2019

Using Shifts in Deployment and Operations to Test for Racial Bias in Police Stops

John M. MacDonald
University of Pennsylvania, johnmm@sas.upenn.edu

Jeffrey A. Fagan
Columbia Law School, jfagan@law.columbia.edu

AEA Papers and Proceedings 2019, 109: 148–151
https://doi.org/10.1257/pandp.20191027

Using Shifts in Deployment and Operations to Test for Racial Bias in Police Stops†

By John M. MacDonald and Jeffrey Fagan*

There is a contentious debate about the extent of racial bias in street and highway stops, the most common form of police-civilian contact. Compared to whites, blacks are more likely to be stopped, searched, and frisked by the police (Pierson et al. 2017). Many studies focus on comparing racial differences outcomes that transpire after a police stop (Ridgeway and MacDonald 2010, Neil and Winship 2019). Outcome tests of this form have a long history in the economics of discrimination literature (Becker 1957). Some scholars contend that conditional on a police stop, outcomes should be similar across race if the police are applying race-neutral standards (Knowles, Persico, and Todd 2001).

However, outcome tests are sensitive to omitted variable bias that may be correlated with the race of the individual stopped (Neil and Winship 2019). Infra-marginality presents an additional challenge (Ayres 2002; Simoiu, Corbett-Davies, and Goel 2017). The average outcome by racial group may be different from those at the margins of a stop outcome if there are racial differences in the underlying crime-suspect risk distributions. Proposed solutions to the infra-marginality problem include estimating outcomes from stops involving different officer-civilian race pairs, using a threshold test of searches for different racial groups, or estimating racial differences in recovery rates from searches that have similar ex ante probabilities of recovering weapons (Anwar and Fang 2009; Antonovics and Knight 2006; Goel, Rao, and Shroff 2016; Pierson et al. 2017). These methods all rely on identifying assumptions that are not directly or easily testable. Also, changes in the factors that determine differences at the margins of a police outcome are difficult to observe. As a result, research that uses outcomes tests from police stops to test for racial bias typically relies on cross-sectional variation.

In this paper, we exploit a policy experiment in the New York Police Department (NYPD) to test for bias in police stops. The NYPD launched Operation Impact in 2003 to change the scale of officer deployments. High crime areas were designated as "impact zones" and saturated with recent police academy graduates. These officers were encouraged to stop, question, and frisk (SQF) crime suspects as part of the NYPD's overall crime-reduction strategy (MacDonald, Fagan, and Geller 2016). We focus on the expansion of impact zones in Brooklyn and Queens in July 2007. We use geographic data on the boundaries of the impact zones and the specific locations of recorded SQF encounters to test for racial bias in the outcomes from police stops.[1] We use a difference-in-difference (D-D) framework that exploits time and place varying sources of variation in police incentives to stop criminal suspects. We combine the D-D identification with a doubly robust estimator to assure that similarly situated stops are compared in areas before and after impact zones were formed. If the police are not discriminating based on race of crime suspects, then changes in stop outcomes in areas affected by

* MacDonald: Department of Criminology, University of Pennsylvania, 3718 Locust Walk, McNeil Building, Suite 483, Philadelphia, PA 19104 (email: johnmm@upenn.edu); Fagan: Columbia Law School, Columbia University, Jerome Greene Hall Room 916, 435 West 116th Street, New York, NY 10027 (email: jaf45@columbia.edu). The opinions expressed in the article reflect those of the authors only and not any other entity. We are grateful to Greg Ridgeway, Stephen Ross, Jennifer Doleac, and Felipe Goncalves for comments.

† Go to https://doi.org/10.1257/pandp.20191027 to visit the article page for additional materials and author disclosure statement(s).

[1] This expansion was during impact-zone period 9. We exclude stops that were located in areas that were previously part of impact-zone period 8.

the impact-zone program should be proportional across racial groups relative to unaffected areas.

I. Empirical Analysis

A. *Data*

We obtained detailed information from the SQF database in NYC for 2007. These records include the date (month, day, year), time (hours), location (latitude, longitude), the crime suspected and suspicious-behavior officers noted, demographics of the individuals stopped, and the outcomes from each stop.[2] Outcomes include whether the stop resulted in an arrest, summons issued, frisk, search, placing hands on suspects, and making suspects stand against walls. We also examine whether any illegal contraband or weapons were recovered from individuals that were frisked or searched. All outcomes are binary indicators of whether $(=1)$ or not $(=0)$ it occurred as a consequence of a stop.

