risk and if one of those populations (say, members of a minority) is persistently treated more harshly and with greater amounts of coercion, then it is reasonable to infer that an observed surplus of coercion directed at that population is unjustified and unnecessary. That is, if a policy imposes an excess of costs on racial minorities in relation to a similarly situated white group, it violates the disparate-impact rule. The “similarly situated” phrase here captures both the second and the third steps of the doctrinal burden-shifting test. Any effort to operationalize this approach necessarily turns on the ability to credibly identify similarly situated groups.

In sum, CBA and disparate-impact analysis each provide a distinct lens through which to analyze and evaluate the problem of unnecessary policing. Neither is complete; they each instead capture a slice of what can plausibly be termed unnecessary state coercion. They provide perspectives on a problem that, to date, has otherwise largely escaped empirical scrutiny.

II. EMPIRICAL TESTS FOR UNNECESSARY POLICING

We now turn to three empirical strategies for identifying unnecessary—or unnecessary and discriminatory—policing. First, by drawing on a CBA framework, we show that a policing practice of stopping hundreds of thousands of people in Nashville and New York each year disproportionately burdened Black individuals and did so without clear gains in public safety. Next, by formalizing a concept of “similarity” between suspects, we show that Black and Hispanic individuals detained under New York’s and Chicago’s stop-and-frisk programs were frisked more often than comparably risky white individuals. This again highlights an unnecessary cost borne by racial minorities. Finally, by adopting a holistic, system-wide perspective, we show that efforts to curtail Chicago’s and Philadelphia’s uses of pedestrian stops likely resulted in a displacement of this problematic practice with another, perhaps equally discriminatory, practice. In combination, these three case studies demonstrate how data analysis can isolate policing tactics that are both unnecessary and discriminatory in effect.

When evaluating a law-enforcement policy using CBA, it is necessary to ask whether its benefits justify its social and financial costs. We start by showing how one can identify policing practices with little or no crime-suppression benefits, making them unlikely to pass muster under CBA.

A. Nashville

In the early 2010s, the Metropolitan Nashville Police Department (MNPd) adopted a policy of pulling over drivers for minor traffic violations.⁴⁷ At the time, MNPd claimed that these stops prevented criminal activity by intercepting individuals driving to and from the scene of a crime, a belief used to justify the department's heavy reliance on this tactic.⁴⁸ In 2012, the MNPd conducted traffic stops up to ten times more frequently per capita than police departments in similar U.S. cities.⁴⁹ These stops were strategically concentrated in high-crime Nashville neighborhoods, many of which were majority-Black communities.⁵⁰ Partially as a result of this geographic concentration, Black drivers in Nashville were stopped at significantly higher rates than white drivers.⁵¹ These disparities were particularly pronounced for stops involving nonmoving violations, such as broken taillights and expired license plates.⁵² For example, in 2017, Black drivers were stopped 68% more often than white drivers for nonmoving violations.⁵³ Higher levels of enforcement in high-crime neighborhoods accounted for some, but not all, of this disparity. After adjusting for these local differences, Black drivers were stopped 37% more often than white drivers.⁵⁴ Black drivers thus bore a disproportionate burden of the costs of this policy.

With these racially disparate burdens in mind, our CBA-informed strategy next considers the policy's potential benefits. The MNPd justified its policy of concentrating stops in high-crime communities by arguing that the widespread use of traffic stops would reduce serious crime, such as burglary.⁵⁵ We evaluate this claim here by drawing on traffic-stop and crime data supplied by the MNPd. For clarity, we focus on the top-line results rather than on the technical details of the analysis.⁵⁶

⁴⁷ For a discussion of MNPd traffic stops during this period, see generally Alex Chohlas-Wood, Sharad Goel, Amy Shoemaker & Ravi Shroff, *An Analysis of the Metropolitan Nashville Police Department's Traffic Stop Practices*, STAN. COMPUTATIONAL POL'Y LAB (2018), <https://perma.cc/5332-2ELP>.

⁴⁸ *See id.* at 6.

⁴⁹ *See id.*

⁵⁰ *See id.* at 4.

⁵¹ *See id.* at 3–6.

⁵² *See* Chohlas-Wood et al., *supra* note 47, at 2.

⁵³ *See id.*

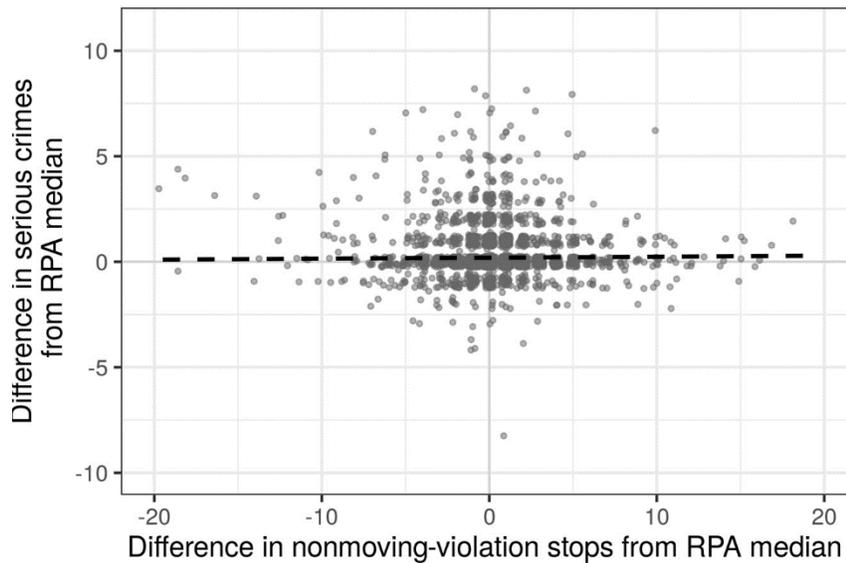
⁵⁴ *See id.* at 5.

⁵⁵ *See id.* at 1, 6.

⁵⁶ For those technical details, see *id.* at 3–8.

We start by examining whether increased use of traffic stops was associated with localized drops in crime within the span of single weeks. In particular, based on the MNPDP's stated rationale, we would expect crime to fall below typical levels in locations and time periods where traffic stops surged above typical levels. By the same token, we would expect crime to surge when traffic stops dropped below typical levels.

FIGURE 1: NONMOVING-VIOLATION STOPS VERSUS SERIOUS CRIME IN NASHVILLE



We first examine the relationship between stops and crime graphically in Figure 1. Each point represents a week in a small MNPDP geographic unit—known as a “reporting area,” or RPA—in 2017. The vertical axis represents departures from each RPA’s median level of serious crime, defined by what the FBI terms “Part I” crimes, which include homicide, rape, robbery, assault, burglary, and theft.⁵⁷ The horizontal axis indicates each week’s departure from the median number of nonmoving-violation stops in that RPA. If stops reduced crime, we would expect to see the dots generally fall along a downward sloping line, corresponding to unusually low levels of crime during periods of unusually high enforcement. Instead, the flat trend line suggests that there was no meaningful local relationship between increased stops and

⁵⁷ See UNIF. CRIME REPS., *Offense Definitions*, FBI, <https://perma.cc/CB2C-B4WK>.

crime levels in Nashville in 2017. This is in contrast to what the MNPDP had offered as justification for the policy.

