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

John MacDonald
University of Pennsylvania

Jeffrey Fagan
Columbia University

1/4/2019

This article develops a new outcomes test for identifying racial bias in the context of police stops. Traditional tests predict that outcomes from police stops should be similar for minorities and whites if police are applying race neutral standards. We rely on a policy experiment in the New York City Police Department (NYPD) to address the well-known problems of omitted variable bias and infra-marginality in traditional outcomes tests of racial bias in police stops. The NYPD over the course of multiple years designated specific areas of the City as impact zones, and deployed extra officers to these areas and encouraged them to engage in more intensive stop, question, and frisk activity. We use a difference-in-differences design and doubly robust estimation to estimate racial bias in outcomes from stops. The results indicate that police are more likely to arrest, summons, and frisk Black and Hispanic suspects after an area becomes an impact zone. The results provide suggestive evidence of racial bias driven by declaring an area an impact zone and directing the police to increase enforcement activities.

1 INTRODUCTION

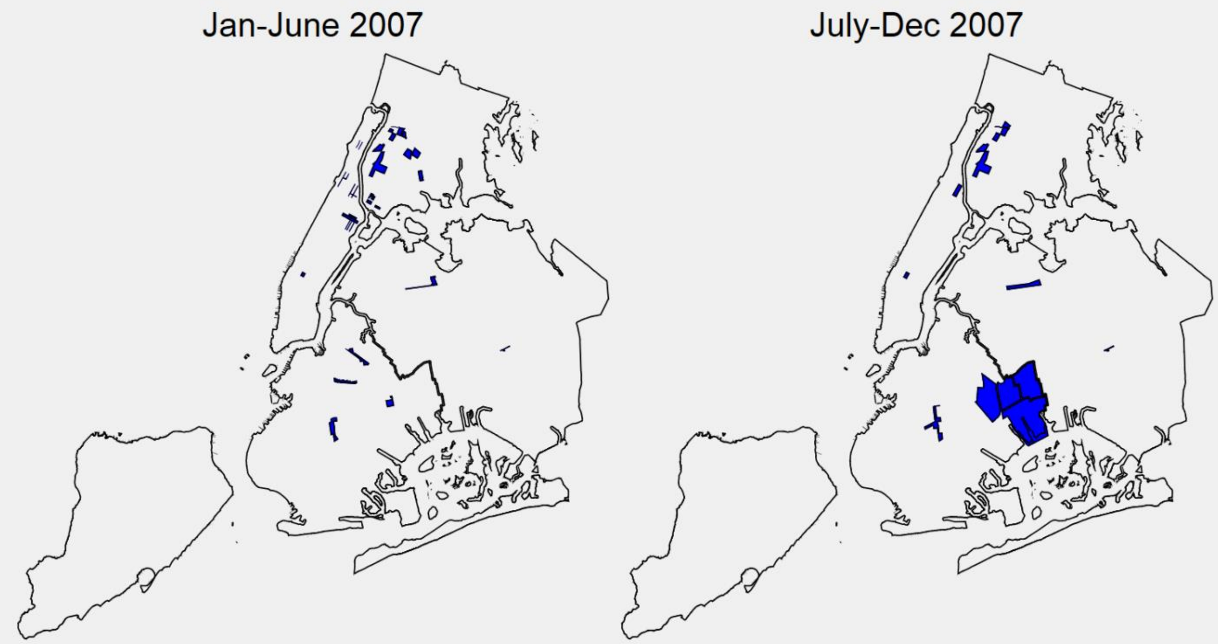
Racial disparities exist at every stage of criminal justice contact in the U.S. Compared to whites, blacks are more likely to stopped, searched, and frisked by the police (Pierson, et al. 2017). A large number of studies attempt to estimate whether the police discriminate based on race when deciding whether to stop, question, and frisk criminal suspects (Gelman, Fagan and Kiss 2007, Coviello and Persico 2015, G. Ridgeway 2006). Many studies focus on comparing racial differences outcomes that transpire after a police stop (Ridgeway and MacDonald 2010, Neil and Winship 2018). This form of an outcome test has a long history in the economics of discrimination literature (Becker 1957). Outcome tests from police stops often include examining racial differences in frisks, searches, arrests, and recovery rates of illegal contraband (Knowles, Persico and Todd 2001, G. Ridgeway 2006, Goel, Rao and Shroff 2016). Some scholars contend that conditional on a stop by the police, the probability of outcomes from those stops (e.g., search, arrest, finding of contraband or weapons) should look similar across race if the police are applying race neutral standards (Coviello and Persico 2015).