B. *Estimator of Racial Bias*

We rely on a potential outcomes framework and estimate the average treatment effect on treated impact-zone areas (ATT). The differences in outcomes from police stops can be expressed as a counterfactual comparison of individuals (i) who are stopped after the expansion of impact zones (denoted by $t=1$) to individuals of the same race or ethnicity that are stopped in the same areas before an impact zone was formed (denoted by $t=0$). We can identify the effect of impact-zone formation on racial bias in stop outcomes for individuals if we assume changes in stop outcomes in impact zones $(D=1)$ should be proportional to areas where impact zones were not formed $(D=0)$. This estimate then takes the form of a D-D estimator as in

$$\tau_D = E[Y_{it}(1,1)\,|\,D=1] - E[Y_{it}(1,0)\,|\,D=1] - E[Y_{it}(0,1)\,|\,D=0] - E[Y_{it}(0,0)\,|\,D=0].$$

To assure that estimates of each racial group's (τ) changes in stop outcomes after an impact zone forms are not biased due to changes in the observed characteristics of stops, we use entropy distance weighting to reweight the distribution of stop features from the pre-impact period to equivalent on the mean, variance, and skew in stops made in the post-impact-zone period. We combine this entropy-weighting procedure with a regression model that includes the full set of covariates so that estimates (τ) are doubly robust (Zhao and Percival 2016). This model yields estimates of the effect of being an impact zone relative to other areas of the city for blacks, Hispanics, and other racial groups.[3] We then compare the estimates of τ for each racial group, resulting in a difference-in-difference-in-difference (D-D-D) estimator.

II. Results

Table 1 shows the race-specific effects of the formation of impact zones in areas relative to unaffected areas of the city on all outcomes. For blacks, impact-zone formation increases arrests, summons, and frisks. For Hispanics, impact-zone formation increases arrests, frisks, and street detention (hands placed on walls). For other races, impact-zone formation does not significantly change $(p < 0.01)$ the risk of any outcome.

Recovery rates in impact zones for weapons increases for blacks, but this difference is not significantly greater than areas that don't receive impact zones. Recovery rates for other contraband are unaffected by impact-zone formation.

[2] To measure the crime suspected, we include indicators $(1=\text{yes},\ 0=\text{no})$ of whether a stop was for a suspected violent, weapons, property, drug, or other offense reason. To measure criminal suspicions, we included indicators $(1=\text{yes},\ 0=\text{no})$ noted on the SQF forms of whether a stopped individual was suspected of carrying an illegal object in plain view, fit a crime description, casing a place or victim, serving as a lookout for a crime, engaging in a drug transaction, exhibiting a furtive movement, observed committing a violent crime, had a suspicious bulge, or any other non-specified criminal suspicion. To measure the general context of stops, we also created indicators for whether $(=1)$ or not $(=0)$ the stop was the result of a radio call, the day of the week the stop occurred, the patrol shift (first, second, or third patrol), and a general age category of individuals stopped (e.g., under 16, 16–19, 20–24, 25–34, 35–64, and 65 or older).

[3] The breakdown of other racial groups is majority white (61 percent), Asians (17 percent), other (21 percent), or unknown (0.5 percent) to police officers.

Table 1—Outcomes from Similarly Situated Stops in Impact-Zone Areas and Other Areas

	Arrested	Summons	Frisked	Searched	Hands	Wall/Car	Contraband	Weapon
Blacks								
Impact areas	1.815	1.235	1.453	1.245	1.224	0.839	1.464	2.202
	(0.198)	(0.084)	(0.088)	(0.111)	(0.090)	(0.127)	(0.258)	(0.530)
Mean before	0.022	0.086	0.544	0.059	0.174	0.019	0.013	0.005
Mean after	0.039	0.104	0.607	0.072	0.202	0.016	0.020	0.010
Observations	26,329	26,329	26,329	26,329	26,329	26,329	14,180	14,296
Other areas	1.154	1.013	1.090	1.050	1.056	1.023	1.010	1.654
	(0.034)	(0.032)	(0.024)	(0.031)	(0.031)	(0.051)	(0.0498)	(0.137)
Mean before	0.061	0.069	0.553	0.091	0.216	0.037	0.035	0.008
Mean after	0.069	0.070	0.568	0.095	0.225	0.037	0.036	0.013
Observations	174,876	174,876	174,876	174,876	174,876	174,876	95,285	96,234
τ	3.29	2.44	3.97	1.65	1.76	−1.35	1.72	1.00
Hispanics								
Impact areas	1.970	1.124	1.787	1.321	1.691	1.212	1.214	1.900
	(0.341)	(0.120)	(0.165)	(0.176)	(0.177)	(0.262)	(0.349)	(0.744)
Mean before	0.025	0.087	0.520	0.073	0.164	0.025	0.019	0.006
Mean after	0.049	0.096	0.621	0.093	0.240	0.049	0.023	0.012
Observations	7,425	7,451	7,451	7,425	7,425	7,451	4,020	4,050
Other areas	1.076	0.994	1.082	1.060	1.106	1.001	0.991	1.896
	(0.037)	(0.037)	(0.025)	(0.035)	(0.034)	(0.053)	(0.0587)	(0.207)
Mean before	0.065	0.074	0.559	0.096	0.209	0.044	0.034	0.007
Mean after	0.069	0.074	0.579	0.102	0.226	0.044	0.033	0.013
Observations	111,633	111,633	111,633	111,633	111,633	111,633	61,519	62,116
τ	2.61	1.03	4.22	1.45	3.24	0.79	0.63	0.01
Other races								
Impact areas	2.516	1.131	1.403	1.389	1.382	1.075	2.501	1.353
	(0.755)	(0.171)	(0.155)	(0.253)	(0.186)	(0.316)	(1.143)	(1.123)
Mean before	0.019	0.074	0.520	0.064	0.166	0.018	0.009	0.003
Mean after	0.044	0.085	0.567	0.087	0.212	0.019	0.022	0.004
Observations	3,214	3,116	3,229	3,214	3,229	3,214	1,639	1,656
Other areas	0.946	0.899	1.069	0.969	1.158	0.978	0.948	1.557
	(0.039)	(0.044)	(0.034)	(0.041)	(0.047)	(0.060)	(0.064)	(0.185)
Mean before	0.062	0.072	0.417	0.081	0.155	0.030	0.045	0.009
Mean after	0.059	0.065	0.430	0.080	0.173	0.029	0.043	0.014
Observations	77,282	77,282	77,284	77,282	77,282	77,282	32,018	32,316
τ	2.07	2.33	2.10	1.63	1.16	0.30	1.35	−0.16