In theory, it is still possible that stops do prevent crime, but the observed lack of correlation between the two stems from other factors that mask their relationship. For example, though it would have been practically challenging to do, the MNPDP could have deployed stops precisely to avoid local spikes in crime. In other words, it is possible that the MNPDP anticipated weeks with higher levels of criminal activity and deployed increased stops to suppress crime rates back to their median levels.

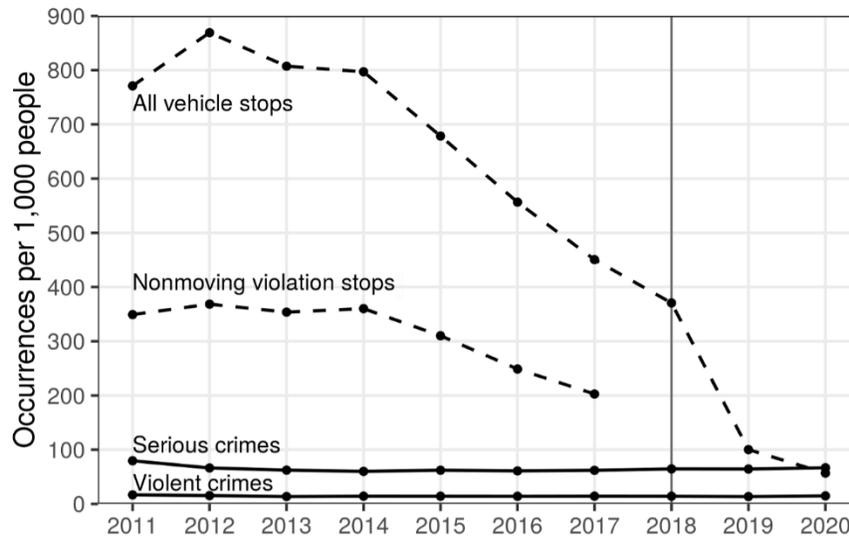
To assess this possibility, we examined the incidence of serious crime in locations and weeks that were comparable along the dimensions that the MNPDP might have systemically considered when making deployment decisions but that varied in the actual number of stops made. Every week, the MNPDP made deployment decisions for the following week that would go into effect each Sunday. It based these decisions in part on the location-specific crime trends from the previous week using CompStat—a popular system among police departments for tracking and responding to crime. Therefore, we model Sunday-to-Saturday crime level as a function of reported crime in the previous week, the number of deployed stops in the previous week, the specific RPA, and the month in which the week begins (to account for seasonality). After adjusting for these factors, we believe that it is reasonable to consider weekly fluctuations in the number of stops made to be approximately random—facilitating estimates of the causal effect of stops on crime.

Formally, using a Poisson regression that adjusts for the above factors, we find no evidence of a meaningful relationship between stops and serious crime. Specifically, we find that a one-standard-deviation increase above the average number of stops for any given RPA is associated with a 2.5% (95% confidence interval: 1.2%–4.4%) increase in per capita crime for that RPA. This supports the inference that we drew from the pattern observed in Figure 1: it suggests that there is minimal causal connection between nonmoving-violation stops and serious crime.

Our analysis above examined the immediate effects of the MNPDP's stop policies, but it is possible that the benefits of traffic stops are only apparent over the course of many years. In Figure 2, we compare annual stop rates—for all vehicle stops as well as the subset of stops for nonmoving violations—to annual crime rates in Nashville between 2011 and 2020. (Data on the number of

stops for nonmoving violations were available only from 2011 to 2017.) If these stop policies were an effective tactic for reducing crime in the long run, we would expect to see that crime rates rose when enforcement of traffic stops was reduced, and vice versa. But even with dramatic changes in stop rates, we see little associated change in crime rates.⁵⁸

FIGURE 2: TRAFFIC STOPS VERSUS SERIOUS CRIME IN NASHVILLE



In aggregate, our results suggest that the MNPd's traffic-stop policy had little to no immediate or long-term crime-fighting benefit, even as it disproportionately burdened the city's Black residents. In 2018—in response to the above findings and a sustained years-long campaign from community groups calling on the department to reconsider its policy—the MNPd significantly reduced traffic enforcement.⁵⁹ In 2019, the department reported conducting approximately 56,000 stops, a nearly 80% reduction from 2017 (and a nearly 90% reduction from 2012).⁶⁰ The city saw no increase in crime rates, providing further evidence that traffic stops had little crime-fighting benefit.⁶¹

⁵⁸ See Chohlas-Wood et al., *supra* note 47, at 6.

⁵⁹ See Samantha Max, *Nashville Police Report Major Drop in Traffic Stops Following Accusations of Racial Bias*, WPLN NEWS (Mar. 25, 2021), <https://perma.cc/8ECA-KA7V>.

⁶⁰ *Id.*

⁶¹ See *Police Data Dashboard: Uniform Crime Reporting Incidents Map*, METRO. GOV'T OF NASHVILLE & DAVIDSON CNTY., <https://www.nashville.gov/departments/police/data-dashboard/ucr-incidents-map>.

B. New York

Similar to the MNPDP's policy of pulling over vehicles for non-moving violations, the New York City Police Department (NYPD) had a long-standing policy of stopping and frisking pedestrians for suspicion of criminal activity.⁶² Stop-and-frisk is predicated—similar to MNPDP's policy's rationale—on the belief that frequent police contact is an effective tool for disrupting more serious crime.⁶³ Stop-and-frisk, however, differs from the MNPDP's practice of stopping vehicles for minor infractions, as stop-and-frisk stops—known as *Terry* stops—are based only on “reasonable suspicion” of illegal activity, rather than the actual observance of some violation.⁶⁴

At the peak of the city's use of stop-and-frisk, NYPD officers reported conducting nearly 700,000 *Terry* stops in 2011 alone, nearly 90% of which involved Black or Hispanic pedestrians.⁶⁵ For comparison, about half of New York City's residents at the time identified as Black or Hispanic.⁶⁶ Thus, as in Nashville, the policy placed substantially disproportionate burdens on racial and ethnic minorities.

One of the primary stated aims of stop-and-frisk was to get weapons off the street as a means to prevent violent crime.⁶⁷ In 2011, about 7,300 of the nearly 700,000 reported stops led to the recovery of a weapon, and only about 800 of those yielded a gun.⁶⁸ In other words, one gun was recovered for approximately every 900 stops. It is hard to measure the precise benefit of taking one illegal gun off the streets, but we suspect that many would conclude that the benefit is outweighed by the social and financial costs of conducting nearly 900 stops.

We take two empirical approaches to more rigorously examine the relative costs and benefits of stop-and-frisk. First, we

⁶² This analysis was aided by the existence of a relatively high-quality data set created as a consequence of constitutional litigation. See Andrew Gelman, Jeffrey Fagan & Alex Kiss, *An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias*, 102 J. AM. STAT. ASS'N 813, 813 (2007); Sharad Goel, Justin M. Rao & Ravi Shroff, *Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy*, 10 ANNALS APPLIED STAT. 365, 365 (2016).

⁶³ See Goel et al., *supra* note 62, at 365.

⁶⁴ See Gelman et al., *supra* note 62, at 814.

