RE: ACLU Matter vs. - REF# 1340012232

Analysis of Chicago Police Department Post-Stop Outcomes during Investigatory Stops July through December 2016 (Period 2): Input to Hon. Arlander Keys' (Ret.) Second First Year Report

REVISED FINAL TECHNICAL REPORT

Ralph B. Taylor, Lallen T. Johnson, Doris Weiland

DATE: March 3, 2018

Acknowledgments. All the material herein represents only the views of the authors and does not reflect the views or policies of any other organization including the City of Chicago, the Chicago Police Department, or ACLU-Illinois. Any mistakes or misinterpretations herein are solely the authors'.
Declaration of Conflicting Interests. The authors declare no potential conflicts of interest with respect to the research, authorship and/or dissemination of this work.
Funding. This work was paid for by a contract to Temple University, with a sub-contract to Drexel University, from the Department of Law, City of Chicago.

1 CONTENTS

2 TABLE OF FIGURES

3 TABLE OF TABLES

4 FOR THE NON-TECHNICAL READER: FAQ

4.1 OUTCOMES OF INTEREST AND DATA SOURCES

Q: What is this report about?

A: This report describes what happens to civilians detained as part of an investigatory stop by officers of the Chicago Police Department (CPD) officers. Of interest are stops during the last six months of 2016, from July 1st through December 31st. Of special interest is how the outcome of the stop may depend on the race and/or ethnicity, also referred to as the ethnoracial category, of the stopped civilian. The influence of the ethnic or racial composition of the community where the stop happened is of interest as well.

The outcomes of interest in Period 2 receiving extensive attention are:

- Whether the detainee received a pat down;
  - If a pat down occurred, whether a weapon or firearm got discovered;
- Whether the detainee was searched; and
- In a stop involving no enforcement action, whether the detainee received a pat down.

Descriptive information is also provided about other outcomes and stop features.

In addition, information is presented about how detainee how race and ethnicity connect to outcomes of interest in the second half of 2016.

Finally, the report also examines changes from the first half of 2016 to the second half of 2016 for a number of outcomes.

Q: How many stopped civilians were there?

A: During the period there was a total of 51,538 investigatory stops. Single individuals may have been stopped multiple times during the period, so the exact number of individual civilians that were stopped is not known.

Most of these stops, 48,831 (94.8 percent), resulted in the CDP officers writing up only one version of the stop. These are called single version investigatory stop reports (ISRs). For 2,707 (5.2 percent) of the stops, officers ended up writing more than one version of the stop report. These are called multiple version or multi-version stop reports.

Q: What data source does this report rely on?

A: The report analyzes records from the Chicago Police Department investigatory stop reports (ISR) databases. These databases provide a wealth of information, only some of which is used here.

One ISR database provided included all versions of all multi-version ISRs for the period. From this, only the last version of each multi-version ISR was retained. CPD expected this group to comprise 2,714 records. The data sets used here yielded only 2,707 multi-version records in their final form.

Another database of ISRs was checked against a master list of ISR report numbers. From this file, 48,831 single version ISRs were retained. This is the same number the CPD expected for the period.

Additional information about the residential composition of the police beat was developed from US Census data for 2011-2015. These Census data come from the American Community Survey, an ongoing survey conducted by the Bureau of the Census. The Bureau releases data annually based on the last five years of surveys at the census block group level for the city of Chicago. Typically, in a city like Chicago a census block group is a collection of four census blocks, and corresponds roughly to a small neighborhood. A census block is the area enclosed by and bounded by the four sides of a city block. The Census information allows us to take into account the demographic structure of the communities in which the stops took place. Such demographic information is routinely included in studies examining patterns of police stops.

Q: In what ways is this analysis different from the analysis for the first period?
A: Three changes merit mention. First, models in this analysis used the police beat as the geographic unit for clustering stops. Beats are geographic sub-units of police districts.

Because there is a sizable number of police beats, around 270, using beats as the spatial unit to capture the context in which the stop happened permits the inclusion of demographic factors at the beat level as predictors. This is a second change from the models used in the first period. These second period models align more closely, in some ways, with the current models that are in use in other recent or ongoing cases, such as Floyd et al. v. New York City, or Bailey et al. v. City of Philadelphia. The models in those cases considered influence of racial and ethnic surround. In Floyd et al., surround was defined as the police precinct or the Census tract (Fagan, 2012). In Bailey et al., analysts started by defining the police district as the level of surround of interest, but have migrated to considering police service areas, or PSAs (Rudovsky, Messing, & Lin, 2017), which are somewhat comparable to the beats-within-districts considered here. The current models, similar to the Floyd and Bailey models, will also take account of the racial composition of geographic surround.

Third, where feasible, the main statistical models used in this analysis account for the clustering of the stops at multiple layers. Stops are clustered within beats, and beats in turn are clustered within districts.

4.2 Takeaway Lessons

Q: What are your most important findings?
A: First, changes over time appeared. In the latter half of 2016 compared to the first half of 2016:

(a) Stops were less likely to involve a pat down.
(b) Stops were less likely to involve a search.
(c) Pat downs proved more likely to produce a weapon (3.5 percent of the time up from 2.5 percent).
(d) The proportion of stops resulting in any type of enforcement action declined (from 32 percent down to 28 percent).
(e) The proportion of stops generating *neither* a pat down *nor* any enforcement action went up (50 percent up from 43 percent).

Interpreting these changes from the first to the last half of 2016, however, proves challenging. We do not know the causes for these changes. For each, as noted in the discussion section, there could be multiple competing explanations for the shift. Lacking further information, it is hard to say why specific changes happened.

The second takeaway lesson is that racial disparities persist when predicting whether a pat down takes place. Black non-Hispanic detainees were more likely to experience a pat down compared to White non-Hispanic detainees after controlling for other factors. This difference appeared as well in the first reporting period. This difference is consistent across different sets of stops and across multiple types of analyses.

Ethnic composition contributed consistently, at least in the main models, to pat down probability differences. Detainees were more likely to be patted down in more predominantly Hispanic police beats. This connection was not re-tested with alternate analytics.

Third, if we turn to stops involving no enforcement action, Black non-Hispanic detainees were more likely to experience a pat down with no enforcement than were White non-Hispanic detainees. This disparity replicated across samples and analyses and was present in the first reporting period as well.

Simply put, two key racial differences on outcomes were observed consistently in both the first half of 2016 and the second half of 2016.

4.3 LIMITATIONS

Q: Does your study have limitations?

A: It has many. These are described in a section of the discussion. Most importantly, though, the results seen here could change if the models we used had considered a different set of factors. Additionally, there were other aspects we wanted to explore, either in terms of different types of analytics or additional diagnostics of the models that we used, but did not have time to complete.

5 KEY FINDINGS

5.1 FIRST VS. SECOND HALF OF 2016

Comparisons between the first half of calendar year 2016 and the second half are noted below. Some changes proved statistically significant. Whether a change was statistically significant or not, we do not know the reasons for the change. For example, changes in the number of stops could arise simply from weather and related outdoor activity pattern differences between the first and last half of the year. Here are the major changes and consistencies observed:

- The relative frequency of the three key ethnoracial groups among the investigatory stops appears largely unchanged from the first half of the year to the second half of the year.
- Approximately 3,500 fewer investigatory stops took place in the last half of the year as compared to the first half.
- The proportion of stops that resulted in a pat down declined significantly, from 34 to 30 percent.
- A significantly higher fraction of pat downs yielded weapons in the last half of 2016 (3.5 percent) compared to the first half (2.5 percent).

- In the second half of the year, a significantly lower fraction of stops resulted in any enforcement action being delivered, 28 percent compared to 32 percent in the first half.

5.2 Patterns in the second half of 2016

Focusing only on stops that took place in the last six months of 2016, the following disparities appeared:

- Considering the post stop outcomes of whether the detainee was patted down by police or not, Black non-Hispanic detainees' odds of getting [patted down versus not patted down] were significantly higher compared to the odds for White non-Hispanic detainees. This significant difference appeared consistently across both random samples and across alternative analyses.
- Odds of getting [patted down versus not patted down] also were higher in more predominantly Hispanic police beats. This significant difference appeared in both random samples in the main model but was not tested with alternative analyses.
- For searches, after removing all stops where an arrest took place, no feature of detainee race or ethnicity, or ethnic or racial composition of locale, significantly affected this outcome.

6 Scope

The unit of analysis is the individual stop. The focus is on understanding the connections between civilian race, ethnicity, and gender differences and each of these outcomes. Connections between community racial and ethnic composition and these outcomes merit attention as well. The connections are considered in different ways.

First, the connections are considered on their own, without taking other factors into account. These represent **gross impacts** of detainee race or ethnicity differences, or gross impacts of racial or ethnic community composition, on the outcome. Gross impacts are simply described.

The connections are also considered with progressively stricter criteria. A second examination asks: If racial or ethnic features of either detainee or surround affect the outcome, do these same connections persist after controlling for other factors? If they do, they are called **net impacts** of race or ethnicity, or of racial or ethnic community composition.

The third examination asks: Is the net impact statistically significant, that is, unlikely to arise from chance alone?

And finally, after conducting alternative analytics, the fourth examination asks: does a statistically significant net impact merit a causal as opposed to correlational interpretation? To answer that, diagnostics of specific models are conducted, and alternative models merit consideration. It is *not* possible for all outcomes to conduct alternative models that bear on the causal interpretation question.

The nature of how the analytic goals here align with policy questions about disparate treatment and disparate impact is not assessed. For *the specific outcomes examined here,* the scholarship does not seem to provide a clear cross referencing. See the discussion section for more commentary.

6.1 Outcomes of Interest

What happens after a stop has been initiated has important practical and policy repercussions. This report considers the racial and ethnic patterning of select post-stop outcomes. Questions of who is stopped where are addressed in a different ecological report.

The following specific post-stop outcomes receive attention here:

A. Is a pat down conducted or not?
B. If a pat down is conducted, is a weapon found?
C. Is a search conducted or not?
D. If a search is conducted, is a weapon found?
E. What are the chances that the stopped civilian experienced a pat down combined with no enforcement action vs. no pat down and no enforcement action?

6.2 Questions Addressed

6.2.1 Descriptive

To provide descriptive context, simple race and ethnicity differences, and district differences, are portrayed for all of the above mentioned outcomes. Although statistical tests are (usually) not applied, these descriptive differences between ethnoracial categories represent an important part of the examination.

6.2.2 Involving statistical inference

For each outcome, the relevant questions are the same:

The race question: Controlling for observed covariates, i.e., other relevant factors, is there a statistically significant net difference in outcome scores between non-Hispanic Black civilians and non-Hispanic White civilians?

The ethnicity question: Controlling for observed covariates, is there a statistically significant net difference in outcome scores between White Hispanic civilians and non-Hispanic White civilians?

Stated differently, each model tests a null hypothesis of no difference between non-Hispanic White civilians, and either non-Hispanic Black civilians or Hispanic civilians, after controlling for observed covariates and district and beat-within-district contexts.

Ethnic and racial composition of the surround: Controlling for other beat features as well as for detainee and stop features, does the racial or ethnic composition of the beat itself significantly influence the outcome in question?

7 BACKGROUND: POLICE POST STOP OUTCOMES

For more background on research about police post stop outcomes, see Taylor and Johnson (2017: Section 6).

7.1 ANALYTIC CONCERNS

7.1.1 Internal replication across independent samples

Two representative independent random samples of data are available. Tests of statistical significance are conducted on both samples. If a key statistically significant finding surfacing with one sample also reappears as significant in the second sample, then the statistical finding has been internally replicated. Internally replicated significant findings inspire more confidence. They suggest the findings are robust across independent random samples. They suggest that the linkage observed does not depend on something about the mix of records found in one sample but not the other. We observe one type of robustness if a specific significance pattern for a key predictor replicates across two independent samples of data.

This technique is called split-sample cross-validation and is widely used in social science

7.1.2 Internal replication across alternative analytic approaches

The main statistical analysis used throughout is **multiple regression.** This is used in many different studies examining potential racial or ethnic disparities in policing. For example, the agreed upon statistical benchmarks because of the consent decree emerging from *Bailey et al. v. City of Philadelphia* use multiple regression models (Rudovsky et al., 2017).

Such models are used here, with some minor improvements. The improvements are in line with the current best practices for scholarship in this area. First, if the outcome is binary, it is modeled as binary rather than normally distributed. [1] Second, where available, mixed effects models separate random variation by district and by beat-within-district on each outcome, and allow for correlated errors both within districts and within beats-within-districts. [2] They also make Empirical Bayes adjustments to district-level means. [3]

In one case the outcome is categorical, thus the model uses multinomial multiple regression rather than several logistic multiple regression models. [4]

In addition to these main multiple regression models, we employed **an alternative analytic strategy for almost every outcome**. Doing so indicates whether a particular statistically significant net impact of a race or ethnicity difference is robust across different models, which

---

[1] Analysts in Bailey et al. use ordinary least squares regression. A long literature suggests it is better to use regression models explicitly anticipating a binary outcome (Long, 1997).

[2] Put simply, each beat-within-district, and each district, is allowed to have its own average score on the outcome examined, before and after predictors enter the model. These are random effects because they just vary randomly across locations. After predictors have been entered, these random variations capture geographic discrepancies between predicted scores on the outcome, based on the predictors entered, and Empirically Bayes adjusted beat and district scores.

[3] Again, there is a long literature speaking to the advantages of this approach (Hox, 2010; Rabe-Hesketh & Skrondal, 2012; Snijders & Bosker, 2012)

[4] See Long (1997: 149-178) for reasons to avoid repeated logistic regression models with a categorical outcome.

may make different assumptions and/or use the data in different ways. This allows for a different type of internal replication to determine if results are robust across different statistical approaches.

7.1.3 Clustered data

The data here represent stops taking place within a specific police district or police beat. That clustering has numerous statistical and analytic implications (Snijders & Bosker, 2012). It is considered in different ways with the different models used.

7.1.4 Statistical power

Because the number of total records, sizes of the random samples, and types of analyses used were generally comparable to the setup for Period 1 (Jan-June 2016), the power analyses previously conducted (Taylor & Johnson, 2017: Section 8) apply here as well. Two-tailed tests with an alpha level of $p < .05$ provide excellent statistical power (> .80) for uncovering small to medium size effects.

7.1.5 Multiple correlated outcomes

This report analyzes multiple outcomes. They do not correlate sizably with one another; all correlations are well below .10. We do not think there is an inflated experiment-wise error rate (Aickin & Gensler, 1996). But if the reader is still concerned, he or she could make his/her own internal Bonferroni adjustment by considering only the effects that are significant at $p < .01$ rather than $p < .05$.

8 METHODOLOGY

8.1 DATA SOURCES AND DATA PROCESSING

8.1.1 Chicago Police Department sourced data

Chicago Police Department (CPD) provided several data files.

A contact card file was generated in January of 2017 containing records for every version of every ISR report generated during the 2016 calendar year.

A master ISR list from CPD indicated which specific investigative stop report (ISR) numbers were generated during the period in question.

From the main contact card file, we retained only those single version ISRs whose ISR numbers appeared in the master ISR list for the period.

A separate file comprised only multi-version ISRs, that is, ISRs that generated more than one report. For each ISR, this file stacked later versions of each ISR after the earlier versions. From this file, only the latest version of each ISR was maintained. We verified that each of these ISR numbers also appeared in the master ISR list for the period.

Subsequently, the single version ISR file was joined to the file containing only the last version of each multiple-version ISR.

Therefore, for Period 2, this post stop analysis concentrates on single version ISRs added together with the last version of each multiple version ISR. This is the same composition of records as was analyzed in the first reporting period.

CPD informed us that these files should result in 48,831 single version ISRs. We obtained the same number. They also told us we should obtain 2,714 multiple version ISRs. However, we only could obtain 2,707.

CPD also provided a separate file of charge details for those detainees who were charged. That information is not used here.

8.1.1.1 Indicator (or dummy) variables were created for gender, race, and ethnicity, various times of day, days of the week, months, and age ranges, where "1" indicates that a quality was present and "0" indicates that a quality was not present. Ethnoracial groups, and ethnoracial groups of interest

The original distribution of race/ethnicity codes used by CPD personnel in the field, **RACE_CODE_CD** appears in Table 1 for both Period 1 (January-June 2016) and Period 2 (July-December 2016). This report will focus on three racial/ethnic groups: non-Hispanic Whites, White Hispanics, and non-Hispanic Blacks. Stops associated with other races or ethnicities are dropped from the analysis. [5] This permits a clean focus on the three mutually exclusive racial/ethnic groups that are most prevalent in Chicago. These three groups represent 50,723 out of 51,538 cases and 98.4 percent of ISRs for the period.

8.1.1.2 Differences in relative frequency of key ethnoracial groups: Period 2 vs. period 1

As seen in Table 1, the relative frequency of the three key ethnoracial groups did not change much between the first and last half of 2016.

As percentages of all ISRs, Black non-Hispanic detainees increased from 70.1 to 70.7 percent, White non-Hispanics increased from 7.7 to 8.3 percent, and non-Black Hispanics, simply referred to as Hispanics hereafter, decreased from 21.1 to 19.3 percent.