However, research shows that outcome tests are sensitive to omitted variable bias that may be correlated with the race of the individual stopped, such as the severity of the crime for which someone was stopped, the intensity of policing in the location, or dynamics of the police-citizen interaction (Dixon, et al. 2008, Neil and Winship 2018). Infra-marginality presents an additional challenge with using outcome tests to estimate racial discrimination by the police in their decision to stop citizens (Ayres 2002, Simoiu, Corbett-Davies and Goel 2017). Comparisons of mean differences from a police stop by race may differ from comparison of the marginal outcomes when racial groups have different underlying crime suspect risk distributions. Several papers show that after controlling for circumstances of searches, such as the location of a stop, the identifying assumptions of the traditional outcome tests of police stops are fragile (Anwar and Fang 2006, Sanga 2009).

Solutions to the infra-marginality problem have been proposed, including comparing officer-civilian race pairs (Anwar and Fang 2006). Others have suggested that controlling for the contexts of police stops (e.g., reason for the stop, the time, and location) allows one to still use outcome tests as an indicator of racial profiling (Coviello and Persico 2015, Goel, Rao and Shroff 2016). Another potential solution to addressing infra-marginality would be to predict stop outcomes should be across the distribution of police stops for blacks and whites, as has been done in wage discrimination (Charles and Guryan 2008). But it is not possible to forecast what the distribution of stop outcomes should be by race. Another challenge to addressing infra-marginality is that the factors that determine differences at the margins of who are going to be frisked, searched, and arrested by the police typically cannot be observed changing over time. As a result, research on racial discrimination in outcomes tests from police stops typically have to rely on cross-sectional variation.

In this paper we address this limitation in outcomes tests by taking advantage of a policy experiment in the New York Police Department (NYPD). The NYPD launched Operation Impact in 2003 as part of its major change to officer deployment. The program involved designating high crime areas as "impact zones" and saturating these areas with police academy graduates during their first year or two on the police force. Once deployed to these areas officers were encouraged to conduct investigatory street stops and to enforce misdemeanor laws as part of the NYPD's approach to reducing serious street crime (MacDonald, Fagan and Geller 2016). We take advantage of the major expansion of impact zones in Brooklyn and Queens in July 2007 shown in Figure 1. We use geographic data on the locations of the impact zones and the precise location of recorded stops, questions, and frisks (SQFs) to assess the effect that having area declared an impact zone had on racial disparities in outcomes from police stops. We use a difference-in-differences framework that exploits time and place varying sources of identification in police incentives to stop criminal suspects. We combine the difference in differences identification with a doubly robust estimator to assure that similarly situated stops are being compared in areas before and after impact zones were formed. If the police are not discriminating based on race of crime suspects, then areas affected by the impact zone program should see similar relative changes in outcomes from stops across racial groups.

Figure 1. Location of Impact Zones After Expansion



2 A Model to Address Infra-marginality

A simple model can explain the basic problem of infra-marginality in an outcome test of police searches for contraband, and the benefit of a policy experiment for addressing this identification problem. If the probability of having a contraband (denoted by *C*) is 0.05 for blacks (denoted by *B*) and 0.15 for whites (denoted by *W*), it is possible for police to appear racially biased for the average case even if they are race neutral. Assume that carrying contraband increases the risk of being searched (*S*) by a factor of 2 regardless of race [$\frac{P(S|C,B)}{P(S|B)} = \frac{P(S|C,W)}{P(S|W)} = 2$]. Then the rate at which police recover contraband would signal race bias, as the recovery rate would be 0.10 for blacks [$P(C|S,B) = \frac{P(S|C,B)P(C|B)}{P(S|B)} = 0.05 \times 2 = 0.10$] and 0.30 for whites [$P(C|S,W) = \frac{P(S|C,W)P(C|W)}{P(S|W)} = 0.15 \times 2 = 0.30$]. If instead, police are 3 times more likely to search blacks than whites with contraband, the recovery rate would be 0.15 for blacks [$P(C|S,B) = 0.05 \times 3 = 0.15$] and 0.15 for whites [$P(C|S,W) = 0.15 \times 1 = 0.15$], suggesting no evidence of racial bias. This example violates a key assumption of the standard outcome test, "that the probability of being guilty is equal among all groups that are searched in equilibrium" (Knowles, Persico and Todd 2001, 215). For outcomes tests to be a valid test of racial bias one has to establish that marginal cases of outcomes from stops are similar between racial groups. Unobservable variables correlated with race make it difficult to establish stops at the margin of a search by the police or any other outcome are similarly situated between whites and blacks.

One approach to solving this identification challenge is to rely on a policy experiment that shifts the incentives of police officers to engage in stop, question, and frisk activities. Assume that the probability of being searched increases (or reduces) after a policy (denoted by τ) starts in a police department that encourages officers to more aggressively stop, question, and frisk suspects. We can then solve for the infra-marginality problem by the identity τ, so that during a race neutral policy regime the change (Δ) in the recovery rate for illegal contraband should be equal for blacks and whites regardless of whether carry rates differ by race, as in:

$$\Delta_B - \Delta_w = \left[\frac{P(S|C,B)P(C|B)}{P(S|B)} - \frac{P(S|C,B)P(C|B)\tau_B}{P(S|B)}\right] - \left[\frac{P(S|C,W)P(C|W)}{P(S|W)} - \frac{P(S|C,W)P(C|W)\tau_W}{P(S|W)}\right] \quad (1)$$