⁶⁵ See N.Y. C.L. UNION, STOP-AND-FRISK 2011, at 5 (2012), <https://perma.cc/D8E9-YQXQ>.

⁶⁶ NYC, DEP'T OF CITY PLAN., NYC2010: RESULTS FROM THE 2010 CENSUS 10 (2011), <https://perma.cc/2NDZ-SMRX>.

⁶⁷ See N.Y. C.L. UNION, STOP AND FRISK: REPORT ON 2011 FINDINGS, <https://perma.cc/QT67-KKXE>.

⁶⁸ See N.Y. C.L. UNION, *supra* note 65, at 8.

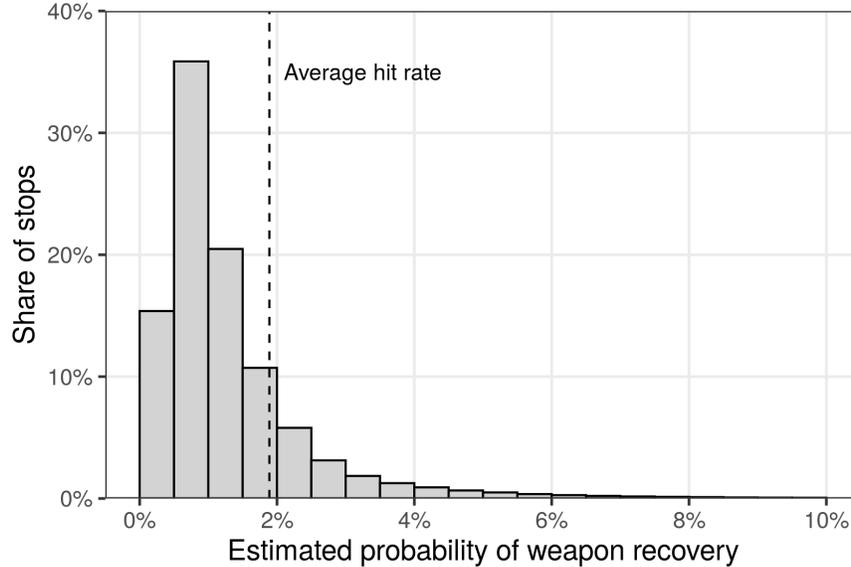
investigate the extent to which an alternative, more tailored stop policy could have achieved the stated benefits of stop-and-frisk at lower costs. In contrast to our examination of Nashville, this first analysis does not show that stop-and-frisk had *no* benefits. Rather, we demonstrate that a modified policy would have struck an arguably better trade-off between its costs and benefits. Our approach is thus akin to the third step of the burden-shifting framework of disparate-impact liability described above.⁶⁹

To develop a tailored stop policy, we first use the historical stop data to build a statistical model that estimates the likelihood that any stop would yield a weapon (also known as the “hit rate”) based solely on information available to officers at the moment immediately before the stop was conducted.⁷⁰ Specifically, we use the wealth of information recorded by officers on the stop form (called a UF-250), including the date, time, location, and recorded circumstances of the stop (e.g., whether the suspect had a “suspicious bulge”). Given that our focus is on the recovery of weapons, we limit our analysis to the 622,826 reported stops in 2009–2010 that resulted in a frisk. In effect, these (historical) data allow us to build a model in which each of the indicia of suspicion is assigned a value that reflects the likelihood that the stop will result in a weapons seizure.

⁶⁹ See *Albemarle Paper Co.*, 422 U.S. at 425.

⁷⁰ For this analysis, we draw on our earlier work in Goel et al., *supra* note 62, and Sharad Goel, Maya Perelman, Ravi Shroff & David A. Sklansky, *Combating Police Discrimination in the Age of Big Data*, 20 NEW CRIM. L. REV. 181 (2017).

FIGURE 3: ESTIMATED PROBABILITY OF WEAPON RECOVERY FOR *TERRY* STOPS IN NEW YORK CITY, 2011–2012



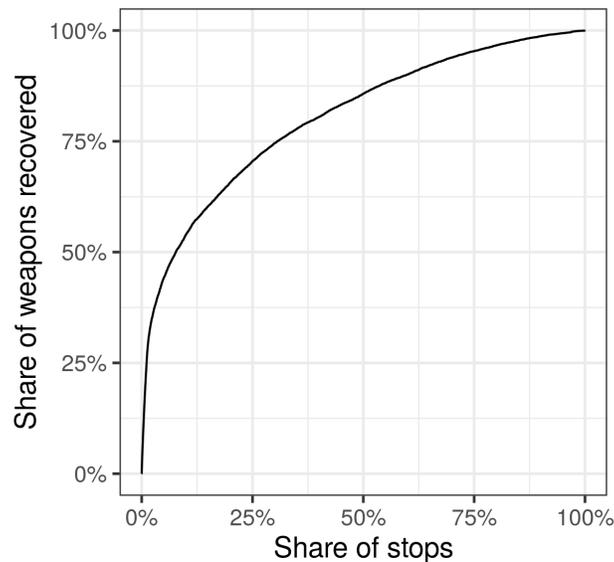
We next use this fitted statistical model to estimate the likelihood that stops conducted in 2011–2012 would yield a weapon based on information available prior to the stop itself. In other words, we use the correlations between different indicia of suspicion and weapons seizures in 2009–2010 to assign a likelihood of a successful weapon recovery to each stop in 2011–2012. Figure 3 shows the distribution of these estimated likelihoods. In particular, the average prediction is 2%, and more than 15% of stops had less than a 0.5% chance of turning up a weapon. At the other end of the spectrum, a modest number of stops had at least a 5% chance of yielding a weapon—which is more than twice the overall average.

This heterogeneity in predictions suggests that there is information available to officers that would allow them to focus on the stops most likely to yield a weapon—namely, those with high estimated probabilities. Changing policing practice thus has the potential to generate efficiencies.

To understand the available scope for such efficiencies, in Figure 4, we estimate the number of weapons we could expect officers to recover under a hypothetical policy in which one only conducts the p -percent of stops deemed *most likely* under the statistical model to result in the recovery of a weapon. In

particular, we estimate that by focusing on the 50% of stops deemed most likely to be successful, one could recover nearly 90% of the weapons recovered under the full set of stops. This demonstrates that the police could obtain nearly all the stated benefit of stop-and-frisk (namely, weapons recovery) while halving the direct costs, as measured by the number of stops. This suggests that the police are not engaged in efficient policing—in fact, much is unnecessary and costly.