[5] The small number of Black Hispanics in the data are among those dropped. Therefore, the three groups of interest are exclusive of one another.

Table 1 Number of investigatory stop reports, by ethnoracial group: Period 1 and period 2

		Period 1			Period 2	
		N	Percent		N	Percent
Category	Code			Code		
Missing					204	0.4
				A	28	0.05
Asian Pacific Islander	API	417	0.76	API	394	0.76
Black				B	1,909	3.7
Black	BLK	38,361	70.13	BLK	34,542	67.02
Hispanic				H	552	1.07
American Indian / Alaskan Native		98	0.18	I	100	0.19
Undocumented code	P	67	0.12	P	84	0.16
Undocumented code				U	1	0
White				W	186	0.36
Black Hispanic	WBH	3	0.01	WBH	4	0.01
White (first code used by police)	WHI	35	0.06	WHI	18	0.03
White (second code used by police)	WHT	4,163	7.61	WHT	4,099	7.95
White Hispanic	WWH	11,557	21.13	WWH	9,417	18.27
		-----------	------------		-----------	------------
Total		54,701	100		51,538	100
Super-category						
		N	Percent		N	Percent
Black non-Hispanic		38,361	70.13%		36,451	70.73%
White non-Hispanic		4,198	7.67%		4,303	8.35%
White Hispanic		11,557	21.13%		9,969	19.34%
All other (including missing)		585	1.07%		815	1.58%
		-----------	------------		-----------	------------
Total		54,701	100.00%		51,538	100.00%

The distribution across districts of the three predominant racial/ethnic groups among stopped civilians appears in Table 2. The table shows, as was seen in Period 1, that the number of investigatory stops varies enormously by district, and that the racial mix of detained civilians also varies by district.

Table 2. Number and percent of investigatory stops by ethnoracial group and district [6]

	White NH		Black NH		Hispanic		Total for district	
District	N	%	N	%	N	%	N	%
1	99	16.05	456	73.91	62	10.05	617	100
2	62	2.04	2,932	96.54	43	1.42	3,037	100
3	13	0.83	1,543	98.09	17	1.08	1,573	100
4	72	2.39	2,382	79.06	559	18.55	3,013	100
5	42	1.72	2,363	96.76	37	1.52	2,442	100
6	31	1.55	1,955	97.95	10	0.5	1,996	100
7	44	1.32	3,220	96.52	72	2.16	3,336	100
8	357	12.66	1,489	52.82	973	34.52	2,819	100
9	327	8.74	1,313	35.11	2,100	56.15	3,740	100
10	96	2.55	2,783	73.96	884	23.49	3,763	100
11	448	6.72	5,925	88.87	294	4.41	6,667	100
12	183	9.36	903	46.17	870	44.48	1,956	100
14	111	11.19	328	33.06	553	55.75	992	100
15	80	2.25	3,374	94.72	108	3.03	3,562	100
16	737	47.58	381	24.6	431	27.82	1,549	100
17	209	23.27	173	19.27	516	57.46	898	100
18	122	12.62	768	79.42	77	7.96	967	100
19	332	24.57	718	53.15	301	22.28	1,351	100
20	203	22.94	404	45.65	278	31.41	885	100
22	74	7.49	900	91.09	14	1.42	988	100
24	354	20.18	894	50.97	506	28.85	1,754	100
25	293	10.71	1,195	43.69	1,247	45.59	2,735	100
31	14	18.67	47	62.67	14	18.67	75	100
Total	4,303	8.48	36,446	71.86	9,966	19.65	50,715	100

Note. Data for July-December 2017 (Period 2). Total shown here differs from total shown in Table 1 because 8 records associated with district 41 (error) and 815 detainees not included in the three central ethnoracial groups of interest are excluded.

8.1.2 Sampling

The data for the period were separated into two independent 50 percent random samples. Random numbers between 0 and 1 were generated for each record. The numbers followed a uniform distribution. A median split on the random numbers generated two independent samples.

[6] According to Officer Joseph A. Candella of the Chicago Police Department, "District 31 is used as a code for "'Out of City'" (email correspondence, August 28th, 2017). These records will be dropped from later analyses, but are shown here in this one table for completeness.

8.1.3 Units of analysis

The unit of analysis is each specific stop, although the same person might have been stopped multiple times over the time period.

8.1.4 Census sourced data

2011-2015 American Community Survey (ACS) (The United States Census Bureau, 2014) data at the census block group level were obtained and re-allocated from the census block group level to the police beat-within-district and police district levels. The re-allocation permitted estimating the demographic fabric of residents and residential households of beats in Chicago.

8.2 CLUSTERING

The clustering of stops within events is ignored here. However, analyses will consider the clustering of stops within beats, and simultaneously, where analytics permit, the clustering of beats within districts. A strong case can be made that such three-level models (stops within beats, and beats within districts) are the most analytically appropriate. [7]

8.3 GEOGRAPHIES AND IMPLICATIONS FOR ANALYSES

8.3.1 Districts

When districts are used as the geographic unit of clustering, there is an analytic limitation (Bryan & Jenkins, 2016; Schmidt-Catran & Fairbrother, 2016). Since there are only 22 districts, district-level demographic predictors cannot be included. Consequently, district models conducted here simply allow the outcome to differ across districts and incorporate district-to-district differences as random effects in these models.

8.3.2 Beats

The limitation noted by Schmidt-Catran (2016) and others (Bryan & Jenkins, 2016) can be overcome if analyses are conducted using the more numerous police beats (over 200) rather than police districts. The main models here nest stops within beats where feasible.

8.3.3 Analytic differences between current work and Floyd and Bailey models

The models used here bear comparing with those performed by Fagan in Floyd et al v. New York City (Fagan, 2012), as well as those used in the monitoring of Bailey et al. v. City of Philadelphia (Rudovsky et al., 2017).

8.3.3.1 Stop and stop outcome classification by ethnoracial group

Perhaps the most important difference between the models here and the Floyd models is in stop classification by ethnoracial group, which is also the most important similarity between the models used here and the Bailey models.

The Floyd models do not characterize stops by the race/ethnicity combination of the detained civilians while the Bailey models do. The practice followed in this report aligns with the Bailey approach. More specifically, stops and therefore stop outcomes are classified here according to

[7] At the same time, such models have limits and are complicated to estimate (Hox, 2010: 32). Most importantly, these models do *not* allow for any random effects of beat-level predictors.

their membership in one of three exclusive ethnoracial groups: White non-Hispanic, Black non-Hispanic, and Hispanic.

8.3.3.2 More numerous spatial units

In Fagan's models (tables 5, 7) he used New York City precincts as the spatial unit of analysis while predicting stop counts. New York City includes 75 precincts, excluding Central Park (Kane, 2002). Because he analyzed several dozen spatial units, he could put spatial demographic predictors, such as percent Black population, in his models without worrying about the limitations arising from analyses with only a couple of dozen geographic units (Bryan & Jenkins, 2016; Schmidt-Catran & Fairbrother, 2016). Because Chicago is smaller than New York City and has well more than 200 beats, the spatial units here (beats) are smaller than Fagan's spatial units of districts. Spatial scaling differences (Taylor, 2015) could contribute to different findings across the two studies. Turning to Bailey et al. which uses 66 PSAs in current models, the number of spatial units used here is larger by comparison.

8.3.3.3 Fixed effects for districts vs. random effects for districts vs. no effects for districts.

Fagan controlled for differences across boroughs by including fixed effects for boroughs. [8] The fixed effects for borough differences translated his model into an examination of intra-borough differences on the outcome. Because the models here are interested in beat-level connections between context and outcome, and not beat-intra-district variation only, we do not control for districts with fixed effects. Rather, districts are allowed to vary randomly.

The Bailey et al. models in Philadelphia in their current incarnation ignore both district variation and beat variation in their main models. That is, they use "flat" models that do not recognize clustering either by district or by PSA-within-district.

Arguably, the approach followed here approximates data features more closely. At the district level, extensive scholarship describes how norms in a police department can vary across district organizations (Klinger, 1997; Taniguchi, 2010). At the beat level, officers may confront civilians and situations that are similar (Chainey & Ratcliffe, 2005).

8.3.3.4 Beat features

The models here control for racial composition, ethnic composition, residential socioeconomic status, residential stability, and percent males under 24. These align relatively closely with the types of features of PSAs controlled for in the Bailey et al. models (Rudovsky et al., 2017).

Sometimes we use categorical variables for racial and ethnic composition, rather than just percentages. In their analyses of violent and property crime at the census tract level in 90 cities in the US, Peterson & Krivo (2010) argued for classifying locations as 70 percent or more Black non-Hispanic as predominantly Black, 70 percent or more Hispanic as predominantly Hispanic, and 70 percent or more White non-Hispanic as predominantly White. They suggest that key dynamics associated with socioeconomic disadvantage and local resource availability intensify when the locale is predominantly Black or Hispanic.

[8] This is typically done by entering a dummy indicator for each borough, save one.

These demographic features were derived by re-allocating Census data at the census block group level, from the 2011-2015 American Community Survey, to police beats. The current ecological report provides more detail (Johnson et al., 2017).

8.4 Outcome variables

8.4.1 Overall descriptive statistics and changes from first to second half of 2016

Descriptive statistics on the binary outcome variables appear in Table 3. Details on the levels and patterns for each variable are described further below. For outcomes also reported for the first six months of 2016 (Period 1), corresponding proportions and totals are noted. Sometimes changes are tested for statistical significance. [9] See the discussion section for thoughts on the challenges of interpreting these changes.

8.4.1.1 Change in number of ISR reports

The first point to note is a drop of 6.3 percent in the total number of ISRs: for the three ethnoracial groups of interest, for stops within city districts, officers filed 50,715 ISRs in the second half (Table 1Table 2) compared to 54,116 ISRs in the first half of the year (Taylor & Johnson, 2017: Table 4).

8.4.1.2 Pat downs and pat down hits

A pat down took place in 29 percent of Period 2 stops, for a total of 14,945 over the period for the three ethnoracial groups of interest (Table 3). The number of pat downs during this period (14,945) was more than 3,000 fewer than the number of pat downs occurring during the first six months of 2016 (18,364). Since the total number of stops is also down noticeably (50,715 compared to 54,116 in Period 1), the decrease in the number of pat downs is not necessarily that informative.

Perhaps of more interest is the decline in the proportion of stops with pat downs. Whereas in the first six months of 2016 almost 34 percent of stops resulted in a pat down, in the latter half of the year only about 30 percent of stops resulted in a pat down.

This drop in proportion of stops with a pat down from the first to the second half of the year proved statistically significant, suggesting that the difference in proportions in the second versus the first period it is not just noise in the data. [10] Investigatory stops produced fewer pat downs per stop in the last half of the year compared to the first half of the year.

---

[9] Strictly speaking, statistical tests comparing two populations rather than two samples are inappropriate. That nicety is ignored here. Where we can, statistical tests are done on random samples from the period(s).

[10] The difference in proportions between first and second period data was tested using a Yates continuity-corrected χ^2 (Blalock, 1979: 290-292). The test was done twice, once using just the first random half sample in each period, and again using just the second random half sample from each period. For the first random half samples, one sample from Period 1 and one sample from Period 2, χ^2 (df=1) = 116.7; $p < .0001$; for the second random half samples χ^2 (df=1) = 123.8; $p < .0001$.

Simply stated, the chances that this drop in the proportion of stops with a pat down from the first half of the year to the second half of the year was due to *just* random fluctuation was less one in ten thousand.

Table 3 Descriptive statistics, binary outcome variables

					Period 2			Period 1	
Variable	Variable name	N	Min.	Max.	Mean	SD	Sum	Mean	Sum
Pat down conducted	dpat	50,715	0	1	0.295	0.456	14,945	0.339	18,364
Pat down→weapon (*)	pathit_w2	14,945	0	1	0.035	0.183	517	0.025	465
Pat down → drugs (*)	pathit_d2	14,945	0	1	0.019	0.136	280		
Pat down → any (*) (a)	pathit_2	14,945	0	1	0.063	0.243	940		
Search conducted	dsearch	50,715	0	1	0.138	0.345	7,002	0.177	9,595
Search →weapon (*)	se_hit_w2	7,002	0	1	0.050	0.217	348		
Search → drugs (*)	se_hit_d2	7,002	0	1	0.146	0.353	1,022		
Search → any (*)	se_hit_2	7,002	0	1	0.188	0.390	1,313		
Any enforcement action taken (b)	denforce_2	50,642	0	1	0.278	0.448	14,066	0.322	17,425

Note. For all binary outcomes, 1 = outcome occurred, 0 = did not occur
Note. "Drugs" means officer found either drugs or contraband or both.
Note. "Any" means officer found either a weapon or (drugs or contraband).
Source: Period 2: July-December 2016 ISRs, CPD. Period 1: January-June 2016 ISR reports, CPD.
(*) = dependent variable depends on selection through another dependent variable. More specifically, the pat down "hit" variables were set to missing if no pat down took place; and, the search "hit" variables were set to missing if no search took place.
(a) The numbers for this variable were taken from the ISR form checkbox "was a weapon or contraband discovered as a result of the pat down?" Items mentioned included, in addition to unspecified weapons: firearms ;specific drugs such as cannabis, heroin, cocaine; other controlled substance; drug paraphernalia; alcohol; stolen property; and unspecified 'other'.,. So, the total for this variable will exceed the total of firearms plus weapons plus drugs found because of items such as "other", "alcohol" and "stolen property".[11]
(b) This indicator is based on the ISR check box "any enforcement action taken." This check box, however, was not checked 73 times when a box for a specific enforcement action *was* checked. More specifically, 70 times when "other" was checked as the enforcement type code, one time when arrest was checked, one time when violation was checked, and one time when personal citation was checked, the overall enforcement check box indicated no rather than yes for any enforcement action taken. Those 73 cases are *not* included here because priority is given to the overall check box, and those 73 cases are set to missing.

Some outcomes are dependent upon another particular post stop outcome taking place and are marked accordingly in the table (*). The pat down "hit" variables and the search "hit" variables are both of this type.

However, ignoring selection and focusing specifically on whether pat downs yielded weapons or not, the proportion yielding weapons in the second half of the year (.0346) was significantly

---

The Yates continuity-corrected χ^2 values were calculated by using "Calculator 3" under "Clinical Research Calculators" on Richard Lowry's VassarStats.net website.
[11] If the term 'firearm' is not mentioned, the term 'weapon' applies to either firearm or non-firearm weapons.

higher than in the first half of the year (.0253). [12] The higher likelihood that pat downs during investigatory stops produced weapons in the latter half of the year as compared to the first half of the year is probably not just reflected in random variation over time.

Of course, the cause of this higher weapon "hit" rate is not known. Did weapon carrying increase among those likely to be stopped and patted down by police from the first half of the year to the second half? Did officers become more discriminating about who might have a weapon if patted down? The reason for the change is unknown.

Almost 2 (1.9) percent of pat downs produced drugs or contraband in the latter half of 2016. [13]

If drugs, firearms, weapons, drug paraphernalia and other items of interest are considered collectively, 6.3 percent of the pat downs in the second half of the year resulted in a "hit." In 940 of the 14,945 pat downs officers recovered something meriting their attention.

8.4.1.3 Searches and search hits

Moving on to searches, in the second half of 2016, officers conducted searches in 13.8 percent of their investigatory stops (7,002 out of 50,715). This proportion of stops with searches proves significantly lower than the corresponding proportion for the first half of the year (17.7 percent); the drop is statistically significant and unlikely to be due *just* to chance fluctuations. [14]

During Period 2, about one out of six searches (1,313 out of 7,002) proved productive of something; 18.8 percent of them generated firearms, weapons, drugs, contraband, or something else noteworthy. [15]

About one in 20 searches or five percent yielded a firearm or a weapon during Period 2.

8.4.1.4 Any enforcement action delivered

Turning to enforcement action during Period 2, investigative stops resulted in some type of enforcement action about 27.8 percent of the time (14,066 out of 50,642 stops). [16] This contrasted with 32.2 percent of investigative stops in the first half of the year with enforcement

---

[12] Yates continuity-corrected χ^2 (df=1) = 9.69; $p < .01$ if data from the first random half in each period are examined; χ^2 (df=1) = 14.69; $p < .001$ if data from the second random half in each period are examined (Blalock, 1979: 290-292).

Stated simply, the chances that the increase in proportion of pat downs yielding a weapon from the first to the second half of the year was due *just* to chance fluctuations in the data are less than one in a hundred.

[13] This outcome not examined in Period 1.

[14] Yates continuity-corrected χ^2 (df=1) = 170.76; $p < .0001$ for first random half sample in each period; χ^2 (df=1) = 132.51; $p < .0001$ for the second random half sample in each period.

In other words, the chances that the decrease in the proportion of stops with searches was due *just* to chance fluctuations in the data are less than one in ten thousand.