As long as we assume that $\tau \neq 0$ one can identify the effect of policy on racial bias in police searches. If $\tau_B \neq \tau_w$ then there is evidence of racial bias in the policy. If the policy causes $\tau_B = \tau_w$ then the effect will be a race neutral policy. For example if τ =1.5 for blacks and whites, the recovery rate will shift from .10 to .15[1] for blacks and from .30 to .45 for whites.[2] If the police were initially more likely to search blacks than whites with contraband, the race neutral policy change would still lead a 1.5 factor increase in search for all groups. On the other hand, if the policy results in racial bias in police search behavior and $\tau_B \neq \tau_W$, then the change (Δ) in recovery rate from searches will differ by group.[3] We rely on this basic approach to our identification, by taking advantage of the NYPD's impact zone program.

3 EMPIRICAL ANALYSIS

3.1 Data

We obtained detailed information from the stop, question, and frisk (SQF) database in NYC for 2007. From these data we have the date (month, day, year), time (hours), location (latitude-longitude), the crime suspected and suspicious behavior officers noted, demographics of the individuals stopped, an anonymized officer identifier, and the outcomes from each stop.

To measure outcomes we focus on whether the stop resulted in an arrest, summons issued, frisk, search, placing hands on suspects (pf_hands), and making suspects stand against walls (pf_walls). We also examine whether any illegal contraband or weapons were recovered from individuals that were frisked or searched. All outcomes reflect binary indicators of whether (=1) or not (=0) it occurred as a consequence of a stop.

---

1 $\frac{P(S|C,B)P(C|B)\tau}{P(S|B)} = 0.05 \times 2 \times 1.5 = 0.15$

2 $\frac{P(S|C,W)P(C|W)\tau}{P(S|B)} = 0.15 \times 2 \times 1.5 = 0.10$

3 Assume that the policy (τ) increases searches by 2 for blacks and by 1.5 for whites, independent of their propensity to carry contraband, then the recovery rate for blacks shifts 0.10 to 0.2, as $\frac{P(S|C,B)P(C|B)\tau}{P(S|B)} = 0.05 \times 2 \times 2 = 0.2$. By contrast the recovery rate for whites shifts from 0.30 to 0.45.

To measure stop features we include a number of measures, including a classification of suspect descriptions represented by whether (=1) or not (=0) the stop was for a suspected violent, weapons, property, drug, or other offense reason. We also include a set of crime suspicions (cs) measures noted on the SQF forms, representing whether (=1) or not (=0) an 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 general context of stops, we also created indicators for whether (=1) or not (=0) the stopped was the result of a radio call, the day of the week the stop occurred, the patrol shift (1, 2, or 3rd patrol), and a general age category of individuals stopped (e.g., under 16, 16-19, 20-24, 25-34, 35-64, 65 or older).

3.2 Estimator for Racial Bias

To estimate racial bias in outcomes from stops 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 (denoted by *Y*) 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 estimate the effect of impact zone formation on racial bias in stop outcomes for individuals, if we assume changes in stop outcomes in impact zones (denoted by *D*=1) should be proportional to areas that did not have impact zones formed (denoted by *D*=0). This estimate then takes the form of a difference-in-differences according to the following form:

$$\tau_D = \mathrm{E}[Yit(1{,}1)|D = 1] - \mathrm{E}[Yit(1{,}0)|D = 1] \text{-} \mathrm{E}[Yit(0{,}1)|D = 0] - \mathrm{E}[Yit(0{,}0)|D = 0] \quad (2)$$

To assure that estimates of (τ) each racial group's changes outcomes from stops after the formation of impact zones are not biased due to changes in the observed characteristics of stops, we reweight stops from the pre-impact period to be identical on observed stop features. Weights were chosen using a reweighing scheme that minimizes an entropy distance metric between an estimated weight (w_i) and a base weight (q_i) (Hainmueller 2012).

$$\min_{w_i} H(w) = \sum_{\{i|t=0\}} w_i \ \log w_i / q_i \quad (3)$$

The base weight begins with uniform set of weights from the stops of blacks, Hispanics, and other racial groups impact zone expansion (t=1) and then minimizes the entropy distance between estimated comparison weights(w_i) based on balancing and normalizing constraints such that the comparison group (t=0) of stops before impact zone expansion are equivalent on the mean, variance, or skew of the distribution of covariates (X_i). Weights are normalized to have a sum of 1 and cannot be negative.