FIGURE 4: ESTIMATED SHARE OF STOPS VERSUS SHARE OF WEAPONS RECOVERED IN NEW YORK CITY



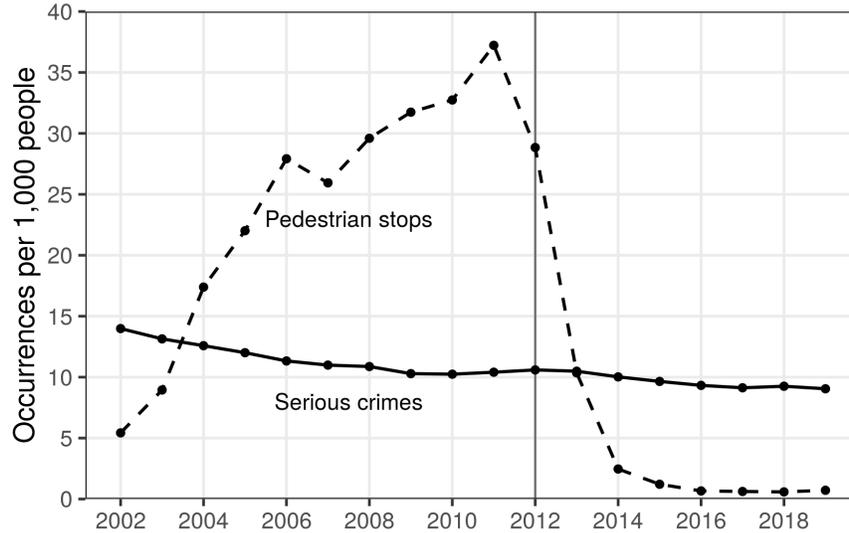
Our analysis above looked only at stop-and-frisk's immediate goal of getting weapons off the streets. More broadly, however, New York City's stop-and-frisk policy sought to reduce serious crime, such as gun violence.⁷¹ We conclude here by directly examining this downstream outcome. Mirroring our long-term analysis in Nashville, in Figure 5 we plot the incidence of serious crime from 2002 to 2019 against the number of pedestrian stops over the same time period. In 2012, in response to federal litigation and a changing political landscape, the NYPD began dramatically curtailing its use of stop-and-frisk.⁷² By 2016, the city reported

⁷¹ See Goel et al., *supra* note 62, at 365.

⁷² See Michael D. White, Henry F. Fradella, Weston J. Morrow & Doug Mellom, *Federal Civil Litigation as an Instrument of Police Reform: A Natural Experiment Exploring the*

12,404 stops, a 98% reduction from its peak in 2011. All the while, Figure 5 shows, serious crime in the city held steady.

FIGURE 5: RELATIONSHIP BETWEEN PEDESTRIAN STOPS AND CRIME IN NEW YORK CITY, 2002–2019



* * *

The empirical evidence suggests that both the MNPDP's intensive use of traffic stops and the NYPD's use of stop-and-frisk were generally unnecessary and had racially disproportionate costs. They were unnecessary because they yielded little apparent reduction in crime. Further, in New York, a more tailored policy could have resulted in similar weapon recovery with substantially lower fiscal and social costs. They were discriminatory in the sense that their burdens fell disproportionately on minority communities.

The approaches used here are ways of identifying a lower bound of unnecessary and discriminatory policing: they do not identify policies that are associated with moderate crime-suppression benefits while imposing high costs. That is, they do not capture all instances of unnecessary and discriminatory policing. Nevertheless, we believe that there is value to even an underinclusive analytic tool given the present absence of more robust alternatives.

Effects of the Floyd Ruling on Stop-and-Frisk Activities in New York City, 14 OHIO ST. J. CRIM. L. 9, 52 (2016).

III. IDENTIFYING POLICIES THAT DISPARATELY IMPACT MINORITY GROUPS

In Part II, we applied CBA to identify policies that burdened racial and ethnic minorities without providing accompanying social benefits. Here, we describe two complementary approaches to directly examine the disparate impacts of policies (rather than considering their broader costs and benefits). We illustrate these approaches in the context of police frisks. But both of the following strategies can be applied more generally to one's choice of policing tactics.

A central challenge in assessing disparities in policing involves the baseline task of identifying similarly situated groups of racial minority individuals and white individuals subject to policing under the same conditions. How, that is, does one appropriately compare officer decisions across race groups when the context of those interactions may also vary across groups?⁷³

Consider the use of a frisk after a *Terry* stop has occurred. Suppose that in some city, police frisk stopped Black individuals more frequently than they frisk stopped white individuals. This pattern may reflect an unjustified racial disparity in officer frisk decisions (whether intentional or not). Alternatively, it may reflect a justified disparity if the stopped white individuals in this city generally pose a lower risk to officer safety than stopped Black individuals. This latter hypothesis need not rest on race-based premises: imagine (again by way of illustration) a city in which Black residents had historically experienced neglect and discrimination and, as a result, experienced higher rates of poverty and thus more favorable conditions for crime.

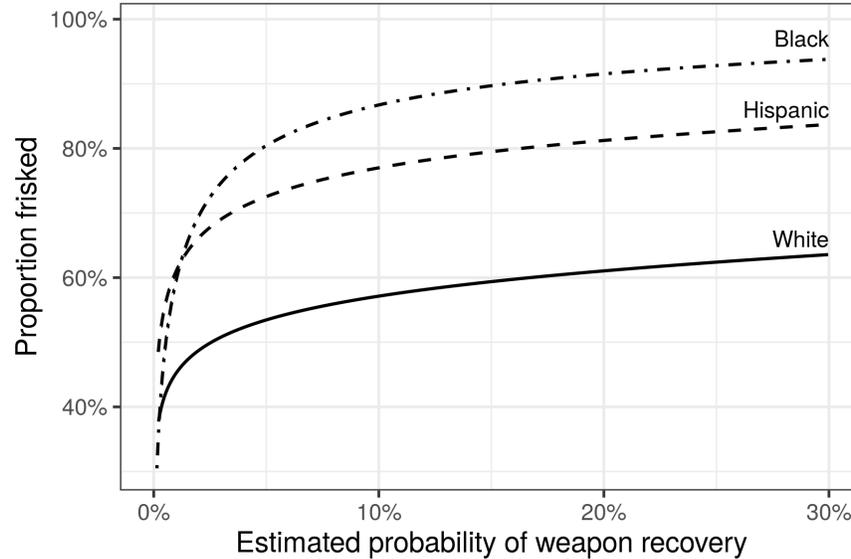
To disentangle these two possibilities, we describe and apply a strategy to measure the extent to which racial disparities among frisked individuals are justified by differences in

⁷³ One complication is that police do not typically record information on individuals whom they potentially could have stopped but ultimately did not. To account for this data gap, here we focus on poststop police coercion that is conditional on an encounter arising and being recorded. Recent work has raised challenges to the estimation of discrimination in post-stop police actions. See generally Dean Knox, Will Lowe & Jonathan Mummolo, *Administrative Records Mask Racially Biased Policing*, 114 AM. POL. SCI. REV. 619 (2020). However, these challenges concern the estimation of disparate treatment rather than disparate impact and, in many settings, are not insurmountable. See generally Johann Gaebler, William Cai, Guillaume Basse, Ravi Shroff, Sharad Goel & Jennifer Hill, *A Causal Framework for Observational Studies of Discrimination*, ARXIV (Oct. 7, 2021), <https://perma.cc/V5HZ-WSUN>.

individual “riskiness.” We call this “risk-adjusted regression.”⁷⁴ Here, we define the risk associated with an individual stop as the chance that a frisk, if conducted, would recover a weapon. To estimate this risk, we use the same statistical model described in Part II.B, trained on a historical data set of police frisks using all documented information that would be available to an officer before a frisk is conducted. Next, this trained statistical model is applied to new stops. It is used for each stop to estimate the likelihood that a frisk, if conducted, would yield a weapon—what we call a stop’s “risk score”—whether or not a frisk actually occurred. Finally, with these risk scores in hand, we compare differences in frisk rates between similarly risky stopped minority individuals and stopped white individuals. Under this approach, differences in frisk rates, or “risk-adjusted disparities,” can be interpreted as *unjustified* disparities—namely, those that are not explained by differences in observable threats to officer safety. Risk-adjusted regression hence offers a race-specific measure of excessive police action in the controversial context of street stops.