[15] The numbers for this variable were taken from the ISR form checkbox "was a weapon or contraband discovered as a result of the search?" Items mentioned included, in addition to unspecified weapons: firearms ;specific drugs such as cannabis, heroin, cocaine; other controlled substance; drug paraphernalia; alcohol; stolen property; and unspecified 'other'.,

[16] This indicator was based on the check box on the ISR form "Any enforcement action taken?" Specific actions officers could have checked were: arrest, personal citation, administrative notice of violation, or other.

actions. This drop in the likelihood that a stop would produce enforcement was statistically significant and unlikely to have arisen from chance fluctuations in the data. [17]

Because pat downs and enforcement actions are also *jointly* considered as an outcome, enforcement action delivered will not be analyzed separately.

8.4.1.5 Pat down and enforcement combinations

Descriptive statistics for the one categorical outcome analyzed appear in Table 4. Of interest here are all four possible combinations of outcomes when enforcement and pat down actions are jointly considered. The proportions of stops falling into different categories shift between the first half of the year and the last half of the year. [18] More specifically, the proportion of investigative stops where there were no pat downs and no enforcement action was 43 percent in the first half of 2016 and 50 percent in the latter half of the year. Given that the results above show that both pat downs and any enforcement action were both lower in the second half of the year, the finding that [no pat down + no enforcement] is up is not surprising.

Table 4 Descriptive statistics: Categorical outcome variable, pat down and enforcement combination

		Period 2		Period 1	
	Category	N	Percent	N	Percent
No pat down delivered, no enforcement action	1	25,528	50.34	23,236	42.94
Pat down delivered, no enforcement action taken	2	11,048	21.78	13,444	24.84
No pat down delivered, enforcement action taken	3	10,193	20.1	12,508	23.11
Pat down delivered and enforcement action taken	4	3,873	7.64	4,917	9.09
Missing	.	73	0.14	11	0.02
	Total	50,715	100	54,116	100.00

Note. For Period 1, there were 11 ISRs where the police checkbox “Enforcement action taken yes/no” was checked “no” but officers did indicate some type of enforcement action (10 instances, other, 1 instance, PSC). In cases where the data were internally in conflict, the variable shown here, which depends in part on whether an enforcement action was taken, was coded to missing. For Period 2 there were 73 ISRs where the police checkbox “Enforcement action taken yes/no” was checked “no” but officers did indicate some type of enforcement action (70 instances “other”, one instance each for arrest, personal citation, and administrative notice of violation).

8.4.2 Pat downs: Across groups and districts

The left-most columns in Table 5 display the number of pat downs in each district in Period 2 and by ethnoracial group within district; and, in the right-most columns, the proportions appear of investigatory stops resulting in a pat down, by district, and within district, by ethnoracial group.

---

[17] Yates continuity-corrected χ^2 (df=1) = 129.99; p < .0001 for the first random half sample in each period; χ^2 (df=1) = 117.96; p < .0001 for the second random half sample in each period.
In other words, the chances that the decline in the proportion of stops with enforcement action was due *just* to chance fluctuations in the data are less than one in ten thousand.

[18] χ^2 (df=1) = 588.35; p < .0001. Lowry’s VassarStats website, under “Clinical Research calculators” and “Chi-square, Cramer’s V, and Lambda” calculated this value using the Period 1 and Period 2 counts shown in the table for this outcome, excluding the row for missing. The observed frequency for Period 2 and the outcome no enforcement, no pat down was eight percent lower than the expected frequency if there was no relationship between period and category.

City-wide in the first half of 2016, the district with the fewest pat downs was District 1 (The Loop), and the district with the most pat downs was District 7 (Taylor & Johnson, 2017: Table 9), mid-south Chicago from 55th Street down to 75th Street. In the second half of the year it was the same: 138 pat downs in District 1 was the lowest and 1,697 pat downs in District 7 was the highest. [19]

The right-most columns in Table 5 show the proportion of investigatory stops where police conducted a pat down. The proportions vary across districts. A typical detained civilian in a typical investigatory stop would have an approximately 1 out of 2 chance of being patted down if the stop happened in District 7 (50.9 percent). By contrast, the odds would be about 1 out of 11 (9.4 percent) if the stop happened in District 16 in northwestern most Chicago.

Table 5 Number of pat downs and proportion of investigative stops with pat downs: By district, and ethnoracial group within district.

	Number of pat downs				Percentage of investigatory stops with pat downs			
Ethnoracial group	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
District								
1	16	112	10	138	16.16%	24.56%	16.13%	22.37%
2	12	440	6	458	19.35%	15.01%	13.95%	15.08%
3	6	592	5	603	46.15%	38.37%	29.41%	38.33%
4	27	911	203	1,141	37.50%	38.25%	36.31%	37.87%
5	10	824	12	846	23.81%	34.87%	32.43%	34.64%
6	11	825	3	839	35.48%	42.20%	30.00%	42.03%
7	19	1,644	34	1,697	43.18%	51.06%	47.22%	50.87%
8	79	284	273	636	22.13%	19.07%	28.06%	22.56%
9	89	477	855	1,421	27.22%	36.33%	40.71%	37.99%
10	23	685	418	1,126	23.96%	24.61%	47.29%	29.92%
11	62	1,301	63	1,426	13.84%	21.96%	21.43%	21.39%
12	22	169	177	368	12.02%	18.72%	20.34%	18.81%
14	28	105	176	309	25.23%	32.01%	31.83%	31.15%
15	20	1,019	38	1,077	25.00%	30.20%	35.19%	30.24%
16	58	26	61	145	7.87%	6.82%	14.15%	9.36%
17	49	50	168	267	23.44%	28.90%	32.56%	29.73%
18	22	157	16	195	18.03%	20.44%	20.78%	20.17%
19	44	179	77	300	13.25%	24.93%	25.58%	22.21%
20	28	83	49	160	13.79%	20.54%	17.63%	18.08%
22	13	332	5	350	17.57%	36.89%	35.71%	35.43%
24	78	291	168	537	22.03%	32.55%	33.20%	30.62%
25	60	365	443	868	20.48%	30.54%	35.53%	31.74%
31	5	27	6	38	35.71%	57.45%	42.86%	50.67%
Total	781	10,898	3,266	14,945	18.15%	29.90%	32.77%	29.47%

[19] Recall District 31 is for outside the city.

Ethnoracial group matters as well. Looking at the overall numbers in the bottom of the table, the chances that a stopped civilian would be patted down depended on the race/ethnicity of the stopped civilian. Whereas around a third of stopped non-Hispanic Black civilians (29.9 percent) and approximately a third of stopped Hispanic civilians (32.8 percent) received a pat down, only about one-sixth of stopped non-Hispanic White civilians received the same (18.2 percent).

To give the reader a sense of odds ratios that are presented in later models consider the following.

The odds of [getting patted down vs. not patted down] for each group are derived by taking the [proportion patted down / not patted down] for each group. These are laid out below in Table 6. For example, on average, the chances of White non-Hispanic detainees being patted down during a stop were 18.2 percent, and their chances of not being patted down were 81.8 percent. This means that their odds of [being patted down vs. not patted down] is the ratio of these two chances: .182/.818 = .22.

Table 6. Patted down vs. not patted down: Proportions and odds, by ethnoracial group, by period

		White-NH	Black-NH	Hispanic
Period 2				
	Proportion patted down (a)	0.182	0.299	0.328
	Proportion not patted down (b)	0.818	0.701	0.672
	Odds of being patted down / not patted down [(a) / (b)]	0.222	0.427	0.488
Period 1				
	Odds of being patted down / not patted down	0.304	0.536	0.531
Change, from Period 1 to Period 2, in odds of [being patted down vs. not patted down]		0.732	0.796	0.919

Odds always explain the chances of [this versus that]. **Odds are different from proportions** because proportions are exclusively about the chances of this, not the chances of [this versus that].

Examining the Period 2 odds reveals that Black non-Hispanic detainees had the second to highest odds of being [patted down vs. not patted down]: .43. About 30 percent were patted down while 70 percent were not. The group with the highest odds, .49, of getting [patted down vs. not], were Hispanic detainees. Approximately 33 percent were patted down while 67 percent were not. The group with the lowest odds were White non-Hispanic detainees: .222.

How much higher are non-Hispanic Blacks' odds of [getting vs. not getting a pat down] compared to non-Hispanics' Whites' odds?

To find out one takes the ratio of the two odds, making an **odds ratio (OR).** The odds ratio indicates how much higher or lower one group's odds are relative to the odds of the other group.

In order to find the odds ratio of Black NH/White NH – or the difference in the odds between the two groups –the two odds are divided.

<u>Odds of Black NH [getting vs. not getting patted down]</u> = Odds Ratio of [Black NH vs. White NH] [getting vs. not getting pat down]
Odds of White NH [getting vs. not getting patted down]

In Period 2: [Black NH odds /White NH odds] = OR = .427/.222 = 1.917

This means that the odds of Black non-Hispanics [getting vs. not getting patted down] were **92 percent higher** than were the odds for White non-Hispanics [getting vs. not getting patted down].

The same odds ratio (OR) for Period 1 was: 1.765.

In Period 2, [Hispanic odds / White non-Hispanic odds] = OR = .488/.222 = 2.194

That is, the odds of detained Hispanics [getting vs. not getting patted down] were **119 percent higher** than the odds for detained White non-Hispanics [getting vs. not getting patted down]. Equivalently, Hispanics' odds were 2.19 times higher than White non-Hispanics' odds.

The same odds ratio (OR) for Period 1 was: 1.749

When an odds ratio is close to 1, it implies that the two groups have close to equal chances of [this vs. that] happening. For example, the odds ratio for [getting vs. not getting patted down] for [Hispanic odds / Black non-Hispanic odds] = OR = .488/.427 = 1.144

This suggests that detained civilians in these two ethnoracial groups had roughly comparable chances of being [patted down vs. not patted down].

Table 7 contains the odds ratios for [getting vs. not getting patted down] when each pair of ethnoracial groups is contrasted against each other group for both the first and last half of 2016.

Speaking only descriptively, it appears that the Black vs. White disparity on this outcome among non-Hispanic detainees increased in the second half of the year (OR=1.92) compared to the first half (1.77). This descriptive increase in disparity on this outcome is present for the Hispanic vs.

White non-Hispanic contrast as well. The disparity went from an odds ratio of 1.75 in the first half of the year to an odds ratio of 2.19 in the second half of the year.

Table 7 Odds ratios depicting ethnoracial differences in odds of getting vs. not getting patted down: By period

Period Comparison of odds (odds ratio)	Second half of 2016 (Period 2) OR	First half of 2016 (Period 1) OR
Black NH vs. White NH	1.917	1.765
Hispanic vs. White NH	2.194	1.749
Hispanic vs. Black NH	1.144	0.991

The roots of this increasing disparity between White non-Hispanic detainees and the two other groups can be found in Table 6. Two features merit attention; the odds of getting [patted down vs. not] dropped more between the first and last half of the year for White non-Hispanic detainees (.222/.304 = 27 percent drop) than they did for Black non-Hispanic detainees (.427/.536 = 20.3 percent drop) or Hispanic detainees (8.1 percent drop). Further, the White non-Hispanic detainees' odds, which are the smallest of the three groups, serve as the denominator for the disparity determination.

Further detailed analyses show that although the Black non-Hispanic vs. White non-Hispanic difference on the pat down outcome *appeared* stronger in the latter half of 2016 compared to the first half, the increase in the race difference did not prove statistically significant. The same held true when the Hispanic vs. White non-Hispanic difference was compared between the two periods. [20]

All that can be said here is that the ethnoracial pat down disparities between White non-Hispanic detainees and Black non-Hispanic detainees, and between White non-Hispanic detainees and Hispanic detainees *appeared* somewhat wider in the second as compared to the first half of the year. But that widening could have been due just to random fluctuations between the two periods. Of course, in future periods these trends bear monitoring.

At the very least then, it seems like the Black/White disparity on pat down rates revealed in the first half of the calendar year continues through the second half of the year. The same holds true for the Hispanic/White disparity.

---

[20] Results not shown. More specifically, we conducted a mixed effects logistic regression predicting the pat down outcome, with stops nested within districts, using data from both periods. We did this for each random sample separately. Further, in each case just the two groups of interest – Black non-Hispanic vs. White non-Hispanic; or Hispanic vs. White non-Hispanic – were retained. The only predictor entered was a (key group x period 2) interaction. For example, pat downs for Black non-Hispanic detainees in Period 2 were contrasted with pat downs of (combined) all Period 1 pat downs and Period 2 pat downs for White non-Hispanics.
In both cases, the plots of the marginal impact of the race or ethnic difference on the probabilities of a pat down showed comparable impacts between the two periods. More specifically, the confidence intervals around the estimated differences in predicted probabilities overlapped. In short, the contrasting chances of Black vs. White detainees being patted down, and Hispanic vs. White detainees being patted down, widened somewhat from the first to the last half of the year, but the widening could have arisen just from chance fluctuations in the data.

8.4.3 Pat down weapons/firearms recovered by group and district

As noted above, 517 pat downs produced weapons or firearms. [21] The distribution of those weapons-producing pat downs across locations and ethnoracial groups appears in Table 8.

Table 8 Number and proportion of pat downs producing weapons: By district and ethnoracial group

Ethnoracial group / District	Number of pat downs producing weapons/firearms				Proportion of pat downs producing weapons/firearms			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	1	4	1	6	0.063	0.036	0.100	0.044
2	1	8	0	9	0.083	0.018	0.000	0.020
3	0	20	0	20	0.000	0.034	0.000	0.033
4	1	35	5	41	0.037	0.038	0.025	0.036
5	0	34	0	34	0.000	0.041	0.000	0.040
6	0	36	0	36	0.000	0.044	0.000	0.043
7	1	49	2	52	0.053	0.030	0.059	0.031
8	3	11	6	20	0.038	0.039	0.022	0.031
9	2	27	14	43	0.023	0.057	0.016	0.030
10	1	27	8	36	0.044	0.039	0.019	0.032
11	3	46	3	52	0.048	0.035	0.048	0.037
12	0	3	8	11	0.000	0.018	0.045	0.030
14	0	1	6	7	0.000	0.010	0.034	0.023
15	1	35	0	36	0.050	0.034	0.000	0.033
16	2	1	1	4	0.035	0.039	0.016	0.028
17	3	1	9	13	0.061	0.020	0.054	0.049
18	1	9	0	10	0.046	0.057	0.000	0.051
19	2	4	5	11	0.046	0.022	0.065	0.037
20	2	2	1	5	0.071	0.024	0.020	0.031
22	0	11	0	11	0.000	0.033	0.000	0.031
24	2	13	7	22	0.026	0.045	0.042	0.041
25	7	6	23	36	0.117	0.016	0.052	0.042
31	0	2	0	2	0.000	0.074	0.000	0.053
Total	33	385	99	517	0.042	0.035	0.030	0.035

Note. Period 2, July-December 2016.
Note. Figures shown only for three ethnoracial groups: White non-Hispanic, Black non-Hispanic, and Hispanic detained civilians. NH = non-Hispanic

[21] For simplicity's sake, if the term firearm is not mentioned, the term weapon applies to either firearm or non-firearm weapons.

Pat downs in the 7th and 11th districts produced the greatest number of weapons/firearms: 52 in each. *Each of these two districts contributed one tenth of all recovered weapons in the entire city*. City-wide, District 16 produced the fewest number of weapons from pat downs.

Descriptively, pat downs of White non-Hispanic detainees were most likely to produce weapons/firearms (4.2 percent) whereas pat downs of Hispanic detainees were least likely (3 percent) to produce weapons/firearms. The pat down weapon hit rate was intermediate for Black non-Hispanic detainees that were patted down.

This ordering of the three ethnoracial groups on pat down weapons hit rates has shifted from the first half of the year (Taylor & Johnson, 2017: Table 12) where the ordering was White non-Hispanic (4.1 percent), Hispanic (2.9 percent) and Black non-Hispanic (2.3 percent). This is only noted *descriptively*.

8.4.4 Searches conducted by group and district

Table 9 displays the number of searches by district and ethnoracial group in the columns on the left with the corresponding proportion of stops with a search on the right. The highest number of searches during an investigatory stop occurred in District 11, and the fewest in District 1.

8.4.5 Weapons/firearms producing searches: By group and district

The number of searches that produced weapons or firearms, and the corresponding proportions, appear in Table 10. As we saw with weapons-producing pat downs, Districts 7 and 11 have the highest numbers of weapons-producing searches as well.

Speaking descriptively, weapons hit rates for searches varied by ethnoracial group. Searches of White non-Hispanic detainees were less likely to be weapons-producing (2.6 percent), compared to the searches of Black non-Hispanic detainees (5.4 percent).

The rates varied by location as well. In districts that had at least 30 searches, weapons hit rates ranged from above three percent to above eight percent.