We combine this entropy weighting procedure with a regression model that includes the set of covariates so that estimates are doubly robust, meaning that if either the estimates from a regression model or those from the entropy balancing weights are correct we will have an unbiased estimate of the impact of impact zone on stops (Zhao and Percival 2017). The final model we estimate takes the following form:

$$\hat{\tau}^{DR} = \sum_{j=1}^{p} \hat{\beta}\ (D=1) \left(\sum_{ti\,=\,0} w_i^{EB} c_j(X_i) - \frac{1}{n_1} \sum_{ti\,=\,1} c_j(X_i) \right) - \sum_{j=1}^{p} \hat{\beta}(D = 0) \left(\sum_{ti\,=\,0} w_i^{EB} c_j(X_i) - \frac{1}{n_1} \sum_{ti\,=\,1} c_j(X_i) \right) \quad (4)$$

From this model we obtain the estimated effect of being an impact zone relative to other areas of the city from the regression and the entropy weighting comparison. We can them compare the estimates of $\widehat{\tau}^{DR}$ for each racial group (blacks, Hispanics, and others[4]).

4 RESULTS

4.1 Doubly robust estimates

Table 1 shows the results from the estimates of effect of an impact zone formation relative to unaffected areas on each racial groups' stops outcomes (see Appendix A2 for unadjusted estimates). For blacks the formation of impact zones increases arrests, summons, and frisks. For Hispanics the formation of the impact zone increases arrests, frisks, and hands placed on walls. For Other races the formation of impact zones does not significantly change ($p<0.01$) the risk of any outcome. Impact zone formation increases frisks by 3.97 and 4.22 standard deviations for blacks and Hispanics compared to only 2 standard deviations for other racial groups, though these differences not statistically significant given the estimates for other racial groups has much less precision.

[4] The breakdown of other racial group is majority white (61%), Asians (17%), other (21%), or unknown (.5%) to police officers.

Table 1. Outcomes from Similarly Situated Stops Impact Zones and Other Areas

	Arrested	Summons	Frisked	Searched	Hands	Wall/Car
Blacks						
Impact	1.815**	1.235**	1.453**	1.245	1.224**	0.839
	(0.198)	(0.0846)	(0.0882)	(0.111)	(0.0904)	(0.127)
Before	0.0229	0.0862	0.544	0.0596	0.174	0.0194
After	0.0399	0.104	0.607	0.0726	0.202	0.0165
N=	26,329	26,329	26,329	26,329	26,329	26,329
Other	1.154**	1.013	1.090**	1.050	1.056	1.023
	(0.0347)	(0.0326)	(0.0240)	(0.0318)	(0.0310)	(0.0512)
Before	0.0617	0.0699	0.553	0.0913	0.216	0.0371
After	0.0698	0.0709	0.568	0.0952	0.225	0.0377
N=	174,876	174,876	174,876	174,876	174,876	174,876
τ	3.29**	2.44**	3.97**	1.65	1.76	-1.35
Hispanics						
Impact	1.970**	1.124	1.787**	1.321	1.691**	1.212
	(0.341)	(0.120)	(0.165)	(0.176)	(0.177)	(0.262)
Before	0.025	0.087	0.520	0.073	0.164	0.025
After	0.049	0.096	0.621	0.093	0.240	0.049
N=	7,425	7,451	7,451	7,425	7,425	7,451
Other	1.076	0.994	1.082**	1.060	1.106**	1.001
	(0.0373)	(0.0375)	(0.0254)	(0.0358)	(0.0341)	(0.0531)
Before	0.065	0.074	0.559	0.096	0.209	0.044
After	0.069	0.074	0.579	0.102	0.226	0.044
N=	111,633	111,633	111,633	111,633	111,633	111,633
τ	2.61**	1.03	4.22**	1.45	3.24**	0.79
Others						
Impact	2.516**	1.131	1.403**	1.389	2.516**	1.131
	(0.755)	(0.171)	(0.155)	(0.253)	(0.755)	(0.171)
Before	0.019	0.074	0.520	0.0644	0.166	0.018
After	0.044	0.085	0.567	0.0879	0.212	0.019
N=	3,214	3,116	3,229	3,214	3,229	3,214
Other	0.946	0.899	1.069	0.969	1.158**	0.978
	(0.0390)	(0.0440)	(0.0344)	(0.0411)	(0.0476)	(0.0603)
Before	0.062	0.072	0.417	0.0819	0.155	0.030
After	0.059	0.065	0.430	0.0800	0.173	0.029
N=	77,282	77,282	77,284	77,282	77,282	77,282
τ	2.07	2.33**	2.10	1.63	1.79	0.97

Exponentiated coefficients; Standard errors in parentheses and clustered on officer ID. τ=difference-in-differences estimates. All estimates also control for radio call, day of week, patrol shift, crime suspected, criminal suspicion factors, and suspect age. Marginal probabilities for each outcome are displayed in the period before and after impact zone expansion. Effective sample size different from reported N= due to weighting.
** $p < 0.01$

Table 2 shows the results for the estimates of changes in recovery rates from frisks or searches in stops conducted after an area receives an impact zone relative other areas of the city not affected by the policy. In areas the receive impact zones the recovery rates for weapons increases for blacks by a factor of 2.2, from just 0.5% to 1.0%. But this difference is not significantly greater than the areas that don’t receive impact zones, suggesting it is just part of a citywide trend. In general, impact zones don’t appear to have a material effect on hit rates from searchers or frisks.