Notes: This table reports exponentiated coefficients. Standard errors are in parentheses and clustered on officer ID. Additionally, $\tau =$ difference-in-difference estimates for each racial group. All estimates also control for radio call, day of week, patrol shift, crime suspected, criminal-suspicion factors, and suspect age. The conditional mean for each outcome is displayed in the period before and after impact-zone expansion. The effective sample size is different from reported observations due to weighting.

III. Discussion

The results suggest that intensifying SQF policy in specific areas of New York City leads to racially disparate frisks of blacks and Hispanics. The absence of effects on recovery rates for weapons and other contraband suggests that the police did not apply different standards in searches in impact zones compared to other areas. Even though there was no racial disparity in the change in recovery rates, the increase in stops of nonwhites implies that the burden of this policy shift occurred primarily for blacks and Hispanics in impact-zone areas. Unproductive frisks and searches from stops could be the basis for the claim of a disparate impact (Manski and Nagin 2017; Gelman, Fagan, and Kiss 2007).

This study is limited in several ways. First, the analysis relies on a policy experiment as our identification strategy to solve the problem of omitted variable bias and infra-marginality in using outcomes tests to estimate racial bias in police stops. If the NYPD were uniformly biased in their SQF activities across all areas, then impact zones do not provide useful variation to estimate biased policing. Second, the study design assumes that the decision to designate areas as impact zones was caused by the policy and there were no other important factors that changed incentives for police officers to change SQF activities in these areas at the same time.

Future research should explore the productivity of searches from policy experiments by estimating whether changes in police policies produces racial disparities in search thresholds and recovery rates from police stops.

REFERENCES

Antonovics, Kate, and Brian G. Knight. 2009. "A New Look at Racial Profiling: Evidence from the Boston Police Department." *Review of Economics and Statistics* 91 (1): 163–77.

Anwar, Shamena, and Hanming Fang. 2006. "An Alternative Test of Racial Prejudice in Motor Vehicle Searches: Theory and Evidence." *American Economic Review* 96 (1): 127–51.

Ayres, Ian. 2002. "Outcome Tests of Racial Disparities in Police Practices." *Justice Research and Policy* 4 (1–2): 131–42.

Becker, Gary S. 1957. *The Economics of Discrimination.* Chicago: University of Chicago Press.

Gelman, Andrew, Jeffrey Fagan, and Alex 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): 813–23.

Goel, Sharad, Justin M. Rao, and Ravi Shroff. 2016. "Precinct or Prejudice? Understanding Racial Disparities in New York City's Stop-and-Frisk Policy." *Annals of Applied Statistics* 10 (1): 365–94.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1): 203–29.

MacDonald, John, Jeffrey Fagan, and Amanda Geller. 2016. "The Effects of Local Police Surges on Crime and Arrests in New York City." *PLoS ONE* 11 (6): e0157223.

Manski, Charles F., and Daniel S. Nagin. 2017. "Assessing Benefits, Costs, and Disparate Racial Impacts of Confrontational Proactive Policing." *Proceedings of the National Academy of Sciences* 114 (35): 9308–13.

Neil, Roland, and Christopher Winship. 2019. "Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing." *Annual Review of Criminology* 2: 73–98.

Pierson, Emma, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Vignesh Ramachandran, Cheryl Phillips, and Sharad Goel. 2017. "A Large-Scale Analysis of Racial Disparities in Police Stops across the United States." Unpublished.

Ridgeway, Greg, and John MacDonald. 2010. "Methods for Assessing Racially Biased Policing." In *Race, Ethnicity, and Policing: New and Essential Readings*, edited by Stephen K. Rice and Michael D. White, 180–204. New York: NYU Press.

Simoiu, Camelia, Sam Corbett-Davies, and Sharad Goel. 2017. "The Problem of Infra-marginality in Outcome Tests for Discrimination." *Annals of Applied Statistics* 11 (3): 1193–1216.

Zhao, Qingyuan, and Daniel Percival. 2016. "Entropy Balancing Is Doubly Robust." Unpublished.