Table 9 Number and proportion of searches: By district and ethnoracial group

	Number of searches				Proportion of stops resulting in searches			
Ethnoracial group District	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	6	47	8	61	0.061	0.103	0.129	0.099
2	3	214	3	220	0.048	0.073	0.070	0.072
3	4	186	2	192	0.308	0.121	0.118	0.122
4	10	342	64	416	0.139	0.144	0.115	0.138
5	3	300	3	306	0.071	0.127	0.081	0.125
6	8	340	1	349	0.258	0.174	0.100	0.175
7	6	681	9	696	0.136	0.212	0.125	0.209
8	28	181	123	332	0.078	0.122	0.126	0.118
9	37	157	257	451	0.113	0.120	0.122	0.121
10	7	302	106	415	0.073	0.109	0.120	0.110
11	93	1,096	58	1,247	0.208	0.185	0.197	0.187
12	14	93	61	168	0.077	0.103	0.070	0.086
14	11	38	60	109	0.099	0.116	0.109	0.110
15	19	386	19	424	0.238	0.114	0.176	0.119
16	57	38	44	139	0.077	0.100	0.102	0.090
17	25	32	107	164	0.120	0.185	0.207	0.183
18	9	94	3	106	0.074	0.122	0.039	0.110
19	50	103	64	217	0.151	0.144	0.213	0.161
20	15	40	27	82	0.074	0.099	0.097	0.093
22	9	180	0	189	0.122	0.200	0.000	0.191
24	52	143	92	287	0.147	0.160	0.182	0.164
25	40	167	208	415	0.137	0.140	0.167	0.152
31	4	8	5	17	0.286	0.170	0.357	0.227
Total	510	5,168	1,324	7,002	0.119	0.142	0.133	0.138

Table 10 Number and proportion of searches resulting in weapons/firearms: By group and district

Ethnoracial group	Number of searches resulting in weapons				Proportion of searches producing weapons			
District	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	0	2	0	2	0.000	0.043	0.000	0.033
2	0	14	0	14	0.000	0.065	0.000	0.064
3	1	20	0	21	0.250	0.108	0.000	0.109
4	1	26	8	35	0.100	0.076	0.125	0.084
5	0	27	0	27	0.000	0.090	0.000	0.088
6	0	25	1	26	0.000	0.074	1.000	0.075
7	1	44	0	45	0.167	0.065	0.000	0.065
8	1	8	6	15	0.036	0.044	0.049	0.045
9	1	9	12	22	0.027	0.057	0.047	0.049
10	1	10	5	16	0.143	0.033	0.047	0.039
11	1	39	2	42	0.011	0.036	0.035	0.034
12	0	0	0	0	0.000	0.000	0.000	0.000
14	0	0	3	3	0.000	0.000	0.050	0.028
15	1	14	1	16	0.053	0.036	0.053	0.038
16	0	0	0	0	0.000	0.000	0.000	0.000
17	1	1	3	5	0.040	0.031	0.028	0.031
18	0	4	1	5	0.000	0.043	0.333	0.047
19	0	7	0	7	0.000	0.068	0.000	0.032
20	0	2	2	4	0.000	0.050	0.074	0.049
22	0	15	0	15	0.000	0.083	0.000	0.079
24	3	8	1	12	0.058	0.056	0.011	0.042
25	1	3	11	15	0.025	0.018	0.053	0.036
31	0	0	1	1	0.000	0.000	0.200	0.059
Total	13	278	57	348	0.026	0.054	0.043	0.050

8.4.6 Is any enforcement action delivered or not?

CPD recorded four types of enforcement actions: arrest, administrative notice of violation, personal citation, and other. Table 11 shows the distribution of types of enforcement for stops where some type of enforcement took place. Speaking only *descriptively*, as a proportion of all enforcement actions, arrests were down somewhat and ANOV are up somewhat compared to Period 1.

Table 11 Frequencies of different enforcement actions

	PERIOD 2		PERIOD 1	
Types of enforcement actions	N	Percent	N	Percent
ANOV (administrative notice of violation)	4,514	31.93	5,141	29.48
ARR (arrest)	6,137	43.40	8,037	46.09
OTH (other)	2,996	21.19	3,386	19.43
PSC (personal service citation)	492	3.48	861	4.94
Total	14,139	100	17,425	100

Note. Period 2 - July-December 2016; Period 1 = January-June 2016.
Note. For Period 2, this descriptive total excludes 73 cases where there was a discrepancy between the overall check box for any enforcement action and specific actions. That is, the ISR check box "any enforcement action taken" was *not* checked 73 times when a box for a specific enforcement action *was* checked. More specifically, 70 times when "other" was checked as the enforcement type code, one time when arrest was checked, one time when violation was checked, and one time when personal citation was checked, the overall enforcement check box indicated no rather than yes for any enforcement action taken.
In keeping with the logic rule that the overall check box receives a higher priority than the subsidiary check boxes, the 73 cases where there is a discrepancy are not included here, but rather are set to missing. In statistical models using this outcome, or this outcome combined with a pat down, these 73 cases are also set to missing.
For Period 1, this descriptive total excludes 11 stops where a specific enforcement action was checked but the overall "any enforcement action taken" box was not checked. In ten of those instances the action was other and in one instance it was personal citation. In statistical models using this outcome, or this outcome combined with a pat down, these 11 cases are set to missing on the outcome.

Counts of enforcement action of any type during stops, and proportion of stops with enforcement action, appear by district and race/ethnicity combination in Table 12. Again, both counts and proportions vary by district and ethnoracial group. Overall, 29 percent of stops with Black non-Hispanic detainees resulted in a specific enforcement action, compared to 20 percent of stops with White non-Hispanic detainees. In a small number of districts – 1, 7, 11, and 22 – at least a third of stops led to some type of enforcement action. This contrasts with less than a fifth of stops linking to specific enforcement actions in districts 24, 14, and 16.

8.4.7 Pat down but no enforcement action

This outcome – being stopped and patted down, but no enforcement action resulting from the stop – appears in the procedural justice literature and considers the relative likelihood of two joint outcomes. [22] An officer mayor may not pat down a detained civilian and may or may not conduct enforcement during the stop. [23] From a procedural justice perspective, the pat down-no enforcement type of stop is thought to have potentially corrosive impacts on civilians' views of police legitimacy (Tyler, Fagan, & Geller, 2014).

[22] See section 6.2.2 in Taylor & Johnson (2017).
[23] This creates four possible outcomes: no pat down, no enforcement action; pat down, but no enforcement action; no pat down, but enforcement occurs; and pat down and enforcement both occur.

Table 12 Number and proportion of stops with specific enforcement action

Ethnoracial group District	Number of stops producing specific enforcement action				Proportion of stops producing specific enforcement action			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	29	183	13	225	0.293	0.401	0.210	0.365
2	14	665	8	687	0.226	0.228	0.186	0.227
3	5	364	4	373	0.385	0.237	0.235	0.238
4	21	607	100	728	0.292	0.255	0.179	0.242
5	5	511	3	519	0.119	0.217	0.081	0.213
6	8	532	3	543	0.258	0.273	0.300	0.273
7	7	1,073	21	1,101	0.159	0.334	0.292	0.330
8	62	419	283	764	0.174	0.282	0.292	0.272
9	76	353	540	969	0.232	0.269	0.257	0.259
10	21	874	175	1,070	0.219	0.314	0.198	0.284
11	105	2,365	82	2,552	0.234	0.400	0.279	0.383
12	27	219	155	401	0.148	0.243	0.178	0.205
14	17	61	108	186	0.153	0.186	0.195	0.188
15	27	1,102	31	1,160	0.338	0.327	0.287	0.326
16	93	54	68	215	0.126	0.142	0.158	0.139
17	59	41	171	271	0.282	0.238	0.331	0.302
18	18	218	15	251	0.148	0.285	0.195	0.260
19	85	174	101	360	0.257	0.243	0.336	0.267
20	50	98	87	235	0.246	0.243	0.313	0.266
22	22	323	1	346	0.306	0.362	0.071	0.354
24	62	174	98	334	0.175	0.195	0.194	0.190
25	56	287	411	754	0.192	0.241	0.330	0.276
31	6	13	3	22	0.429	0.277	0.214	0.293
Total	875	10,710	2,481	14,066	0.204	0.294	0.249	0.278

Note. Period 2 - July-December 2016
Note. Cases not shown (n=73) if overall enforcement action check box conflicted with individual enforcement action check box patterns.

Simultaneously considering whether the stopped civilian is patted down, and whether the stopped civilian receives any enforcement action, reveals four possible sets of outcomes. Analyses reported later will contrast two outcomes: a pat down but no enforcement action vs. no pat down and no enforcement action. Of descriptive interest is the proportion of stops where civilians were patted down by police but did not receive any enforcement action from the officer.

8.4.7.1 Temporal change

Focusing only on investigatory stops that involved no specific enforcement action of any type, the numbers in Table 13 suggest that enforcement actions occurred relatively less frequently in the latter half of 2016, dropping from 36 percent in the first half of the year to 30 percent in the second half. With the focus just on stops where no enforcement took place, it is unlikely that the drop in the type of stop thought to be most corrosive from a procedural justice perspective was due to chance. [24]

If the focus is only on stops where no enforcement actions took place, the proportion of stops where civilians were patted down before being released was lower in the second half of the year, although it is not yet clear whether the three different ethnoracial groups benefited equally from this reduction.

Table 13 Investigatory stops with no enforcement: Proportion with pat down, by period

	Period 2		Period 1	
	N	Percent	N	Percent
	----------	-----------	---------------	-------
No pat down, no enforcement	25,528	69.79	23,247	63.36
Pat down, no enforcement	11,048	30.21	13,444	36.64
	----------	-----------	---------------	-------
	36,576	100	36,691	100

Note. Period 2 - July-December 2016; Period 1 = January-June 2016.
Note. Only stops with no specific enforcement actions shown.

8.4.7.2 Overall Districts and ethnoracial groups

Returning to all stops and not just those where no enforcement took place, Table 14 shows counts and proportions of stops with pat down but no enforcement, by district and ethnoracial group. Again, the outcome varies both by location and ethnoracial group.

Overall, the proportion of stops without enforcement but with pat downs appears larger for stopped Black non-Hispanic civilians (22 percent) than for stopped White non-Hispanic civilians (14 percent). The corresponding proportion for stops with Hispanic civilians (25 percent) is the highest of the three groups.

How do these proportions align with the overall representation of the three ethnoracial groups in all the stops, that is, their respective overall stop shares?

[24] The Yates continuity-corrected χ^2 (df=1) = 178.86, p < .0001 taking the first random sample for Period 1 and Period 2; and χ^2 (df=1) = 161.48, p < .0001 using the second random sample for each period. Chances that this shift from Period 1 to Period 2 could have arisen due *just* to random variation, are less than one in 10,000. Values were calculated using "Calculator 3" under "Clinical Research Calculators" on Richard Lowry's VassarStats.net website. The focus is only on no-enforcement stops.

Table 14 Counts and proportions of stops where civilians receiving pat down but no enforcement action, by district and race/ethnicity

Ethnoracial group District	Number of stops with pat down but no specific enforcement action				Proportion of stops with pat down but no specific enforcement action			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	13	77	6	96	0.131	0.169	0.097	0.156
2	10	312	3	325	0.161	0.107	0.070	0.108
3	3	451	2	456	0.231	0.293	0.118	0.291
4	19	681	167	867	0.264	0.286	0.299	0.288
5	10	646	10	666	0.238	0.274	0.270	0.273
6	7	600	2	609	0.226	0.308	0.200	0.306
7	15	1,205	24	1,244	0.341	0.375	0.333	0.373
8	61	187	182	430	0.171	0.126	0.188	0.153
9	73	391	671	1,135	0.223	0.298	0.320	0.304
10	15	510	336	861	0.156	0.183	0.381	0.229
11	48	913	49	1,010	0.107	0.154	0.167	0.152
12	19	123	136	278	0.104	0.136	0.156	0.142
14	20	73	144	237	0.180	0.223	0.260	0.239
15	16	773	30	819	0.200	0.229	0.278	0.230
16	32	14	43	89	0.044	0.037	0.100	0.058
17	36	36	117	189	0.172	0.209	0.227	0.211
18	16	102	11	129	0.131	0.133	0.143	0.134
19	34	123	57	214	0.103	0.172	0.189	0.159
20	22	61	39	122	0.108	0.151	0.140	0.138
22	11	198	5	214	0.153	0.222	0.357	0.219
24	66	218	138	422	0.186	0.244	0.273	0.241
25	46	268	292	606	0.158	0.225	0.234	0.222
31	5	20	5	30	0.357	0.426	0.357	0.400
Total	597	7,982	2,469	11,048	0.139	0.219	0.248	0.218

Note. Period 2 - July-December 2016
Note. Cases not shown (n=73) if overall enforcement action check box conflicted with individual enforcement action check box patterns.
Note: The counts in the columns at left reflect the total number of stops where the civilian was patted down and no enforcement action was recorded. The proportions in the right most columns express those counts as fractions of all stops.
Note. NH = non-Hispanic.

8.4.7.3 Ethnoracial share: Overall stops and stops with pat down but no enforcement

Table 15 contrasts the proportions of each ethnoracial group represented in overall stops with their representation in stops with pat downs but no enforcement. That comparison appears in the last column of the table. If the three groups were represented proportionally, or the same way in both all stops and stops with pat downs but no enforcement, the ratios for each group in the last column would be 1. If a group was *under* represented in stops with pat downs but no enforcement, given their share of all stops, the ratio of the two proportions in the last column would go *below* 1.0. If a group was *over* represented in stops with pat downs but no enforcement, given their share of all stops, the ratio of the two proportions in the last column would go *above* 1.0. Figures for Period 2 appear in the top rows of the table, and Period 1 figures appear in the bottom portion.

For Period 2, detained non-Hispanic Black civilians contributed equally to stops (72 percent) and stops with pat downs but no enforcement (72 percent). This is closely comparable to the figures for Period 1.

By contrast, proportionally, detained non-Hispanics Whites contributed far less frequently to the outcome of interest (5 percent), compared to their share of overall stops (8 percent). A similar contrast appeared for this group with the Period 1 data.

One group that may be appearing more frequently in stops with pat downs but no enforcement are Hispanic detainees. Their share of stops with this pattern was fourteen percent higher than their overall stop share. This represents a shift for this group compared to the first half of the year.

The differences noted here are *descriptive only*.

Table 15. Proportional representation, three ethnoracial groups: All stops vs. stops with (pat down and no enforcement action)

PERIOD 2

	Racial / ethnic group	N: All stops	Proportional representation : all stops	N: PD+NEA	Proportional representation : PD+NEA	Ratio: [PR (Pat + NEA) / PR(All)]
	White NH	4,303	8.48	597	5.4	0.64
	Black NH	36,446	71.86	7,982	72.25	1.01
	Hispanic	9,966	19.65	2,469	22.35	1.14
Total		50,715	100	11,048	100	

Note. NH = non-Hispanic. Source: July-December 2016 ISRs, CPD. PD = pat down; NEA = no enforcement action taken. PD+NEA = stops where civilian was patted down but no enforcement actions were taken. PR = proportion

PERIOD 1

	Racial / ethnic group	N: All stops	Proportional representation : all stops	N: PD+NEA	Proportional representation : PD+NEA	Ratio: [PR (Pat + NEA) / PR(All)]
	White NH	4,198	7.76	706	5.25	0.68
	Black NH	38,361	70.89	9,828	73.1	1.03
	Hispanic	11,557	21.36	2,910	21.65	1.01
Total		54,116		13,444	100	

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD. PD = pat down; NEA = no enforcement action taken. PD+NEA = stops where civilian was patted down but no enforcement actions were taken. PR = proportion

8.5 INDEPENDENT VARIABLES

8.5.1 Stop-level predictors

Descriptive statistics for stop-level independent variables appear in Table 16.

8.5.2 Contextual predictors

As explained above, it proves unwise to include district features as predictors in our models. This is simply because there are too few districts. However, models *can* incorporate features of the locale where the stop happened, if the analysis uses beat-within-district as the geographic grouping unit. Excluding the two beats that are outside the city (District 31), each remaining district has anywhere between nine and seventeen beats within them, accounting for a total of 275 beats. It *is* feasible to include contextual predictors at the beat level, since the number of units is sizable enough to meet model assumptions in mixed effects models.

Beat-level contextual predictors permit capturing the demographic structure of the community within which the stop took place. Decades of work on the fabric of urban neighborhoods suggests three fundamental demographic dimensions: racial/ethnic composition, socioeconomic status, and residential stability (Golledge & Stimson, 1997; Hunter, 1971). The following variables, constructed from 2011-2015 American Community Survey data provided by the U.S. Census, describe these dimensions. Descriptive statistics for contextual predictors appear in Table 17. As explained above, these were based on 2011-2015 ACS data at the census block group level re-allocated to police beats.