Table 2. Hit Rates from Stops Impact Zones Compared to Other Areas

	(Black) Contraband	(Black) Weapon	(Hisp) Contraband	(Hisp) Weapon	(Other) Contraband	(Other) Weapon
Blacks						
Impact	1.464	2.202**	1.214	1.900	2.501	1.353
	(0.258)	(0.530)	(0.349)	(0.744)	(1.143)	(1.123)
Before	0.0138	0.00520	0.0194	0.00683	0.00926	0.00306
After	0.0203	0.0107	0.0236	0.0127	0.0228	0.00413
N=	14,180	14,296	4,020	4,050	1,639	1,656
Other	1.010	1.654**	0.991	1.896**	0.948	1.557**
	(0.0498)	(0.137)	(0.0587)	(0.207)	(0.0646)	(0.185)
Before	0.0359	0.00855	0.0342	0.00731	0.0453	0.00934
After	0.0360	0.0138	0.0339	0.0138	0.0430	0.0145
N=	95,285	96,234	61,519	62,116	32,018	32,316
τ	1.72	1.00	0.63	0.01	1.35	-.16

Exponentiated coefficients; Standard errors in parentheses and clustered on officer ID. τ=difference-in-differences estimates. All estimates also control for radio call, day of week, patrol shift, crime suspected, criminal suspicion factors, and suspect age. Marginal probabilities for each outcome are displayed in the period before and after impact zone expansion. Effective sample size different from reported N= due to weighting.
** $p < 0.01$

It is important to note that all estimates are comparing stops in the period before and after impact zone formation with stops that are on average statistically similar contexts, so these changes in outcomes are not the result of the changing composition of stops. When we examine unadjusted estimates, the results shown in Appendix A2 show larger effects, suggesting that stop context changes are important to control for in comparing diff-in-diff estimates.

5 CONCLUSIONS

The estimates presented in this study suggest that a more intensive police SQF policy in specific areas of New York City as a result of Operation Impact lead to racially disparate frisks of blacks and Hispanics. Recovery rates for illegal contraband and weapons did not change materially by race after impact zones are formed, suggesting that declaration of an area as an impact zone did not change racial disparities in propensity to search individuals in these areas. Even though there was no racial disparity in hit rates from searches and frisks among individuals stopped by the police as a result of Operation Impact, it is important to acknowledge that the burden of this policy shift occurred primarily for blacks and Hispanics in these 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 infra-marginality in using outcomes tests of racial bias in police stops. If, however, the NYPD were uniformly biased in their propensity to stop, search, and frisk across all areas then the impact zone formation does not provide any useful variation to estimate biased policing. The documentation on the policy and its emphasis on increased SQF activity does suggest that the program increased the incentives to make extra stops (MacDonald, Fagan and Geller 2016). Second, the identification strategy here like other difference-in-differences designs rests on the assumption that the decision to shift SQF tactics in the areas that became impact zones was exogenously caused by the policy. While the areas that became impact zones were dictated by police commanders and not patrol officers, we cannot know for certain if there were other circumstances that may have also changed incentives for patrol officers to stop citizens in these areas at the same time that impact zones were formed. Finally, while we can test for average differences in stop outcomes before and after impact zones formed for blacks, Hispanics, whites, and others relative to other areas of New York City, we cannot assign a true value of searches. Future research should explore the productivity of searches from policy experiments, by estimating whether policies produce differences in search thresholds and hit rates by race. This would be a useful extension on recent work jointly estimating the probability of searches and the risk distribution by racial groups (Simoiu, Corbett-Davies and Goel 2017).

Acknowledgements: We thank Greg Ridgeway for his helpful comments on the model and statistical methods. The data were made available under court order: Floyd et.al. v. City of New York et al., 08 Civ. 1034, Dkt # 66 (S.D.N.Y. June 9, 2009). The opinions expressed in the article reflect those of the authors only and not any other entity. The authors received no compensation for this work from outside parties Jeffrey Fagan was an expert witness for the plaintiffs in the case of Floyd et al. v. City of New York, 08 Civ. 1034 (S.D.N.Y.). John MacDonald currently serves on the federal monitoring committee on the settlement agreements in Floyd et al. v. City of New York, et al., 08 Civ. 1034 (AT), Ligon, et al., v. City of New York, et al., 12-CV-2274 (AT), and Davis et al., vs. City of New York, et al., 10-CV-00699 (AT). The opinions expressed in the article reflect those of the authors only and not any other entity.

6 References

Anwar, Shemena, and Hanming Fang. 2006. "An alternative test of racial prejudice in motor vehicle searches: Theory and evidence." *American Economic Review 96* 127-151.

Ayres, I. 2002. "Outcome tests of racial disparities in police practices." *Justice Research and Policy* 4 (1-2): 131-142.