⁷⁴ Additional theoretical details of this approach can be found in Jung et al., *supra* note 19.

FIGURE 6: FRISK RATES OF STOPS VERSUS RISK BY RACE IN NEW YORK CITY, 2011–2012



In Figure 6, we display frisk rates (vertical axis) by estimated risk score (horizontal axis) for a data set of approximately 1.1 million stops of Black, Hispanic, and white individuals in New York City between 2011 and 2012. As in Part II.B, risk scores for each individual stop were calculated using a statistical model trained on the set of all frisks that took place between 2009 and 2010, using information recorded by the officer at the time the frisk occurred. Figure 6 shows the frequency of frisks in 2011–2012, disaggregated by race, as a function of this estimated risk. It provides a simple graphical representation of differences in frisk rates among similarly risky individuals of different races. In particular, Figure 6 demonstrates that, at any given level of risk, Black 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 in risk between stopped individuals of different race groups.

Risk-adjusted regression is a powerful strategy for identifying disparate police practices. Nevertheless, it is important to flag several limitations of the approach. To begin with, risk (e.g., of possessing a weapon) may not be the only legitimate factor informing officer 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 took public transit more often than white individuals (and were frisked there) observed risk-adjusted disparities may reflect this potentially justifiable difference rather than systemic discrimination. In the New York data, 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.⁷⁵

In addition, the estimated risk that we calculate using historical data may not fully account for the information in fact available to officers in the field. It is possible that officers can in practice 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. For example, an officer's estimate of risk may be legitimately affected by a stopped individual's response to questioning (which is not recorded in our data), and thus the officer's decision to frisk may be altered by that new and unrecorded information. If this were the case, given two individuals that appear identical on the attributes that exist in our data set, an officer may reasonably be able to identify and choose the riskier of these two individuals to frisk. In consequence, a statistical risk model trained only on the limited sample of frisked pedestrians—a necessary limitation, since frisks identify whether the pedestrian was carrying a weapon—may systematically overestimate risk for individuals who were *not* frisked, which would likely distort estimates of risk-adjusted disparities. As such, it is important to assess the sensitivity of results to varying forms and degrees of such unrecorded information. In the case of the New York data here, however, the large gaps in risk-adjusted frisk rates across race groups indicate that officers would need access to substantial unrecorded information to erase the apparent disparities. We think that unobserved information, therefore, is unlikely to explain these observed disparities.

Another distinct statistical strategy to measure the disparate impacts of policing decisions is to look at the success rates of such

⁷⁵ Officer perceptions of location-based risk, however, may themselves be inflected by race. See Ben Grunwald & Jeffrey Fagan, *The End of Intuition-Based High-Crime Areas*, 107 CALIF. L. REV. 345, 385–87 (2019).

decisions by race rather than the decision rates themselves.⁷⁶ Intuitively, the success rate of a police tactic when applied to individuals of a given race (assuming “success” makes conceptual sense for the tactic under consideration) may approximate the standard of evidence officers use when deciding whether to use that tactic. In particular, if in some jurisdiction weapons were found on frisked minorities less often than on frisked white individuals (i.e., if frisks of minorities were less successful than frisks of white individuals), it would suggest that minorities were frisked on the basis of less evidence. Such differences in success rates, also known as “hit rates,” may constitute evidence of the disparate impacts of police frisk practices in that jurisdiction.⁷⁷ This approach mitigates the problem of omitted variables, since one need only observe the outcome of a frisk decision, not the factors that prompted the frisk itself.⁷⁸

We illustrate this outcome-based approach by analyzing 102,118 records of investigatory stops of Black, Hispanic, and white adults from the Chicago Police Department (CPD) in 2017. In Chicago, following an investigative stop of a pedestrian, an electronic Investigative Stop Report is filled out for each individual under suspicion. These reports contain detailed information about the stopped individual, stop context, and stop disposition. Overall, we find that the rates at which frisks recover weapons are significantly lower for frisked Black individuals (3.8%) and Hispanic individuals (3.4%) compared to white individuals (5.7%).⁷⁹ By the logic of this approach, these results provide

⁷⁶ See, e.g., Gary S. Becker, *Nobel Lecture: The Economic Way of Looking at Behavior*, 101 J. POL. ECON. 385, 386, 401–03 (1993).

⁷⁷ See Ian Ayres, *Outcome Tests of Racial Disparities in Police Practices*, 4 JUST. RSCH. & POL'Y 131, 133 (2002).

⁷⁸ One limitation of outcome tests is that they may suffer from the problem of “infra-marginality,” a statistical phenomenon in which the stop success rate for a group may be a poor proxy for the actual level of evidence used by officers in the field. Newly developed “threshold tests” have recently leveraged advances in black-box Bayesian inference to address this issue of infra-marginality. See generally Emma Pierson, Sam Corbett-Davies & Sharad Goel, *Fast Threshold Tests for Detecting Discrimination*, 84 PROC. MACH. LEARNING RSCH. 96 (2018); Camelia Simoiu, Sam Corbett-Davies & Sharad Goel, *The Problem of Infra-marginality in Outcome Tests for Discrimination*, 11 ANNALS APPLIED STAT. 1193 (2017). In our case, outcome and threshold tests yield comparable results, and so, for simplicity, we restrict our analysis to the former.

⁷⁹ For this outcome-based analysis, we combine data on both consensual and non-consensual frisks in line with analysis of the Oakland Police Department’s stop practices. See REBECCA HETHEY, BENOIT MONIN, AMRITA MAITREYI & JENNIFER L. EBERHARDT, *DATA FOR CHANGE: A STATISTICAL ANALYSIS OF POLICE STOPS, SEARCHES, HANDCUFFINGS, AND ARRESTS IN OAKLAND, CALIF., 2013-2014*, at 15 (2016). We note, however, that we see

suggestive evidence that Chicago's practice of frisking stopped individuals had a disparate impact on minorities.

Using outcome-based approaches to measure discrimination in police activity can also deploy a granular geographic focus. As explained above, officers may have a legitimate reason to apply a lower standard of evidence when deciding to frisk individuals in some locations instead of others. In Figure 7, we display the success rate of frisks in each Chicago police district (vertical axis) against the racial composition of the district (horizontal axis), along with a trend line, for districts with over twenty-five frisks for each race group. We define the district racial composition as the proportion of the population that is nonwhite. Figure 7 makes apparent that districts with a greater share of minority residents had lower hit rates on average, suggesting that the bar for frisking individuals was lower in such areas.

qualitatively similar results on each subset of frisks, with lower hit rates for Black and Hispanic individuals relative to white individuals. Specifically, on the subset of nonconsensual frisks, hit rates are 5.9% for white individuals, 5.0% for Black individuals, and 4.0% for Hispanic individuals; and on the subset of consensual frisks, hit rates are 5.6% for white individuals, 2.6% for Black individuals, and 3.0% for Hispanic individuals.