Table 16 Descriptive statistics: Independent variables

	Variable	N	MIN	MAX	MEAN	SD	MED.
Black non-Hispanic civilian (d)	dblack	50,715	0	1	0.719	0.450	1
Hispanic civilian (d)	dhisp	50,715	0	1	0.197	0.397	0
White civilian (*) (d)	dwhite	50,715	0	1	0.085	0.279	0
Male civilian (d)	dmale	50,715	0	1	0.863	0.344	1
Age in years (*)	age2	50,715	7	92	31.221	13.908	27
Age in years (centered) (*)	c_age2	50,715	-24.221	60.779	0.000	13.908	-4.221
Age 10-17 (*) (d)	age1017	50,715	0	1	0.140	0.347	0
Age 18-25 (d)	age1825	50,715	0	1	0.316	0.465	0
Age 25-35 (d)	age2635	50,715	0	1	0.227	0.419	0
Age 36-45 (d)	age3645	50,715	0	1	0.122	0.327	0
Age 46 and up (d)	age46pl	50,715	0	1	0.195	0.396	0
Weekend (Sat, Sun) (d)	wknddum	50,715	0	1	0.271	0.445	0
Midnight to 3 AM (*) (d)	dhr0003	50,715	0	1	0.096	0.294	0
3 AM – 6 AM (d)	dhr0306	50,715	0	1	0.023	0.149	0
6 AM – 9 AM (d)	dhr0609	50,715	0	1	0.042	0.200	0
9 AM – noon (d)	dhr0912	50,715	0	1	0.121	0.326	0
Noon – 3 PM (d)	dhr1215	50,715	0	1	0.156	0.363	0
3 PM – 6 PM (d)	dhr1518	50,715	0	1	0.128	0.335	0
6 PM – 9 PM (d)	dhr1821	50,715	0	1	0.224	0.417	0
9 PM – 11:59 (d)	dhr2123	50,715	0	1	0.210	0.407	0
Vehicle stop (d)	dvehstop	50,715	0	1	0.260	0.439	0
ISR missing event no. (d)	eventmis	50,715	0	1	0.061	0.240	0

Note. (d) = binary variable; 1 corresponds to variable name, 0 otherwise. (*) = variable not used in multivariate analyses. MED = median.
Note. Monthly dummy variables for August-December not shown.
Source: July-December 2016 ISRs, CPD.

8.5.2.1 Racial and ethnic composition at the beat level
The following variables captured racial and ethnic composition:

- Percent of residential population that is Black and non-Hispanic
- Percent of residential population that is White and non-Hispanic
- Percent of residential population that is Hispanic
- Percent of population that is Asian
- Percent of population that is other than the above

Further, although descriptive statistics are not shown for them, some models might use Peterson & Krivo's (2010) categorical versions of racial composition (> 70 percent Black non-Hispanic = 1 else 0) and ethnic composition (> 70 percent Hispanic = 1 else 0).

8.5.2.2 Socioeconomic status at the beat level
Socioeconomic status was captured with an internally consistent (Cronbach's $\alpha = .91$) multi-item index using median house value, median income, the percentage of households reporting less than $20,000 in earnings (reversed), and the percentage of households reporting greater than $60,000 in earnings.

8.5.2.3 Residential stability at the beat level
Residential stability was captured with an internally consistent (Cronbach's $\alpha = .73$) multi-item index using percentage owner occupied households; percentage of all households, currently owner occupied, where the household moved in before 1990; and percentage of all households, currently renter occupied, where the household moved in before 2000.

8.5.3 Multicollinearity concerns with predictors
Problems can arise in regression when some predictors in the model correlate too strongly with one another. This is a well-known potential problem of multicollinearity in regression models (Darlington, 1990; Gordon, 1968). [25]

There are no multicollinearity problems if predictors are just stop level predictors. All variance inflation factors (VIFs) are below 4.0 and all tolerances are above .3. [26]

This is not true, however, among the contextual predictors. As a set of predictors, several (percent Black non-Hispanic, percent Hispanic, SES index) have variance inflation factors above 4 and tolerances of below .30. It is also not true if stop and beat level predictors enter a model together. This results in VIF values up to 13, due largely to strong correlations between detainee race and beat racial composition.

---

[25] Collinearity is a problem because it increases the standard errors associated with coefficients, making it more difficult to find statistical significances; further, in extreme cases, it can cause coefficients to "flip" to an opposite direction, a problem Gordon (1968) calls "beta bounce."
[26] Tolerance reflects the portion of the predictor that is independent of the other predictors. Tolerance greater than .30 means that at least 30 percent of the variation in that predictor is not shared with other predictors.

Table 17 Contextual predictors: Police beats

		n	min	max	mean	sd	med
Percent Black NH	p_blnonh	273	0.050	98.443	43.888	40.477	24.939
Percent White NH (*)	p_whnonh	273	0.000	87.307	26.904	28.438	12.467
Percent Hispanic	p_hisp	273	0.059	98.015	22.611	26.896	9.643
Percent Asian	p_asian	273	0.000	72.152	4.827	7.760	1.488
Percent Other	p_other	273	0.000	5.331	1.770	1.269	1.515
Residential SES	sesindx2	273	-1.732	1.749	0.041	0.768	0.017
Residential stability	stabindx	273	-1.914	2.259	0.000	0.806	-0.080

Source: 2011-2015 American Community Survey data from U.S. Census. Census block group data re-allocated to CPD beats.
Note: (*) = not used in contextual models. sd = standard deviation. med=median
Residential socioeconomic status (SES) index (Cronbach's α = .91) is the average of the following z scored items: median household income (logged), median house value (logged), percent reporting household income less than $20,000 (reversed), and percent reporting household income greater than $60,000. A higher score indicates higher socioeconomic status. Residential stability index (Cronbach's α = .73) is the average of the following z scored items: percent owner occupied households, percent of all households where current owner-occupied household moved in before 1990, and percent of all households where current renter occupied household moved in before 2000.

8.6 Analytics: Rationale and details

8.6.1 Outcomes where there is no necessary selection process

8.6.1.1 Regression models: Binary outcomes

Non-conditioned outcomes, that is outcomes where a prior selection process is not logically needed, included:

- whether a pat down took place;
- whether a search took place;
- whether a pat down occurred in a stop in which no enforcement action took place

For the first two outcomes, mixed effects logistic regression models were run.

The main models reported were three level models: stops were nested within beats, and beats in turn were nested within districts. Models also were run using just clustering by beat, and just clustering by district. These are not reported unless their results were discrepant for racial and ethnic predictors.

Where the data permitted, random variation across beats-within-districts, and across districts, were both allowed. Models were run separately for the two random samples.

For each outcome, models with three sets of predictors were run.

(1) (detainee model) The model used only stop level predictors. Of key interest were impacts of race and ethnicity of detainee on the outcomes.
(2) (beat model) The model used only contextual predictors. Of key interest were impacts of racial and ethnic composition of the beat where the stop took place on the outcomes.
(3) (combined model) The model used both step level and contextual predictors. Of key interests were race and ethnicity of the detainee, and racial and ethnic composition of the beat where the stop took place.

Impacts of race, ethnicity, beat racial composition, and beat ethnic composition are reported only for the combined model. However, if the significance pattern for a specific predictor of key interest differed between the three types of models, more details are reported.

As noted above, since contextual predictors have a high degree of multicollinearity, tests of race and ethnicity, or racial and ethnic composition, have lower statistical power. Therefore, when looking at **any models with beat-level contextual predictors included**, ***non*-significant findings of race or ethnicity of detainee, or of race or ethnic composition of the beat, should be interpreted *with extreme caution.*** [27]

Regardless of type, each three-level model controlled both for district-level random variation on the outcome, and for within-district-cross-beat random variation on the outcome.

8.6.1.2 *Propensity score matching models.*

Separate propensity score matching models were run for each of the two key contrasts: White non-Hispanic versus Black non-Hispanic detainees; and White non-Hispanic vs. Hispanic detainees.

Propensity score matching models used the exact same set of predictors used in the multiple regression models, except that race or ethnicity necessarily were treated differently.

The propensity score matching (PSM) models followed this sequence. First, the third ethnoracial group not involved in the contrast was dropped. Then, three level logistic regression models used all the predictors employed in the regression model, except the race or ethnicity contrast in question, to predict that racial or ethnic contrast. These regression models created propensity scores, for example, the propensity of a detainee to be Black non-Hispanic, given his/her age and gender; given stop features such as time, day of the week, and month; and given features of the beat where the stop took place including residential stability, residents' socioeconomic status, percent Black non-Hispanic, percent Hispanic, percent Asian, and percent other races/ethnicities. These features are generally called observed covariates. Using these propensity scores, a matching algorithm found one detainee who was White non-Hispanic whose propensity score, based on the observed covariates, was closest to the Black non-Hispanic detainee being matched, or the Hispanic detainee being matched. Sometimes more than one non-Hispanic White detainee

---

[27] Here is why.

> Collinearity affects only the power of tests on regression slopes – not their validity. The standard errors of the partial regression slopes are increased for collinear variables. This widens the confidence bands on those values of b_j, and makes it harder to find statistically significant values of b_j. But a significant value of b_j is just as conclusive when collinearity is present as when it is absent (Darlington, 1990: 130).

had the closest match to the non-White detainee in question. After doing the matching up, *only* the non-Hispanic White detainees who were matched to a non-Hispanic Black detainee in the White vs. Black analysis, or matched to a Hispanic detainee in the White vs. Hispanic analysis, were kept. All other detainees were discarded. At that point, the analysis, another three level mixed effects logistic regression analysis, was done trying to predict the outcome in question, like a pat down.

Matching was done using a caliper matching protocol. Matched cases needed to be within 6 100^{ths} of a standard deviation on the propensity score.

The prediction model for the outcome controlled for the observed covariates in a different way than did the plain mixed effects logistic regression. It simply discarded stops involving Whites where stop, detainee, and surround features were dissimilar from the features of the non-White cases in question.

Examining diagnostics for the PSM models indicated whether the approach successfully matched on observed covariates, and the degree to which selection on *unobserved* covariates remained a concern.

8.6.1.3 Regression model: Categorical outcome

For the last of these outcomes, a series of multinomial mixed effects logit models were run. As with the other outcomes, different models included detainee, context, or both sets of predictors.

The alternative analysis applied to the multinomial outcome was a canonical discriminant function analysis.

8.6.2 Outcomes where there is sequential selection

Two outcomes – whether a search generated a weapon, and whether a pat down generated a weapon -- were observed only if something prior took place. This brings up the problem of sequential selection (Berk, 1983).

Conceptually, the issue is that there are multiple layers of selection that prove relevant to an outcome like weapons recovered or not because of a pat down (Taylor, 2017). See Figure 1. Officers select where to patrol, and when, which can lead to racial and ethnic disparities in the residential composition of the places patrolled vs. not patrolled. Officers select whom to stop from among those they view passing in vehicles and on foot. This can lead to racial and ethnic disparities in the composition of those stopped vs. passing by but not stopped. No data are available to gauge these two selection dynamics, so dashed lines are shown.

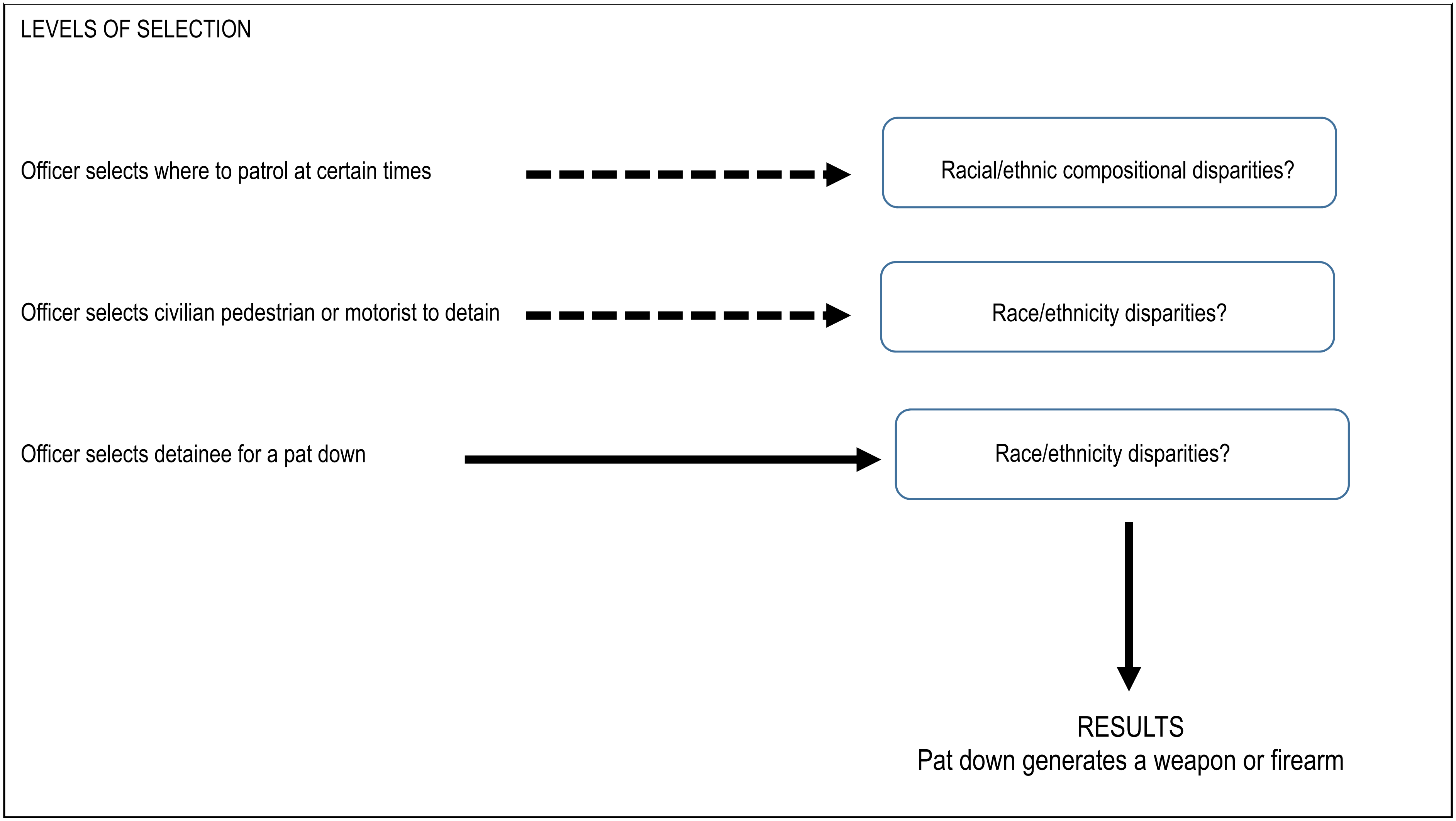
Once a detainee is stopped, the officer selects whether to pat down or frisk the person. This, as results will show, leads to disparities.

Finally, there is the question of whether the pat down results in a weapon or firearm being recovered. If racial disparities surface at that point, usually in these studies in the form of a lower weapon hit rate for Black non-Hispanic or Hispanic detainees, as compared to White non-Hispanic detainees, then the logic is that the race or ethnicity of the detainee, or some feature of the situation, some race- or ethnicity-linked unobserved covariate, led to frisks or searches being less likely to result in the officer finding something relevant. For example, Hannon found that frisks of Black detained pedestrians in Philadelphia had about 20 percent lower odds of [finding weapons or drugs vs. not finding weapons or drugs].

The frisk or pat down outcome, however, only can be gauged if the detainee is selected for a pat down in the first place. Scholars (Berk, 1983; Bushway, Johnson, & Slocum, 2007; Bushway & Reuter, 2008) argue that this selection dynamic itself deserves its own *conceptual* and *empirical* consideration.

A problem arises, however, because theory in this area does not tell us which predictors go where, outside from requiring that individual race or ethnicity, or location racial or ethnic composition, be included in the hit part of the model (Taylor, 2017). How do we decide *theoretically* which factors go into the selection portion of the model, and which factors go into the main portion? When modeling search hits among detained pedestrians in Philadelphia, Taylor (2017) found that decisions about which variables to include in the selection part of the model and which to add to the hit part of the model altered whether the results showed a significantly lower hit rate for more predominantly Black non-Hispanic locations. Consequently, here, different varieties of selection models will be used to see if placing predictors in different places in the model affects racial or ethnic disparities in the hit outcome.

Figure 1 Levels of selection (Dashed line indicates data not available)



9 RESULTS: OUTCOMES WITH NO SELECTION

9.1 PAT DOWN

9.1.1 Regression models

Impacts of detainee race and ethnicity, and beat-level racial and ethnic composition, appear in Table 18.

Detailed results are shown only for detainee-level and beat-level race and ethnicity. Additional predictors not shown at the stop level included detainee gender and age, whether it was a vehicle or pedestrian stop, whether event number was missing, whether the stop occurred on a weekend, and the month and time of day of the stop. Additional predictors not shown at the beat level included socioeconomic status, residential stability, percent Asian population in the beat, and percent other races in the beat.

Results from both random samples show a significant net impact of detainee race on the chances of being patted down. Controlling for other beat and stop factors, and for random variation across beats and districts, Black non-Hispanic detainees' chances of being [patted down vs. not patted down] were 29 percent higher in the first random sample (OR = 1.296), and 52 percent higher in the second random sample (OR = 1.52). In both samples these disparities between Black non-Hispanic and White non-Hispanic detainees' chances of receiving a pat down proved statistically significant ($p < .001$). This means that a Black-White disparity like this would occur *just* due to chance variation less than one time in a thousand.

Detainee ethnicity produced a statistically significant impact in only one of the two random samples. In the second sample, but not the first, Hispanic detainees' odds of being [patted down vs. not patted down], compared to the odds for White non-Hispanic detainees, proved significantly ($p < .01$) higher (OR = 1.22). In the first random sample, the same difference was not statistically significant (OR = 1.06).