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

Charles, Kerwin Kofi, and Jonathan Guryan. 2008. "Prejudice and wages: An emperican assessment of Becker's Economics of Discrimination." *Journal of Political Economy 116* 773-809.

Coviello, Decio, and Nicola Persico. 2015. "An Economic Analysis of the Black-White Disparities in the New York Police Department's Stop-and-Frisk Program." *Journal of Legal Studies 44* 315-360.

Dixon, Travis L., Terry L. Schell, Howard Giles, and Kristin L. Drogos. 2008. "The Influence of Race in Police–Civilian Interactions: A Content Analysis of Videotaped Interactions Taken During Cincinnati Police Traffic Stops." *Journal of Communication* 58: 530–549.

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* 813-823.

Goel, S., J. M. Rao, and R. Shroff. 2016. "Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy." *Annals of Applied Statistics* 10 (1): 365-394.

Hainmueller, Jens. 2012. "Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies." *Political Analysis 20* 35-46.

Knowles, J., N. Persico, and P. Todd. 2001. "Racial bias in motor vehicle searches: Theory and evidence." *Journal of Political Economy* 109 (1): 203-229.

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* 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)* 9308-9313.

Neil, Roland, and Christopher Winship. 2018. "Methodological Challenges and Opportunities in Testing for Racial Discrimination." *Annual Review of Criminology*.

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." *arXiv preprint arXiv:1706.05678.*

Ridgeway, G., and J. MacDonald. 2010. "Methods for assessing racially biased policing." In *Race, ethnicity, and policing: New and essential readings*, by Stephen K. Rice and Michael D. White, 180-204. New York: NYU Press.

Ridgeway, Gregory. 2006. "Assessing the effect of race bias in post-traffic stop outcomes using propensity scores." *Journal of Quantitative Criminology 22(1)* 22 (1): 1-29.

Sanga, Sarath. 2009. "Reconsidering racial bias in motor vehicle searches: Theory and evidence." *Journal of Political Economy, 117(6)* 1155-1159.

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. 2017. "Entropy blanacing is doubly robust." *Journal of Causal Inference 5* doi:10.1515/jci-2016-0010.

A1. Similarity of Stop Features Pre and Post Weighting on Impact Zone Formation

Impact	Black Post	Black Pre	Sdiff	Black W	Hisp Post	Hisp Pre	Sdiff	Hisp W	Other Post	Other Pre	Sdiff	Other W
stop4violence	0.174	0.171	0.006	0.173	0.236	0.206	0.070	0.236	0.203	0.209	-0.016	0.203
stop4weapons	0.423	0.315	0.218	0.422	0.354	0.255	0.206	0.354	0.341	0.265	0.160	0.340
stop4property	0.171	0.179	-0.023	0.171	0.246	0.285	-0.090	0.246	0.276	0.295	-0.042	0.276
stop4drug	0.097	0.107	-0.033	0.097	0.080	0.099	-0.071	0.080	0.104	0.105	-0.004	0.104
cs_objcs	0.031	0.036	-0.034	0.031	0.035	0.039	-0.021	0.035	0.044	0.037	0.036	0.044
cs_descr	0.145	0.134	0.033	0.145	0.149	0.154	-0.013	0.149	0.145	0.110	0.099	0.145
cs_casng	0.226	0.239	-0.031	0.226	0.291	0.291	0.000	0.291	0.261	0.294	-0.076	0.261
cs_lkout	0.153	0.165	-0.034	0.153	0.183	0.200	-0.043	0.183	0.186	0.206	-0.052	0.186
cs_cloth	0.058	0.050	0.033	0.058	0.066	0.055	0.045	0.066	0.057	0.044	0.054	0.057
cs_drgtr	0.085	0.098	-0.045	0.085	0.083	0.092	-0.031	0.083	0.096	0.108	-0.043	0.096
cs_furtv	0.510	0.459	0.102	0.510	0.526	0.475	0.103	0.526	0.473	0.441	0.063	0.473
cs_vcrim	0.073	0.078	-0.019	0.073	0.100	0.082	0.061	0.100	0.070	0.094	-0.092	0.070
cs_bulge	0.192	0.131	0.156	0.192	0.182	0.103	0.204	0.181	0.197	0.122	0.190	0.197
cs_other	0.179	0.252	-0.190	0.179	0.172	0.219	-0.127	0.172	0.179	0.222	-0.110	0.179
cs_probcause	0.345	0.373	-0.059	0.345	0.420	0.423	-0.005	0.420	0.389	0.452	-0.129	0.389