FIGURE 7: HIT RATES VERSUS POLICE-DISTRICT RACIAL COMPOSITION IN CHICAGO

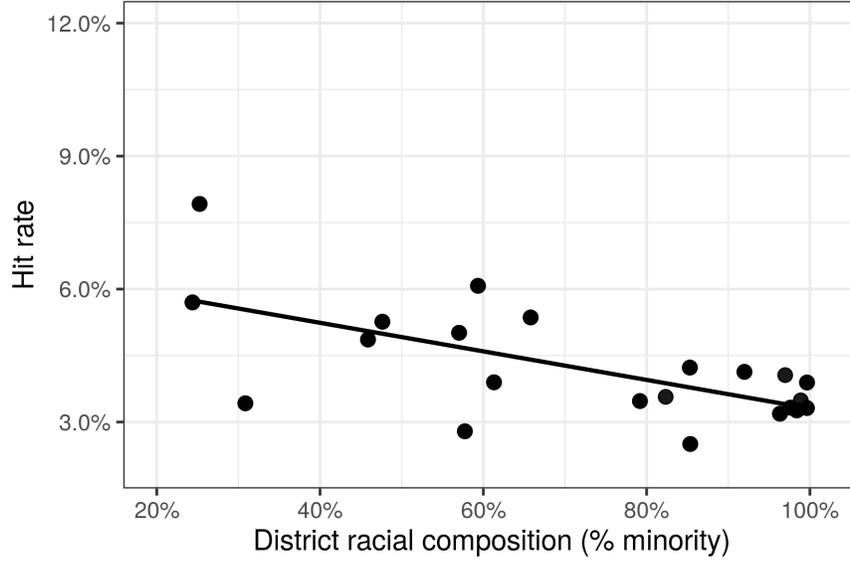
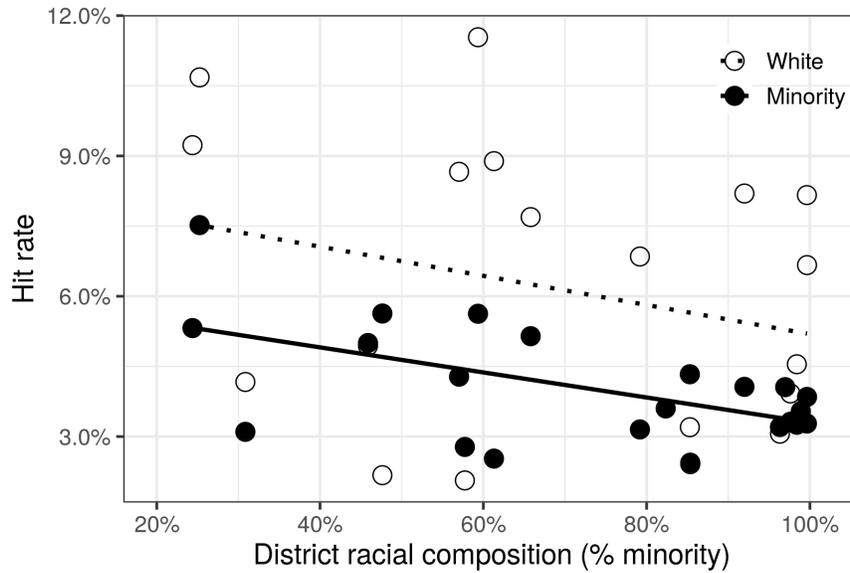


FIGURE 8: HIT RATES VERSUS POLICE-DISTRICT RACIAL COMPOSITION BY STOPPED PERSON'S RACE IN CHICAGO



In Figure 8, we show a similar relationship but now separately plot hit rates for white and minority residents within each

police district (open points/dotted trend line and solid points/solid trend line, respectively). Regardless of the racial composition of a district's population, hit rates within each district were generally lower for minorities than for white individuals. Minorities in predominantly minority neighborhoods were thus doubly impacted by Chicago police-frisk practices: first, because frisks in high-minority areas were, on average, carried out on the basis of less suspicion than frisks in predominantly white neighborhoods; and second, because within a given area minorities were, on average, frisked on the basis of less suspicion than white individuals.

We emphasize that our application of both risk-adjusted regression and outcome analysis differs from a traditional disparate-treatment analysis in that we do not attempt to account for the full set of factors that may potentially provide a race-neutral explanation for the observed differences.⁸⁰ Suppose, hypothetically, that (1) officers have a policy of frisking individuals entering or exiting public-housing complexes, regardless of an individual's race; (2) 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 weapons. In this scenario, differences in frisk rates across groups may not be driven by racial animus, but such a policy nonetheless would create a facially unjustified burden on racial minorities, and so it is a form of discriminatory impact.

IV. IDENTIFYING THE CIRCUMVENTION OF ANTIDISCRIMINATION MANDATES

As we have seen above, analyses of particular policing practices can yield important insights about unnecessary and discriminatory policing. But no policing practice exists in isolation. Policing tactics that are unnecessary or discriminatory may have close parallels elsewhere in the department, and communities harmed by policing feel the combined impacts of all these practices. Focusing exclusively on individual tactics can thus obscure the overall harms of a broader policing strategy. With this understanding in mind, we conclude by holistically analyzing related pedestrian- and traffic-stop practices, discussing how oversight narrowly focused on pedestrian stops in Chicago and Philadelphia

⁸⁰ See generally Ian Ayres, *Three Tests for Measuring Unjustified Disparate Impacts in Organ Transplantation: The Problem of "Included Variable" Bias*, 48 *PERSPS. BIOLOGY & MED.* S68 (Supp. 1 2005).

may have led the departments to use substitute means to enact similarly discriminatory tactics.⁸¹

A. Chicago

In 2015, the CPD entered into a settlement agreement with the ACLU of Illinois,⁸² following many years of local activism and the publication of a report by the ACLU of Illinois about the CPD's stop-and-frisk practices.⁸³ Subsequently, the CPD also came under a consent decree mandating broad changes related to an investigation of the department after Laquan McDonald, a Black seventeen-year-old, was killed by a CPD officer in 2014.⁸⁴

The settlement agreement with the ACLU took effect on January 1, 2016.⁸⁵ For 2016, the CPD reported a total of approximately 100,000 pedestrian stops, a sharp drop from the roughly 600,000 stops reported for 2015 (Figure 9).⁸⁶ At the same time, the number of traffic stops made by the CPD began to rise. The CPD reported around 100,000 traffic stops in 2014 and a similar amount in 2015, but by 2019, the CPD reported nearly 600,000 traffic stops, with large increases occurring each year from 2016 to 2019.⁸⁷

These traffic stops came to closely resemble the pedestrian stops that the CPD was contemporaneously under pressure to curtail. For example, in 2014, most of the CPD's traffic stops were

⁸¹ For a similar analysis and conclusion, see generally David Hausman & Dorothy Kronick, *When Police Sabotage Reform by Switching Tactics* (Feb. 16, 2021) (unpublished manuscript), <https://perma.cc/8SU5-8LTS>.

⁸² See generally *Investigatory Stop and Protective Pat Down Settlement Agreement*, ACLU ILL. (Aug. 6, 2015), <https://perma.cc/BZ76-7U8U>.

⁸³ See generally ACLU OF ILL., *STOP AND FRISK IN CHICAGO* (2015), <https://perma.cc/3PMN-4Q53>.