Table 18 Impacts of detainee and contextual race and ethnicity on pat down

Sample							95 % Confidence Interval			
							b		OR	
	Fixed effects	OR	b	se of b	z test	p <	LCL	UCL	LCL	UCL
1	% Black NH (p_blnonh)	1.0061	0.0060	0.0029	2.1000	.05	0.0004	0.0117	1.0004	1.012
	% Hispanic (p_hisp)	1.0108	0.0108	0.0029	3.6900	.001	0.0050	0.0165	1.005	1.017
	Detainee Black NH (dblack)	1.2963	0.2595	0.0700	3.71	.001	0.1223	0.3967	1.130	1.487
	Detainee Hispanic (dhisp)	1.0573	0.0557	0.0732	0.76	0.447	-0.0878	0.1993	0.916	1.221
	Random effects									
		variance	se of variance							
	District	0.1580	0.0571							
	Beat within district	0.0938	0.0161							
Wald χ² (df=28) = 2143.17; p < .001; Number of observations = 25,218 (22 districts, 270 beats within districts)										
	Fixed effects									
2	% Black NH (p_blnonh)	1.0036	0.0036	0.0029	1.2500	ns	-0.0020	0.0092	0.9980	1.009
	% Hispanic (p_hisp)	1.0084	0.0084	0.0029	2.8700	.01	0.0027	0.0141	1.003	1.014
	Detainee Black NH (dblack)	1.5238	0.4212	0.0710	5.94	.001	0.2821	0.5603	1.326	1.751
	Detainee Hispanic (dhisp)	1.2203	0.1991	0.0740	2.69	.01	0.0541	0.3441	1.056	1.411
	Random effects									
		variance	se of variance							
	District	0.1275	0.0466							
	Beat within district	0.1057	0.0169							
Wald χ² (df=28) = 2057.38; p < .001; number of observations = 25,198 (22 districts, 270 beats within districts)										

Note. Data from July-December 2016. NH = non-Hispanic. Results from three level mixed effects models with stops nested within beats, and beats nested within districts. Random effects permitted at both the beat and district levels. Specific predictors only allowed fixed effects. Source: CPD ISR data and ACS census data.
Note. Outcome = pat down occurred (1) or did not (0). Additional stop level variables included as predictors were: age, gender, time of day, weekend day, month, vehicle stop, and event number missing. Details on these additional predictors not shown. Additional contextual variables included as predictors were: percent Asian, percent other races/ethnicities, socioeconomic status index, and residential stability index. Details on these additional predictors also not shown. Contextual variables are at the beat level. OR = odds ratio; b = coefficient
Note. In Sample 2, impacts of racial composition (p_blnonh) statistically significant in model with only contextual predictors.

Turning to contextual predictors describing police beats, pat downs are more likely in more predominantly Hispanic locations in *both* random samples ($p < .001$ in Sample 1; $p < .01$ in Sample 2). Each additional percentage Hispanic composition increased the odds of getting [patted down vs. not patted down] by about one percent (OR = 1.01 in Sample 1; OR = 1.008 in Sample 2).

Racial composition, the percentage of the beat residential population that was Black non-Hispanic, significantly affected pat down chances in the first random sample ($p < .05$), but not the second random sample. In a Sample 2 model with only contextual predictors, racial composition did have a significant and positive impact on pat down chances ($p < .05$). This discrepancy between racial composition results in the contextual model in Sample 2 vs. the full model arises probably from the multicollinearity concerns described above.

In sum, focusing just on consistent results across both samples, pat downs were significantly more likely:

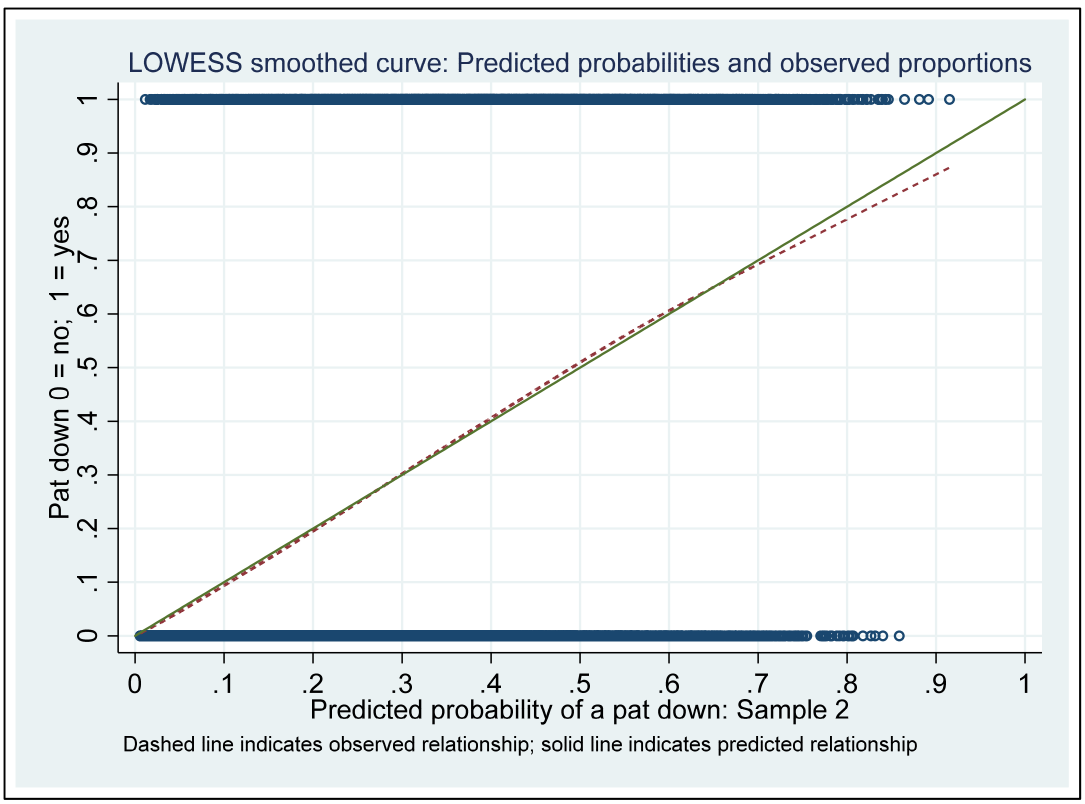
- if the detainee was Black and non-Hispanic rather than White and non-Hispanic; or
- if the stop took place in a predominantly Hispanic police beat.

9.1.2 Regression model diagnostics

Only one regression diagnostic is undertaken here. Model fit is gauged by "comparing predicted probabilities to a moving average of the proportion of cases that are one [on the outcome]" (Long & Freese, 2006: 156).

For both random samples, LOWESS plots revealed that predicted probabilities deviated from observed probabilities when predicted probabilities climbed above about .7. Above this value, the fractions of observed cases were lower than the predicted probabilities of cases expected in this range. This discrepancy appeared for both random samples. The discrepancies are shown for Sample 2 model results in Figure 2. These discrepancies suggest potential concerns about the model. In short, it does not fit the data as well at the higher end of the predicted probabilities.

Figure 2 Predicted probabilities fit to observed proportions: pat down outcome, sample 2



9.1.3 Caliper matched propensity score models: Non-Hispanic Black vs. matched White civilians

Caliper matching propensity score models provide an alternative analysis (Apel & Sweeten, 2010; Austin, 2011; Guo & Fraser, 2015).

9.1.3.1 Key results

Table 19 shows results of the model predicting a pat down using just matched Black and White non-Hispanic stops. The predicted pat down rate is significantly higher for Black non-Hispanic detainees as compared to matched White non-Hispanic detainees ($p < .01$ in Sample 1; $p < .001$ in Sample 2). Given the statistical significance, the difference in the pat down rate between these two groups is unlikely to have occurred just due to chance variation between the two groups of detainees.

For Sample 1, the predicted pat down rate was 4.8 percent higher for Black detainees (26.5 vs. 21.7 percent) and in Sample 2 the predicted pat down rate was 7.3 percent higher for Black detainees (26.1 vs. 18.8 percent).

Alternatively, Black detainees' odds of [being patted down vs. not patted down] were 30 percent higher than matched White detainees in Sample 1, and 52 percent higher in Sample 2.

Table 19 Pat down outcome: White and Black propensity score caliper matched samples

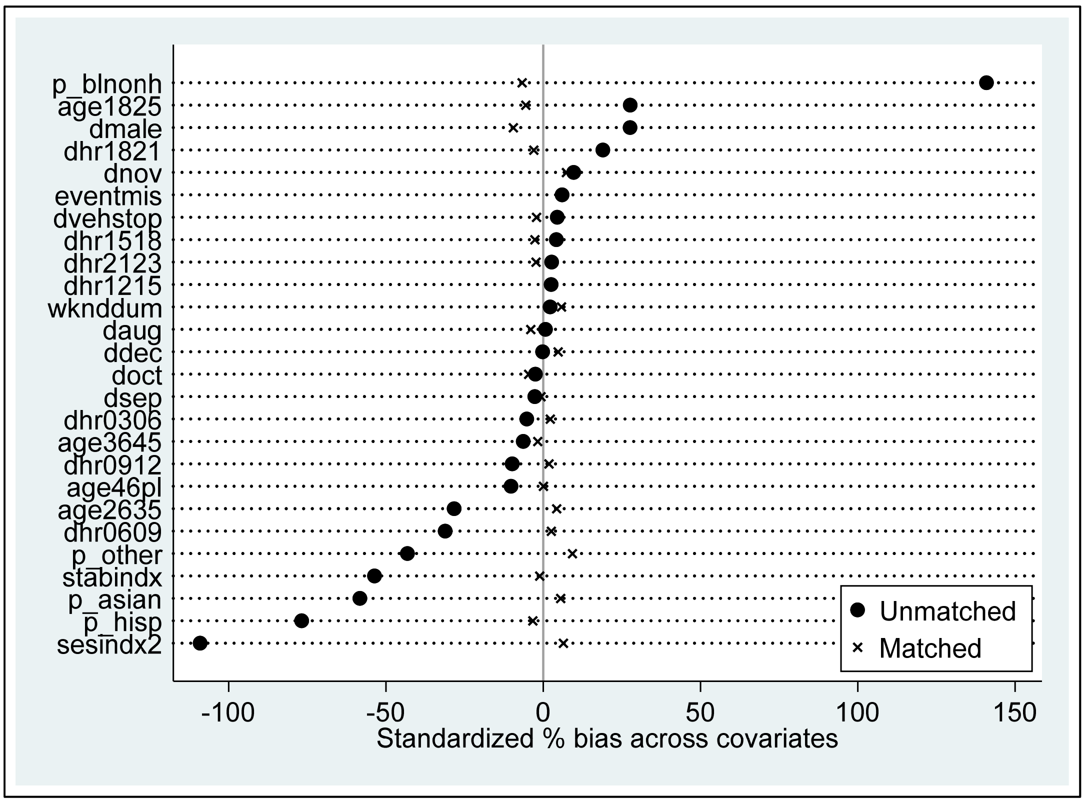
Sample			OR	b	se	z	p <	95% LCL of b	95% UCL of b
1	Fixed effects								
		Black NH	1.3029	0.2646	0.0856	3.09	.01	0.0967	0.4324
		Constant (White NH)	0.2768	-1.2844					
	Random effects			Variance	se				
		District		0.2705	0.1157				
		Beat within district		0.3822	0.0858				
Wald χ2 (df=1) = 9.54; p < .01									
n=3,749									
			OR	b	se	z	p <	95% LCL of b	95% UCL of b
2	Fixed effects								
		Black NH	1.5222	0.4201	0.0844	4.98	.001	0.2547	0.5856
		Constant (White NH)	0.2321	-1.4605					
	Random effects			Variance	se				
		District		0.0911	0.0517				
		Beat within district		0.2257	0.0632				
Wald χ2 (df=1) = 24.75; p < .001									
n=3,798									

Note. Source: CPD ISR data, Period 2 (July-December, 2016). Matched cases based on .06 SD caliper matching protocol, only matched pairs retained. Sometimes a stop of a Black non-Hispanic detainee has more than one matching White non-Hispanic detainee with a comparably close propensity score. Results of three level (district and beats within district) mixed effects logistic regression model (outcome = 1 if pat down; 0 if no pat down). Hispanic detainees excluded.

9.1.3.2 Diagnostics

The caliper matching protocol used to select matched cases did a good job removing bias between the two groups on the observed covariates. See Figure 3. Covariates are sorted, top to bottom, starting with those where Black non-Hispanic detainees scored much higher than White non-Hispanic detainees, like the percent of the residential population in the beat that was Black non-Hispanic, down to those covariates where Black non-Hispanic detainees scored much *lower* than White non-Hispanic detainees, like the socioeconomic status of the beat residents. The dot for each covariate shows how much bias there was when the two groups were compared on each covariate.

Figure 3 Covariate bias before and after matching: Black non-Hispanic and White non-Hispanic Sample 1 detainees



Note. Source: CPD ISR data. Period 2, July-December 2016. Two groups are Black-non-Hispanic detainees and White non-Hispanic detainees. Matching used caliper matching within .06 of a standard deviation. Results from Sample 1 only. Sample 2 results closely comparable.

After matching, however, the bias on observed covariates between the two groups of detainees was extremely low. This is shown in the figure by the “x”s. These show the bias between the two groups after matching. Standardized bias on each covariate is less than ten percent. Overall balance statistics, Rubin's B and Rubin's R were within acceptable ranges. [28] The implication is that selection bias, working through observed covariates, is not a concern as an alternative explanation for these differences between the two groups.

Sensitivity to unobserved selection bias, however, is still a potential concern as an alternative explanation. [29] The two groups of detainees still did differ markedly on their propensity-to-be-Black-and-non-Hispanic scores, although the differences were far less marked than they were before matching (details not shown). Table 20 shows critical gamma values from the Mantel-Haenzel (1959) test.

Table 20 Sensitivity analysis, propensity score models, pat down outcome, Black vs. White non-Hispanic civilians

		Gamma (Γ) value where race impact becomes non-significant
Caliper match = .06		
	Sample 1	1.15
	Sample 2	1.30

In the first random sample, if two individuals who were similar on the observed covariates differed in their odds of being Black and non-Hispanic versus White and non-Hispanic by only about 15 percent, then there was no significant impact of race on the pat down outcome. Given that this value of gamma (Γ) is relatively close to 1.0, the significant race impact seen is "sensitive to unobserved selection bias" (Aakvik, 2001: 30). Results from the second random sample are less sensitive to unobserved selection bias because the two individuals similar on observed covariates would have to differ in their odds of being Black vs. White by at least 30 percent.

In sum, especially from the first random sample, unobserved selection bias is a potential alternative explanation of the differences observed. Stated differently, there was something else going on, some feature of the detainee, the stop itself, the stop context, or all three, that is not captured by the variables used here, that might be "behind" the significant race differences observed at least in Sample 1.

9.1.4 Caliper matched propensity score models: Hispanic vs. matched White civilians

9.1.4.1 Key results

Table 21 shows the results from the three level mixed effects model using only matched Hispanic and White non-Hispanic detainees. In both random samples Hispanic detainees were significantly more likely ($p < .05$ in Sample 1, $p < .001$ in Sample 2) to get a pat down. There was a predicted 3 percent difference in Sample 1 with a predicted pat down rate of 23.7 percent

[28] For more details see section 10.1.2.3 in Taylor & Johnson (2017)
[29] For more details see section 10.1.2.4 in Taylor & Johnson (2017)

for Hispanic detainees vs. 20.7 percent for matched White non-Hispanic detainees. There was a 7.2 percent difference in Sample 2 with a predicted pat down rate of 25.7 percent for Hispanic detainees and 18.5 percent for matched White non-Hispanic detainees.

In short, in both samples, stopped Hispanic civilians were significantly more likely to get a pat down compared to matched White non-Hispanic civilians. These differences were unlikely to have arisen from chance variation in the data.

Table 21 Pat down outcome: White and Hispanic propensity score caliper matched samples

Sample			OR	b	se	z	p <	95% LCL of b	95% UCL of b
1	Fixed effects								
		Hispanic	1.1917	0.1754	0.0864	2.03	.05	0.0061	0.3447
		Constant (White NH)	0.2613	-1.3421					
	Random effects			Variance	se				
		District		0.1486	0.0724				
		Beat within district		0.2507	0.0657				
Wald χ2 (df=1) = 4.21; p < .05									
n=3,621									
			OR	b	se	z	p <	95% LCL of b	95% UCL of b
2	Fixed effects								
		Hispanic	1.5215	0.4197	0.0858	4.89	.001	0.2516	0.5878
		Constant (White NH)	0.2268	-1.4836					
	Random effects			Variance	se				
		District		0.0979	0.0572				
		Beat within district		0.2080	0.0626				
Wald χ2 (df=1) = 23.94; p < .001									
n=3,652									

Note. Source: CPD ISR data, Period 2 (July-December, 2016). Matched cases based on .06 SD caliper matching protocol, only matched pairs retained. Sometimes a stop of a Black non-Hispanic detainee has more than one matching White non-Hispanic detainee with a comparably close propensity score. Results of three level (district and beats within district) mixed effects logistic regression model (outcome = 1 if pat down; 0 if no pat down).