Male	0.910	0.912	-0.005	0.910	0.906	0.890	0.054	0.906	0.572	0.561	0.021	0.572
Radio	0.058	0.046	0.052	0.058	0.099	0.098	0.002	0.099	0.080	0.067	0.046	0.079
1.dow	0.109	0.107	0.007	0.109	0.108	0.117	-0.029	0.108	0.124	0.103	0.064	0.124
2.dow	0.124	0.127	-0.010	0.124	0.130	0.141	-0.034	0.130	0.117	0.134	-0.051	0.117
3.dow	0.139	0.126	0.037	0.139	0.135	0.130	0.013	0.135	0.141	0.136	0.014	0.141
4.dow	0.138	0.151	-0.038	0.138	0.144	0.151	-0.020	0.144	0.179	0.164	0.038	0.179
5.dow	0.154	0.172	-0.052	0.154	0.159	0.171	-0.033	0.159	0.143	0.170	-0.077	0.143
6.dow	0.179	0.174	0.012	0.179	0.167	0.164	0.009	0.167	0.160	0.167	-0.019	0.160
2.shift	0.201	0.201	0.000	0.201	0.229	0.251	-0.052	0.229	0.281	0.300	-0.043	0.281
3.shift	0.526	0.597	-0.143	0.526	0.476	0.555	-0.159	0.476	0.416	0.504	-0.178	0.416
3.agenew	0.043	0.047	-0.017	0.043	0.033	0.028	0.030	0.033	0.031	0.039	-0.049	0.031
4.agenew	0.317	0.301	0.035	0.317	0.320	0.284	0.077	0.320	0.303	0.265	0.082	0.303
5.agenew	0.193	0.185	0.019	0.193	0.207	0.185	0.054	0.207	0.171	0.183	-0.031	0.171
6.agenew	0.223	0.228	-0.012	0.223	0.234	0.254	-0.048	0.234	0.216	0.250	-0.082	0.216
7.agenew	0.219	0.233	-0.034	0.219	0.203	0.245	-0.106	0.203	0.273	0.260	0.031	0.273
Other Areas												
stop4violence	0.185	0.173	0.031	0.185	0.186	0.162	0.063	0.186	0.151	0.137	0.039	0.151

stop4weapons	0.279	0.256	0.053	0.279	0.237	0.215	0.052	0.237	0.132	0.121	0.032	0.132
stop4property	0.253	0.250	0.006	0.253	0.320	0.339	-0.041	0.320	0.479	0.498	-0.040	0.479
stop4drug	0.146	0.131	0.044	0.146	0.127	0.122	0.014	0.127	0.123	0.113	0.031	0.123
cs_objcs	0.027	0.027	0.006	0.027	0.033	0.033	-0.002	0.033	0.034	0.036	-0.008	0.034
cs_descr	0.192	0.175	0.043	0.192	0.193	0.168	0.062	0.193	0.209	0.198	0.027	0.209
cs_casng	0.252	0.248	0.008	0.252	0.279	0.284	-0.009	0.279	0.317	0.325	-0.018	0.317
cs_lkout	0.144	0.148	-0.010	0.144	0.167	0.174	-0.020	0.167	0.180	0.183	-0.006	0.180
cs_cloth	0.045	0.039	0.029	0.045	0.036	0.036	0.000	0.036	0.031	0.030	0.004	0.031
cs_drgtr	0.125	0.116	0.029	0.125	0.112	0.110	0.007	0.112	0.103	0.100	0.012	0.103
cs_furtv	0.443	0.414	0.058	0.443	0.423	0.393	0.062	0.423	0.386	0.358	0.059	0.386
cs_vcrim	0.065	0.060	0.023	0.065	0.073	0.066	0.026	0.073	0.062	0.059	0.012	0.062
cs_bulge	0.103	0.092	0.038	0.103	0.085	0.079	0.020	0.085	0.045	0.041	0.018	0.045
cs_other	0.196	0.224	-0.070	0.196	0.185	0.206	-0.054	0.185	0.185	0.198	-0.035	0.185
cs_probcause	0.407	0.389	0.037	0.407	0.430	0.423	0.015	0.430	0.447	0.452	-0.010	0.447
Male	0.925	0.927	-0.009	0.925	0.925	0.926	-0.004	0.925	0.833	0.826	0.017	0.833
Radio	0.188	0.182	0.017	0.188	0.220	0.205	0.036	0.220	0.295	0.278	0.038	0.295
1.dow	0.099	0.100	-0.005	0.099	0.100	0.097	0.011	0.100	0.097	0.100	-0.008	0.097