⁸⁴ See generally *Consent Decree, Illinois v. City of Chicago*, No. 17-cv-6260 (N.D. Ill. Jan. 31, 2019).

⁸⁵ ARLANDER KEYS, *THE CONSULTANT'S FIRST SEMIANNUAL REPORT ON THE INVESTIGATORY STOP AND PROTECTIVE PAT DOWN AGREEMENT FOR THE PERIOD JANUARY 1, 2016 – JUNE 30, 2016*, at 6 (2017), <https://perma.cc/S8RM-D6A6>.

⁸⁶ For the Illinois traffic- and pedestrian-stop data from 2004 through 2020, see *Illinois Traffic and Pedestrian Stop Study*, ILL. DEP'T TRANSP., <https://perma.cc/ZY2Z-DUU4>. Before 2016, the CPD recorded only stop-and-frisk stops where the subject of the stop was released without further enforcement action, so the actual number of such stops in 2014 and 2015, while unavailable, is higher than the 2014 and 2015 numbers reflect. See KEYS, *supra* note 85, at 21.

⁸⁷ A recent ACLU of Illinois report highlighted the increase in traffic stops by the CPD—and associated racial disparities—from 2015 to 2017. See generally ACLU OF ILL., *RACISM IN THE REAR VIEW MIRROR, ILLINOIS TRAFFIC STOP DATA 2015-2017* (2019), <https://perma.cc/3ADH-WJWX>.

for moving violations.⁸⁸ But by 2019, our analysis of the data shows, most stops were for equipment, license plate, or registration violations. The shift was driven largely by an increase in stops for equipment violations, which comprised roughly 40% of traffic stops in 2019. This focus on nonmoving violations is reminiscent of Nashville's traffic-stop practices, and it suggests that traffic stops in Chicago became a pretext to search for evidence of unrelated criminal activity.

Further, though racial disparities in the CPD's traffic stops were evident in 2014—the beginning of the time period that we consider here—they grew significantly from 2014 to 2019. In 2019, Hispanic drivers were almost twice as likely to be stopped as white drivers relative to their share of the adult population, and Black drivers were more than five times as likely to be stopped as white drivers.⁸⁹ These growing racial disparities—alongside the general rise in traffic stops—yielded race-specific stop rates for traffic and pedestrian stops combined that roughly mirrored their pre-settlement-agreement numbers, as shown in Figure 10.

⁸⁸ *Illinois Traffic and Pedestrian Study*, *supra* note 86.

⁸⁹ Chuck Goudie, Barb Markoff, Christine Tressel, Ross Weidner, Jonathan Fagg, Yun Choi & Adriana Aguilar, *Chicago Police More Likely to Stop Black Drivers Without Citing Them, Data Investigation Reveals*, ABC 7 CHI. (Sept. 9, 2020), <https://perma.cc/7CLA-ALSE>.

FIGURE 9: YEARLY STOP COUNTS IN CHICAGO, 2014–2019

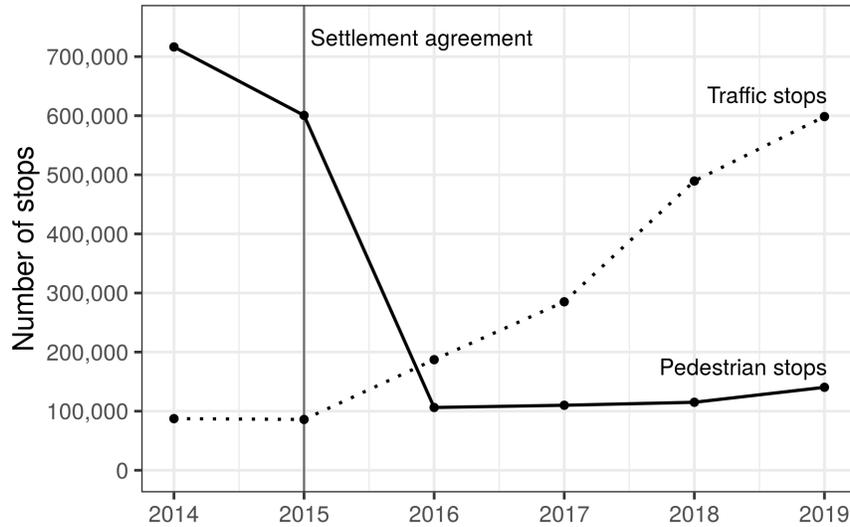
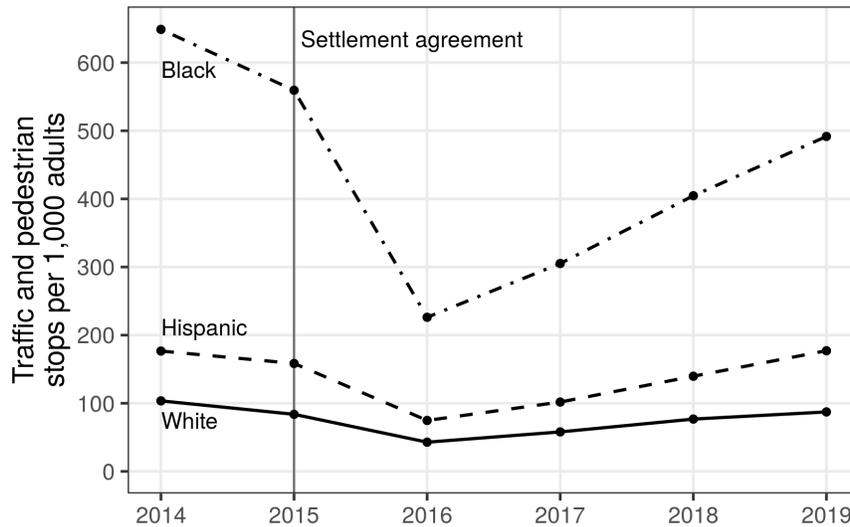


FIGURE 10: RACIAL DISPARITIES IN CHICAGO STOPS, 2014–2019



Poststop searches likewise exhibit racial disparities. Among drivers stopped by the CPD in 2014–2019, Black and Hispanic drivers were searched more than twice as often as white drivers.⁹⁰

⁹⁰ Officers also sometimes conduct dog sniffs or dog searches of vehicles. We do not discuss those types of searches here due to their infrequency; the data indicate that the

In addition, recovery rates of contraband, such as drugs, alcohol, weapons, or stolen property, were lower for searched Black and Hispanic drivers than for searched white drivers. This gap suggests that minority drivers were searched by the CPD on the basis of less evidence than white drivers (Table 1).⁹¹

B. Philadelphia

A similar story holds for the Philadelphia Police Department (PPD). Following a consent decree and settlement in 2011,⁹² pedestrian stops fell from more than 200,000 reported stops in 2014 (the earliest year for which we have data released publicly by the city) to fewer than 100,000 reported stops in each of 2018 and 2019, while traffic stops almost doubled in the same period, as shown in Figure 11.⁹³

We found similarities in the race, age, and gender distributions of stopped individuals for both traffic and pedestrian stops in Philadelphia; in particular, from 2014 to 2019, 49% of traffic stops and 61% of pedestrian stops were of Black men. As shown in Table 1, Black and Hispanic drivers experienced higher search rates and lower recovery rates of contraband.⁹⁴ In the long run, despite successful legal action to curb pedestrian stops, we find

CPD employed dog sniffs in approximately ten to twenty of these stops per year from 2014 to 2019, representing an extremely small proportion of all stops documented.