9.1.4.2 Diagnostics

Figure 4 shows covariate bias before and after matching for Sample 2, the sample providing the strongest contrast on the outcome between the two groups of interest. After matching, the standardized bias on each covariate was less than |10| percent. Further, for both samples the bias statistics summarizing across all the covariates were within acceptable ranges (Rubin's B, Rubin's R).

Given the excellent matching on observed covariates, selection on observed covariates does not seem to be a significant concern.

That said, the two groups of detainees still did differ markedly on their propensity-to-be-Hispanic scores, although the differences were far less marked than they were before matching (details not shown). The Mantel-Haenzel bounds test similarly suggested concern. See Table 22. In Sample 2, which generated the most marked difference in predicted pat down rates, a gamma (Γ) value of just 1.15 rendered the differences non-significant. Stated differently, in the second random sample if two individuals who were similar on the observed covariates differed in their odds of being Hispanic versus White and non-Hispanic by only about 15 percent, then there was no significant impact of ethnicity on the pat down outcome for this sample.

Figure 4 Covariate bias before and after matching: Hispanic and White non-Hispanic Sample 2 detainees

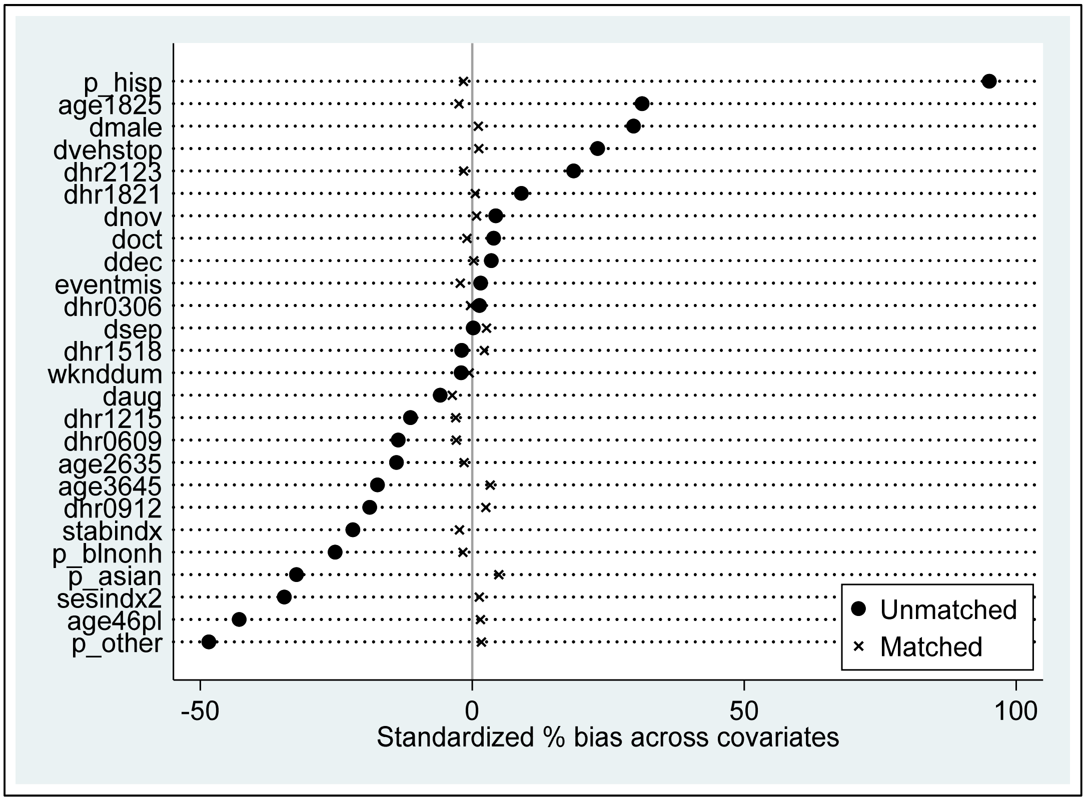


Table 22 Sensitivity analysis, propensity score models, pat down outcome, Black vs. White non-Hispanic civilians

	Gamma (Γ) value where race impact becomes non-significant
Caliper match = .06	
Sample 1	1
Sample 2	1.15

9.1.5 Pat down: Summary

For the Black vs. White non-Hispanic contrast:

- Regression models for both samples suggest significant impacts of detainee race (p s < .001), with Black non-Hispanic detainees more likely than White non-Hispanic detainees to receive a pat down.
- Using propensity models to narrow the focus just to comparable Black non-Hispanic and White non-Hispanic detainees confirmed the significant differences in both samples (p < .01 Sample 1, p < .001 Sample 2). For these matched cases, the predicted differences in the pat down rate ranged from 5 to 7 percent depending on the sample.
- Although the propensity models seemed to take care of concerns about selection on observed covariates, selection on unobserved covariates remained a concern, at least in one sample.

For the Hispanic vs. White non-Hispanic contrast:

- Regression models found significant differences only in Sample 2 (p < .001).
- Looking just at comparable Hispanic and White non-Hispanic detainees, the propensity models showed significant differences in the chances a pat down would take place for both random samples (p < .05 in Sample 1; p < .001 in Sample 2). The predicted difference in pat down rates was 3 percent in Sample 1 and 7 percent in Sample 2.
- Although the propensity models seemed to take care of worries about selection on observed covariates, selection on unobserved covariates as a potential alternative explanation remains a concern.

9.2 SEARCH

9.2.1 Search and arrest exclusions

CPD officers are required to conduct a search prior to taking a civilian into custody for an arrest or transport.

This dramatically reduced the volume of searches examined by roughly two thirds in each sample.

In the first random sample there were 3,427 searches in 25,325 stops. After removing stops with arrests (n=3,040), only 1,012 searches remained among 22,285 stops. [30]

The corresponding numbers in the second random sample were 3,558 searches among 25,315 stops. After removing stops with arrests (n=3,088) there were 1,049 searches among 22,227 stops.

Therefore, after removing stops involving arrests, searches occurred in only about four to five percent of stops.

[30] Models with beat predictors exclude stops taking place outside the city (District 31) so the total number of searches shown here across the two samples will not match the total shown in Table 3.

The reason for dropping these searches was that some of these searches were *incident* to taking the arrestee into custody. Removing these is appropriate. The officer did not decide whether to search but rather just followed department procedures in these cases. But there were other stops where the officer decided to do a search, did such a search, and based in part on what turned up in the search, decided to arrest. One can argue that removing the latter group of searches was *inappropriate*. Such an inappropriate exclusion may render non-significant what would otherwise have been a significant net impact of race or ethnicity.

This is plausible. This concern represents a significant limitation of our search analyses.

9.2.2 Regression models

Results appear in Table 23. They tell a simple story. Among the variables of key interest, none shows a significant impact on the outcome in either of the two random samples.

In the first random sample, searches were more likely to take place in beats where the population was more predominantly Hispanic. This significant impact, however, did not re-appear in the second random sample. Given the inconsistent results across samples, the ethnic composition impact on whether a search happened is not robust.

9.2.3 Regression model diagnostics

As before, model fit is gauged by "comparing predicted probabilities to a moving average of the proportion of cases that are one [on the outcome]" (Long & Freese, 2006: 156). Results for the two random samples appear in Figure 5 and Figure 6. Observed scores, on the vertical (y) axis, of course can only be 0 (no search) or 1 (search). Predicted probabilities that a search took place for a stop, are arranged from low to high, left to right on the horizontal (x) axis. The two variables are connected by the dashed line, which represents the LOWESS smoothed curve in each figure. It "shows the fraction of observed cases that equal 1 [search took place] at each level of the model's predicted probability of observing a [search]" (Long & Freese, 2006: 157). If the model was predicting well at all levels of predicted probabilities, the LOWESS curve, the dashed line, would track close to the expected relationship, shown by the solid line, at all different levels of predicted probabilities.

The model diagnostics in each sample reveal concerns.

In Sample 1, above predicted search probabilities of about .3, the fractions of observed stops with a search were lower than the predicted fractions of stops with a search. The dashed line (observed relationship between predicted probabilities and observed probabilities) starts to deviate below the solid line above .3 on the horizontal (x) axis. Above predicted probabilities of .4, there were no observed stops where a search took place, i.e., the "bubbles" at 1 on the vertical (y) axis stop.

The reverse happens in Sample 2. Here, the dashed line starts to curve upward and away from the solid line above predicted search probabilities of around .2. So here, the fractions of observed stops with a search were *higher* than the predicted fractions of stops with a search above this predicted search probability.

In both samples, these discrepancies suggest potential concerns about the model. In short, it does not fit the data well, in either sample, for predicted search probabilities in the moderate to high range.

Table 23 Impacts of detainee and contextual race and ethnicity on search

Sample							95 % Confidence Interval			
							b		OR	
	Fixed effects	OR	b	se of b	z test	p <	LCL	UCL	LCL	UCL
1	% Black NH (p_blnonh)	1.0056	0.0055	0.0044	1.2600	ns	-0.0031	0.0142	0.997	1.014
	% Hispanic (p_hisp)	1.0095	0.0094	0.0045	2.0800	0.05	0.0006	0.0183	1.001	1.018
	Detainee Black NH (dblack)	1.1149	0.1088	0.1460	0.7400	ns	-0.1774	0.3949	0.837	1.484
	Detainee Hispanic (dhisp)	1.1536	0.1429	0.1517	0.9400	ns	-0.1544	0.4402	0.857	1.553
	Random effects									
			variance	se of variance						
	District		0.1427	0.0603						
	Beat within district		0.0905	0.0379						
Number of observations = 22,189										
	Fixed effects									
2	% Black NH (p_blnonh)	1.0005	0.0005	0.0040	0.1200	ns	-0.0074	0.0083	0.993	1.008
	% Hispanic (p_hisp)	1.0037	0.0037	0.0041	0.9100	ns	-0.0043	0.0117	0.996	1.012
	Detainee Black NH (dblack)	1.0738	0.0712	0.1390	0.5100	ns	-0.2013	0.3436	0.818	1.410
	Detainee Hispanic (dhisp)	0.9101	-0.0942	0.1456	-0.650	ns	-0.3796	0.1913	0.684	1.211
	Random effects									
			variance	se of variance						
	District			0.0966	0.0419					
	Beat within district			0.0723	0.0293					
Number of observations = 22,127										

Note. Data from July-December 2016. All investigatory stops resulting in an arrest have been removed from each sample. NH = non-Hispanic. Results from three level mixed effects models with stops nested within beats, and beats nested within districts. Random effects permitted at both the beat and district levels. Specific predictors only allowed fixed effects. Source: CPD ISR data and ACS census data. Outcome = search occurred (1) or did not (0). Results from three-level mixed effects logit models, with stops nested within beats and beats nested within districts. Additional stop level variables included as predictors, but now shown, include: age, gender, time of day, weekend day, month, vehicle stop, and event number missing. Additional contextual variables included as predictors but not shown include: percent Asian, percent other races/ethnicities, socioeconomic status index, and residential stability index. Contextual variables are at the beat level. OR = odds ratio; b = coefficient

Figure 5 Search regression model diagnostics: Sample 1

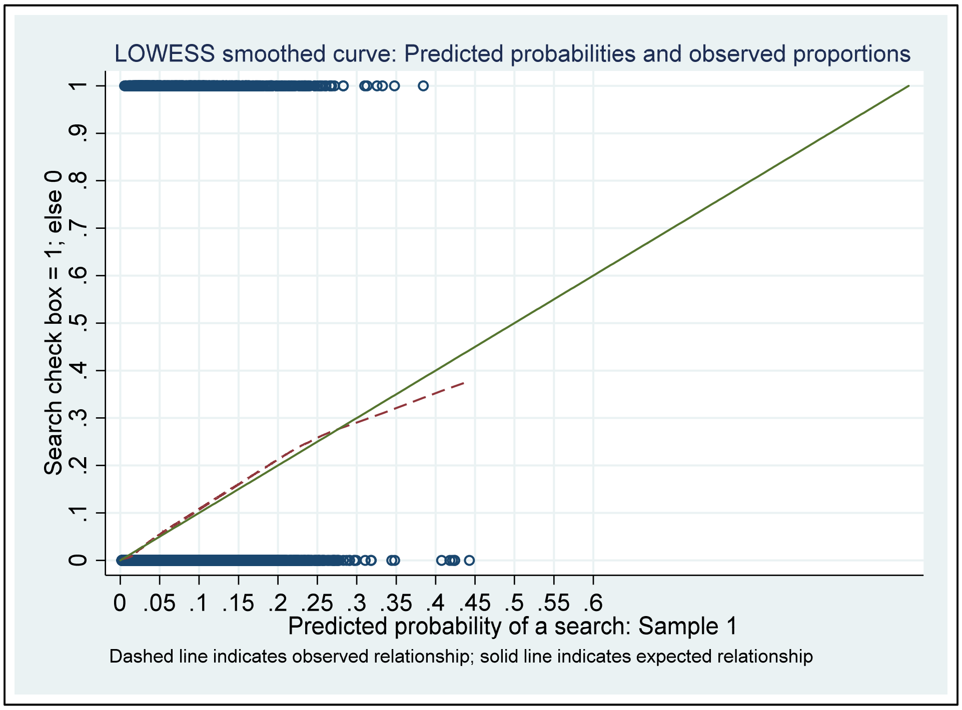
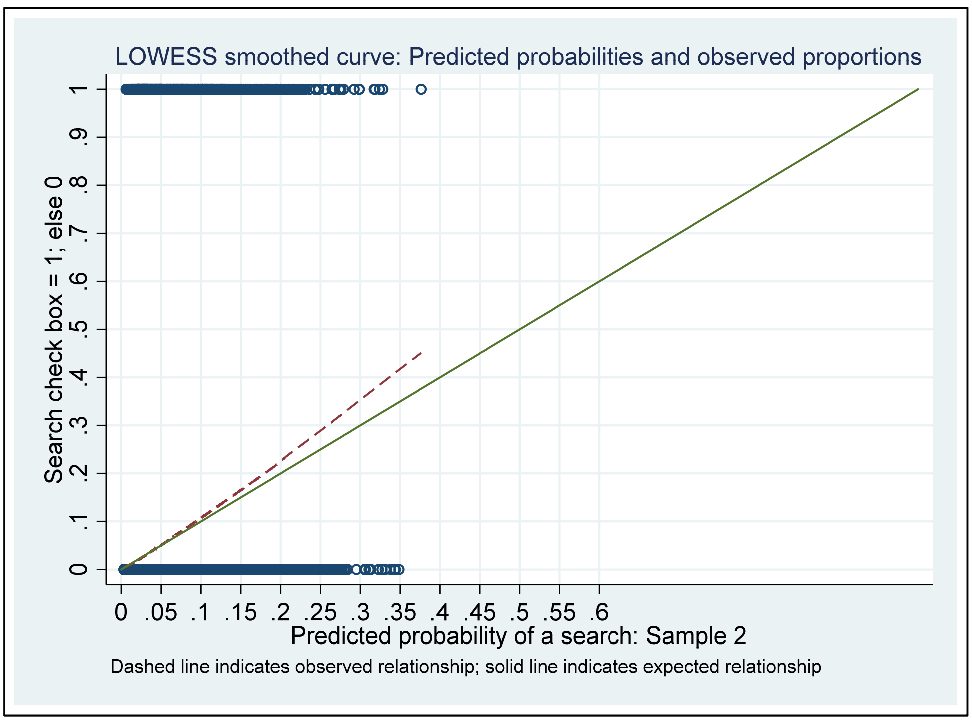


Figure 6 Search regression model diagnostics: Sample 2



9.2.4 Propensity score models

Given that neither detainee race nor detainee ethnicity showed a significant net impact on whether a search took place, for either sample, propensity score models, since they provided an even more stringent test of race or ethnicity differences, were even less likely to show significant differences between the two ethnoracial groups when only matched cases were considered. This was checked using Sample 1 data (detailed results not shown). No discernible differences emerged between matched White and Black non-Hispanic detainees on the search outcome. For example, for Sample 1 $z = .56$, $p = .57$ for the impact of the Black non-Hispanic variable on the outcome.

9.3 If no enforcement took place, what determined whether a pat down took place?

9.3.1 Main modeling approach

These analyses used mixed effects multinomial models. In contrast to the models for the outcomes described so far, these models do not simultaneously nest stops within beats, and beats within districts. Rather, the models just nest stops within beats, and allow for random effects at the beat level. [31] There are four outcome categories for all possible combinations of enforcement and pat down. Results reported here focus just on two categories and one contrast: the prediction of:

[pat down and no enforcement vs. no pat down and no enforcement]

Results appear in Table 24. Three factors linked consistently to higher chances of detainees experiencing a pat down vs. no pat down in stops where police took no enforcement actions: race of detainee, ethnicity of detainee, and ethnic composition of the stop context. Again, these impacts apply *only* to stops where no enforcement action took place. [32]

Compared to White non-Hispanic detainees, Black non-Hispanic detainees had significantly higher ($p < .001$ in both samples) relative risks of experiencing a pat down. Their relative risk was 44 percent higher in the first sample and 63 percent higher in the second sample.