2.dow	0.150	0.135	0.041	0.150	0.155	0.140	0.041	0.155	0.155	0.147	0.020	0.155
3.dow	0.159	0.143	0.045	0.159	0.162	0.146	0.042	0.162	0.171	0.153	0.047	0.171
4.dow	0.152	0.169	-0.048	0.152	0.153	0.164	-0.029	0.153	0.158	0.170	-0.031	0.159
5.dow	0.164	0.171	-0.019	0.164	0.162	0.179	-0.047	0.162	0.163	0.179	-0.045	0.163
6.dow	0.157	0.161	-0.010	0.157	0.153	0.165	-0.034	0.153	0.146	0.152	-0.016	0.146
2.shift	0.230	0.221	0.022	0.230	0.244	0.248	-0.009	0.244	0.261	0.275	-0.031	0.261
3.shift	0.498	0.576	-0.157	0.498	0.489	0.549	-0.119	0.489	0.473	0.503	-0.060	0.473
2.agenew	0.046	0.045	0.004	0.046	0.041	0.044	-0.014	0.041	0.045	0.047	-0.011	0.045
3.agenew	0.259	0.273	-0.031	0.259	0.262	0.272	-0.023	0.262	0.281	0.280	0.002	0.281
4.agenew	0.206	0.199	0.016	0.206	0.227	0.218	0.021	0.227	0.204	0.193	0.028	0.204
5.agenew	0.226	0.220	0.014	0.226	0.248	0.247	0.004	0.248	0.210	0.210	-0.001	0.210
6.agenew	0.259	0.259	0.000	0.259	0.218	0.214	0.008	0.218	0.254	0.263	-0.021	0.254
7.agenew	0.004	0.004	0.001	0.004	0.003	0.004	-0.011	0.003	0.006	0.006	-0.004	0.006

Sdiff=standardized difference of means; W=after entropy weighting

Appendix A2. Unadjusted Outcomes from Stops After Impact Zone Expansion Compared to Other Areas

	Arrested	Summons	Frisked	Searched	Hands	Wall/Car	Contraband	Weapon
Blacks IZ	1.768**	1.298**	1.731**	1.406**	1.469**	0.919	1.146	2.075**
	(0.183)	(0.0895)	(0.0911)	(0.123)	(0.105)	(0.133)	(0.219)	(0.489)
Before	0.0231	0.0819	0.471	0.0527	0.147	0.0179	0.0180	0.00515
After	0.0401	0.104	0.607	0.0726	0.202	0.0165	0.0205	0.0106
N=	26,461	26,461	26,461	26,461	26,461	26,461	14285	14404
Other Areas	1.137**	1.012	1.189**	1.083**	1.142**	1.114	0.973	1.769**
	(0.0332)	(0.0315)	(0.0244)	(0.0320)	(0.0329)	(0.0544)	(0.0475)	(0.143)
Before	0.0619	0.0702	0.525	0.0885	0.202	0.0339	0.0370	0.00790
After	0.0698	0.0709	0.568	0.0952	0.225	0.0376	0.0361	0.0139
N=	175,518	175,518	175,518	175,518	175,518	175,518	95,632	96,587
τ	3.39**	3.03**	5.74**	2.54	2.97**	1.35	1.50	0.60
Hispanic IZ	1.946**	1.123	2.048**	1.579**	1.971**	1.332	0.999	2.188
	(0.314)	(0.111)	(0.161)	(0.188)	(0.196)	(0.288)	(0.265)	(0.873)
Before	0.0258	0.0868	0.444	0.0610	0.138	0.0206	0.0236	0.00584

After	0.0491	0.0964	0.621	0.0930	0.239	0.0273	0.0236	0.0127
N=	7,479	7,479	7,479	7,479	7,479	7,479	4032	4062
Other Areas	1.105**	1.003	1.216**	1.123**	1.197**	1.101	0.968	2.060**
	(0.0372)	(0.0375)	(0.0285)	(0.0374)	(0.0367)	(0.0584)	(0.0555)	(0.215)
Before	0.0637	0.0740	0.531	0.0917	0.196	0.0407	0.0349	0.00676
After	0.0699	0.0743	0.579	0.102	0.226	0.0446	0.0338	0.0138
N=	112,023	112,023	112,023	112,023	112,023	112,023	61,752	62,354
τ	2.66**	1.02	0.68	0.40	0.64	0.20	0.11	0.14
Others IZ	2.521**	1.106	1.631**	1.772**	1.679**	1.339	1.950	0.946
	(0.642)	(0.157)	(0.158)	(0.295)	(0.197)	(0.364)	(0.823)	(0.725)
Before	0.0181	0.0749	0.445	0.0510	0.138	0.0142	0.0118	0.00435
After	0.0444	0.0822	0.567	0.0869	0.211	0.0189	0.0228	0.00412
N=	3,240	3,240	3,240	3,240	3,240	3,240	1,645	1,662
Other Areas	0.970	0.926	1.128**	1.023	1.211**	1.048	0.942	1.626**
	(0.0393)	(0.0459)	(0.0332)	(0.0419)	(0.0477)	(0.0647)	(0.0639)	(0.193)
Before	0.0611	0.0705	0.401	0.0783	0.147	0.0285	0.0455	0.0090

After	0.0594	0.0656	0.430	0.0799	0.173	0.0298	0.0430	0.0146
N=	77,567	77,567	77,567	77,567	77,567	77,567	32,137	32,439
τ	2.41**	1.10	3.12**	2.51**	2.31**	0.79	1.19	-1.47

Exponentiated coefficients; Standard errors in parentheses and clustered on officer ID. Marginal probabilities for each outcome are displayed in the period before and after impact zones were expanded.

** $p < 0.01$