⁹¹ Broken out by year, contraband recovery rates were higher for white drivers than they were for Black drivers each year from 2014 to 2019 and were higher for white drivers than they were for Hispanic drivers in each year except 2019. In 2019, hit rates were 17% for Black drivers, 27% for Hispanic drivers, and 24% for white drivers.

⁹² See generally Settlement Agreement, Class Certification, and Consent Decree, *Bailey v. City of Philadelphia*, No. 10-5952 (E.D. Pa. June 21, 2011). Before *Bailey*, the PPD's stop-and-frisk practices were monitored for several years in the 1990s and 2000s through a settlement stemming from an earlier federal lawsuit. See *Case Profile: NAACP v. City of Philadelphia*, UNIV. MICH. L. SCH.: C.R. LITIG. CLEARINGHOUSE, <https://perma.cc/Gx5E-RM5J>.

⁹³ In our Philadelphia analysis, we include searches incident to arrest, as those cannot be separated out from other search types in the data we received.

⁹⁴ The PPD recorded information on the race of drivers differently than the CPD did; the PPD indicated drivers who are both Black and Hispanic while the CPD did not. To present results from Philadelphia using the same groups as those in the CPD data, we group Black Hispanic drivers with Black drivers. Stops of such drivers are relatively rare, comprising about 1.5% of all stops of Black drivers in Philadelphia. The rates presented in Table 1 do not meaningfully change if Black Hispanic drivers are reclassified as Hispanic drivers instead of Black drivers. Search rates for Black Hispanic drivers mirror those for both Black drivers and Hispanic drivers. In addition, contraband recovery rates for Black Hispanic drivers are slightly higher than contraband recovery rates for either Black or Hispanic drivers, with contraband recovered in 30% of searches for Black Hispanic drivers.

minimal change in the number of combined pedestrian and traffic stops per capita for each race group (Figure 12).

FIGURE 11: YEARLY STOP COUNTS IN PHILADELPHIA, 2014–2019

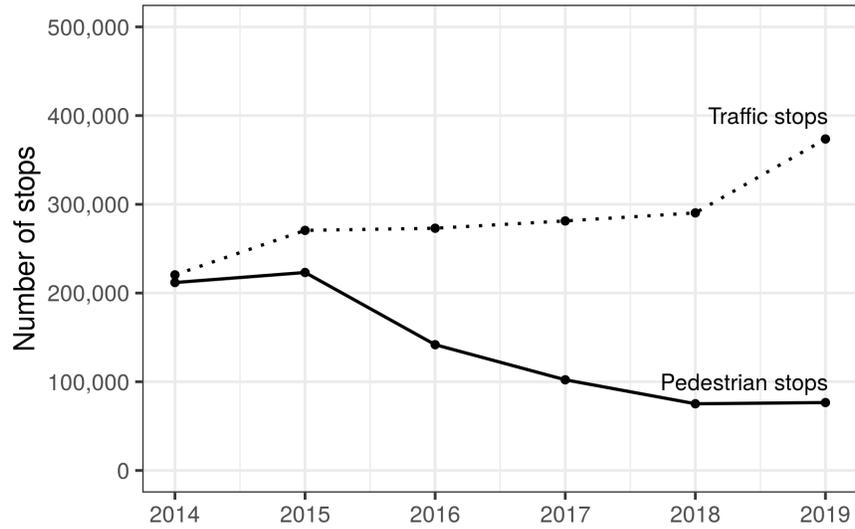


FIGURE 12: RACIAL DISPARITIES IN STOPS IN PHILADELPHIA, 2014–2019

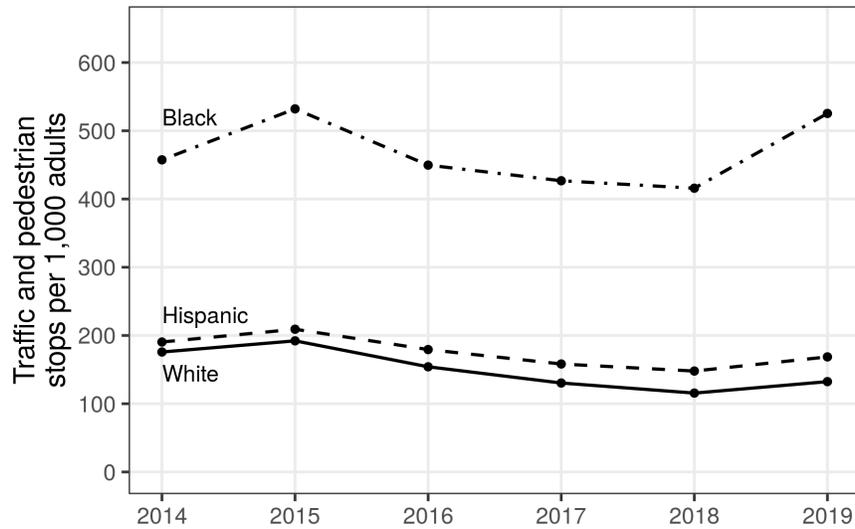


TABLE 1: SEARCH RATES AND CONTRABAND RECOVERY RATES BY RACE IN CHICAGO AND PHILADELPHIA, 2014–2019

	Chicago			Philadelphia		
	White	Black	Hispanic	White	Black	Hispanic
Stops	261,643	1,036,913	375,041	320,120	1,215,108	160,189
Searches	1,676	16,029	9,439	10,615	66,769	9,283
CR*	457	2,663	1,841	3,598	16,008	2,137
Search Rate						
	0.6%	1.5%	2.5%	3.3%	5.5%	5.8%
CRR**	27%	17%	20%	34%	24%	23%

* Contraband Recovered

** Contraband Recovery Rate

* * *

The cumulative evidence from both Chicago and Philadelphia thus tell similar stories: traffic stops, intentionally or not, took on the role of stop-and-frisk in the face of pressure to curb the latter. The CPD's traffic stops—which became increasingly racially disparate, rarely resulted in the recovery of contraband, and, by 2019, were largely made on the basis of nonmoving violations—came to function as a form of police coercion similar to stop-and-frisk. Similar data from Philadelphia suggest that this is not an isolated occurrence. The displacement of one form of coercion with another warrants further investigation in these and other cities. These examples illustrate the perils of overly narrow policy interventions and further underscore the need for holistic police reform.

CONCLUSION

Police departments have the data to evaluate both the efficacy of their tactics and the existence of unnecessary or racially disparate policy choices. We have demonstrated the potential both to identify tactics that have produced no social gain in terms of lower crime rates (as in Nashville and New York) and to isolate those policing measures for which it is possible to identify large quanta of surplus coercion being used against racial minorities (in New York and Chicago). The data also allow us to identify police efforts to circumvent legal interventions—as we see in both Chicago and Philadelphia. In an era in which the transformation of U.S. policing has seemed, at least briefly, in the cards, we conclude that the

importance of accurate and comprehensive data-driven analysis on policing—its benefits and its costs—cannot be understated.