Compared to White non-Hispanic detainees, Hispanic detainees had significantly higher relative risks of experiencing a pat down ($p < .05$ in first sample; $p < .001$ in second sample). Their relative risk was 23 percent higher in the first sample and 37 percent higher in the second sample.

[31] Additional models using district as the grouping level were completed (results not shown). Those results were generally consistent with those shown here.
[32] They did, however, control for the other outcomes and other contrasts, which is the whole point of doing a multinomial model rather than a series of logit models (Long, 1997: 151). So all four groups were simultaneously considered.

If civilians involved in a non-enforcement action stop were stopped in more predominantly Hispanic beats, they were more likely to be patted down ($p < .001$ in first sample; $p < .01$ in second sample). Each additional percentage Hispanic residential composition increased relative risk of a pat down by about one percent (RR = 1.014 in first sample; 1.009 in second sample).

Table 24 Impacts of detainee and contextual race and ethnicity on pat down+ no enforcement vs. no pat down + no enforcement

Sample							95 % Confidence Interval			
							b		OR	
	Fixed effects	RR	b	se of b	z test	p <	LCL	UCL	LCL	UCL
1	% Black NH (p_blnonh)	1.0106	0.0105	0.0026	4.0200	.001	0.0054	0.0156	1.005	1.016
	% Hispanic (p_hisp)	1.0140	0.0139	0.0026	5.3500	.001	0.0088	0.0191	1.009	1.019
	Detainee Black NH (dblack)	1.4357	0.3616	0.0772	4.6800	.001	0.2103	0.5130	1.234	1.670
	Detainee Hispanic (dhisp)	1.2292	0.2063	0.0814	2.5400	.05	0.0469	0.3658	1.048	1.442
	Random effects									
			variance	se of variance						
	Beat		0.1763	0.0219						
Number of observations = 25,179										
	Fixed effects									
2	% Black NH (p_blnonh)	1.0051	0.0050	0.0026	1.9300	NS	-0.0001	0.0102	1.000	1.010
	% Hispanic (p_hisp)	1.0087	0.0086	0.0026	3.3200	.01	0.0035	0.0137	1.004	1.014
	Detainee Black NH (dblack)	1.6205	0.4827	0.0787	6.1300	.001	0.3285	0.6370	1.389	1.891
	Detainee Hispanic (dhisp)	1.3743	0.3179	0.0827	3.8400	.001	0.1558	0.4801	1.169	1.616
	Random effects									
			variance	se of variance						
	Beat		0.1726	0.0215						
Number of observations = 25,164										

Source: CPD ISR data and ACS census data.
Note. Data from July-December 2016. Investigatory stops with inconsistencies between the overall enforcement check box and the individual enforcement check boxes (n=73) have been removed from the analysis.
Outcome = (1) no enforcement+no pat down vs. (2) no enforcement+pat down vs. (3) enforcement+no pat down vs. (4) enforcement+pat down. Group (1) is the reference category. Results from mixed effects multinomial logit models, with stops nested within beats. Results shown only for the contrast of (2) vs. (1). Additional stop level variables included as predictors, but now shown, include: age, gender, time of day, weekend day, month, vehicle stop, and event number missing. Additional contextual variables included as predictors but not shown include: percent Asian, percent other races/ethnicities, socioeconomic status index, and residential stability index. Contextual variables are at the beat level.
RR = relative risk ratio; b = coefficient; se = standard error; NH = non-Hispanic

9.3.2 Alternative modeling approach

Canonical discriminant functions provided an alternative analytic approach. This analysis uses the same set of predictors as described above to classify individual stops into one of the four groups. The analysis was repeated for each random sample. For each sample the analysis was done with either continuous or categorical variables for percent Black non-Hispanic racial and percent Hispanic ethnic residential composition. The categorical variables used the 70 percent cutoff recommended by Peterson & Krivo (2010). Bear in mind these models are "flat" and do not recognize nesting of stops within beats and districts.

The discriminant functions generated significantly separated the four groups, or the two groups (all multivariate F's had a significance of $p < .001$). Further, when analyzing four groups, it appeared that the Black non-Hispanic variable had a sizable loading of .40 or higher on at least one function, suggesting it was contributing noticeably to discriminating between the groups.[33] Further, the univariate one-way ANOVA for the Black non-Hispanic variable always was significant ($p < .001$). But, regardless of which type of variables were used for racial or ethnic residential beat composition, the models did poorly at predicting to which of the four groups a stop belonged, save for the first group, no pat down and no enforcement. The models correctly classified over 90 percent of stops belonging to the first group. That is, over 90 percent of the stops where there actually were no pat downs and no enforcement actions were predicted to be in that group of stops. For the pat down but no enforcement stops, none of the models correctly classified more than about 20 percent of the stops in this group.

In short, these models showed that detainee race mattered for separating these four groups of stops, but did a poor job of predicting if the stop involved a pat down but no enforcement.

These alternative analytics confirm that detainee race is relevant generally to stop type if the stop is classified based on both enforcement and pat down. In that way these models provide support for the pattern seen in the multinomial models, although the support is not specific to contrasting the two key groups of interest.

10 Results: Outcomes with Selection

10.1 Pat down results in a weapon

The approach used here was a single level model or "flat" model, but it did recognize that error terms clustered at the beat level. This was a Heckman selection model for a binary (probit) outcome with robust standard errors (Baum, 2006). [34] The selection portion of the model estimated whether someone was selected for a pat down, while, simultaneously, the main portion of the model predicted whether the pat down generated a weapon. Table 25 shows the relationship between pat downs and pat down weapon hits. [35] About 3.5 percent of pat downs result in weapon recovery.

The model was set up four different ways. In each model, all the predictors that were used in main models for the pat down outcome were used to try to predict whether a pat down took place. The models varied in what was put into the weapons hit portion, or main portion, of the model.

- Only detainee race and ethnicity were included in the hit portion of the model
- Detainee race, ethnicity, gender, and age were included in the hit portion of the model

[33] This cutoff for "sizable" is arbitrary, but conforms to the tradition of not showing factor loadings in a factor analysis when those loadings are less than .4.
[34] In Stata this is heckprobit.
[35] The variable pathit_w2 is used here.

- Beat racial and ethnic composition were included in the hit portion of the model
- Beat racial and ethnic composition as well as percent Asian, percent other, SES and residential stability were included in the hit portion of the model

Table 25 Pat down outcome and pat down weapon hit outcome: Cross tabulation

		Pat down			
Pat down finds weapon or firearm		No	Yes	Missing	Total
No	N	0	14,392	0	14,392
	Percent	0	96.55	0	28.42
Yes	N	0	515	0	515
	Percent	0	3.45	0	1.02
Missing (no pat down)	N	35,733	0	1	35,734
	Percent	100	0	100	70.56
-----------		-----------	-----------	-----------	----------
Total	N	35,733	14,907	1	50,641
	Percent	100	100	100	100

Note. Source: CPD ISR data, Period 2, July-December 2016. Percentages shown are column percentages. Stops outside city (District 31) excluded.

Models of each type were run separately for each sample. Results tell a simple story. In each sample, in every model variety, neither detainee race or ethnicity, nor racial or ethnic composition of the beat, significantly affected whether the pat down generated a weapon. Table 26 shows the results of the hit portion of the model for the four different varieties, for Sample 1 only. The odds ratio associated with being Black and non-Hispanic rather than White and non-Hispanic indicated about 15-17 percent lower odds of a hit vs. no hit for weapons. But this impact always proved highly *non*significant, meaning it could just be chance variation.

In short, once the contributions of race and ethnicity of the detainee or of the beat to officers' deciding to pat down were taken into account, there were no additional racial or ethnic disparities in whether the pat down resulted in a weapon.

Table 26 Selection model predicting pat down discovers weapon or firearm: Sample 1 only

		OR	b	se (robust)	z	p =	b 95% LCL	b 95% UCL
Main model includes: only detainee race and ethnicity								
Black non-Hispanic	dblack	0.858	-0.1526	0.1101	-1.39	0.166	-0.3684	0.0632
Hispanic civilian (d)	dhisp	0.818	-0.2010	0.1316	-1.53	0.127	-0.4590	0.0569
Constant	_cons		-1.8956					
rho = .231; p < .05								
Main model includes: all detainee features								
Black non-Hispanic	dblack	0.828	-0.1888	0.1133	-1.67	0.096	-0.4109	0.0332
Hispanic	dhisp	0.798	-0.2260	0.1356	-1.67	0.096	-0.4917	0.0397
Male	dmale	1.190	0.1739	0.1485	1.17	0.242	-0.1172	0.4650
Age 18-25	age1825	0.818	-0.2011	0.0775	-2.6	0.009	-0.3530	-0.0492
Age 25-35	age2635	0.877	-0.1318	0.0949	-1.39	0.165	-0.3178	0.0542
Age 36-45	age3645	1.051	0.0493	0.1059	0.47	0.642	-0.1583	0.2569
Age 46 and up	age46pl	1.190	0.1744	0.1612	1.08	0.279	-0.1415	0.4903
Constant			-1.7971					
rho = .06, ns								
Main model includes: beat residential racial and ethnic composition								
Percent Black NH	p_blnonh	1.0003	0.0003	0.0016	0.2	0.844	-0.0027	0.0033
Percent Hispanic	p_hisp	0.9998	-0.0002	0.0019	-0.1	0.923	-0.0039	0.0035
Percent Other	p_other	1.014	0.0134	0.0277	0.49	0.627	-0.0408	0.0677
Percent Asian	p_asian	1.004	0.0036	0.0036	1.02	0.308	-0.0033	0.0106
Constant			-2.1062					
rho=.255; p < .05								
Main model includes: beat residential racial and ethnic composition, SES and stability								
Percent Black NH	p_blnonh	0.999	-0.0014	0.0023	-0.61	0.543	-0.0060	0.0032
Percent Hispanic	p_hisp	0.999	-0.0014	0.0023	-0.6	0.549	-0.0059	0.0032
Percent Other	p_other	1.020	0.0195	0.0268	0.73	0.465	-0.0329	0.0720
Percent Asian	p_asian	1.002	0.0024	0.0037	0.66	0.506	-0.0048	0.0097
Residential SES	sesindx2	0.902	-0.1029	0.0795	-1.29	0.196	-0.2588	0.0530
Residential stability	stabindx	1.094	0.0897	0.0511	1.75	0.08	-0.0106	0.1899
Constant			-2.0239					
rho=.284; p < .05								

Note. Results from Heckman probit model with robust standard errors clustered at the beat level, Period 2 ISR data from CPD. Results only shown for first random sample. N = 25,218 with 7,461 selected for a search. Robust error terms clustered by beat-within-district. ONLY main part of model shown. Selection part of model not shown. Predictors in selection always included detainee features (race, ethnicity, gender, age categories), stop features (weekend, vehicle stop, event number missing, month, time of day blocks) and beat context (percent Black non-Hispanic, percent Hispanic, percent Asian, percent other non-white non-Hispanic races/ethnicities, residential socioeconomic status, residential stability).

10.2 Search results in a weapon

Leaving out District 31, based on the search check box, 8,985 searches were conducted, and 347 of them (4.97 percent) generated a weapon or a firearm.

But if searches taking place during stops which resulted in an arrest are removed, then only 2,061 searches remained, of which only 35 (1.7 percent) resulted in a weapon being uncovered.

These were too few search weapons hits to analyze with selection models.

11 Discussion

Discussion organizes findings into two groups: changes from the first half of 2016 to the second half, and ethnoracial connections with post-stop outcomes only in Period 2, noting for the latter how these links align with those seen in the first period.

11.1 Changes and stability from first half to last half of 2016.

11.1.1 Changes

11.1.1.1 Volume

Stop volume in the last of 2016 was down 6.3 percent from the first half of the year. This is just a descriptive result. Whether that shift arises from seasonal differences between the first and last half of the year that affected pedestrian or vehicular patterns, from alterations – perhaps policy linked -- in how CPD officers acted on the streets, from changes in serious and/or disorder crimes, from other shifts in the mix of matters confronting patrolling officers, or from all of the above, is not known. Whether that shift is meaningful in a practical manner is not commented on here. Whether that shift is meaningful statistically is not assessed. That said, an interesting and important question is: what would be the appropriate volume of investigatory stops? Addressing that question links to matters far beyond the scope of this report. These would include but are not limited to community and individual detainee impacts of stops; impacts of specific stops on later nearby crime patterning; and questions about alternative uses of patrolling officers' time if those officers were freed up by engaging in fewer stops. [36] Expanding for a moment on the latter: with roughly 3,000 fewer stops and 3,000 fewer reports to write, how much officer time was freed up, if one assumes ten minutes per stop and ten minutes per report, that is 3,000 x 1/3 hours per stop, or 1,000 additional available officer discretionary hours. How is that additional time being redirected? This assumes overtime was not being used as part of making or reporting on these stops.

11.1.1.2 Pat downs

The fraction of stops with pat downs declined a statistically significant four percent from 34 to 30 percent between the first half of the year and the last half. Whether this decline is practically significant deserves discussion. Gauging the practical significance requires finding the *reason* for

[36] This question was raised by Dr. Michael White in his remarks as a discussant at a session entitled "Stop and Frisk: Some Outstanding Questions" on 15 November, 2017, at the annual meetings of the American Society of Criminology, Philadelphia, PA.

the decline in relative chances that a detainee would be patted down. Finding that reason is beyond the scope of the current effort. Again, the shift could arise from an enormous range of factors. To take just one, if the average temperature was higher in the latter half of the year as compared to the first half, detainees may have been dressed with fewer outer garments, reducing detaining officers' uncertainty about whether detainees were carrying weapons. [37] In short, the causes of the pat down drop remain unknown. That uncertainty hinders discussion of the practical and policy significance of the shift.

11.1.1.3 Pat down weapon recovery and non-recovery

A statistically significant higher rate pat down weapon recovery in the second half of the year appears when Period 2 is compared to Period 1, up to 3.5 percent from 2.5 percent. Although at face value the higher pat down weapon recovery seems a good thing, uncertainty clouds interpretation. Were civilians who were walking or driving in locations and at times when police were likely to make an investigatory stop just more likely to be carrying weapons in the second half of the year compared to the first? Or was the base rate of weapon carrying unchanged, and officers were becoming more adept at estimating whether a detainee was carrying a weapon? We don't know.

Of course, of equally clear practical and policy significance is the rate of pat downs that did *not* lead to a weapon being recovered. Looked at from that end, the non-recovery rate dropped only slightly from 97.5 percent to 96.5 percent. That probably remains a concern to many.

11.1.1.4 Enforcement, no enforcement, no enforcement+no pat down

The fraction of stops resulting in some type of enforcement action dropped a statistically significant amount from 32 to 28 percent. Practical and policy significance is unclear because, again, the reasons for the drop are unknown. Is it a good thing that stops are less likely to result in some type of enforcement? Or is it a bad thing, suggesting that police are making even more un-needed investigative stops? The answer is not clear.

Similar uncertainty surrounds the statistically significant increase from 43 to 50 percent of stops in which the officer both did *not* conduct a pat down and did *no* enforcement actions. Does this mean officers were showing more restraint in the latter half of the year? Or does it mean something else?

11.1.2 Summary

It is hard to *definitively* gauge the practical and policy significance of each of these noted changes. Further assessments, which would be beyond the scope of this effort, are needed to reduce that uncertainty. Only when that uncertainty is reduced, can we better estimate the practical and policy significance of each of these changes.

11.1.3 Consistencies

Two consistencies spanning the first and second half of calendar year 2016 stand out. First, the relative ethnoracial mix of detainees remains about the same. In both parts of the year, considering only the three major ethnoracial groups, detainees are about 70 percent Black non-Hispanic and about 20 percent Hispanic.

[37] The first author thanks Dr. Jerry H. Ratcliffe for an enlightening conversation on this point.

Additionally, geographical stability appears at the district level in the relative volume of investigatory stops. Districts with a lot of stops in the first half of the year had a lot in the second half. The same applies for districts with few stops. Of course, given the large area of each district, and what we know generally about crime continuities (Taylor, Ratcliffe, & Perenzin, 2015), such stability in investigatory stop patterns over a short term proves unsurprising.

11.2 Ethnoracial links in Period 2

The following two ethnoracial links with specific post-stop outcomes appeared a) in both random samples in the main analysis and b) in both random samples in the alternative analysis. A link replicating across two independent random halves of the data, and across different statistical models making different assumptions, suggests each connection is somewhat robust.

- Black non-Hispanic detainees are more likely to be patted down, after controlling for other factors, than are White non-Hispanic detainees.

This result appears for both random samples in the main regression models (Table 18) and in both random samples in the alternate propensity score models (Table 19). The latter seems to take care of selection on observed covariates (Figure 2). But selection on unobserved covariates as an alternative explanation remains something of a concern (Table 20) as an alternative explanation. It is, therefore, not clear the relationship is causal.

This suggested conclusion replicates the findings from the first period, where a consistent race link with the pat down outcome appeared, for both samples, across main and alternative analytics (Taylor & Johnson, 2017: Table 55 p. 116).

- Among stops where no enforcement action takes place, compared to White non-Hispanic detainees, Black non-Hispanic detainees are more likely to be patted down.

This result appears for both random samples in the main regression models (Table 24). Further, the Black non-Hispanic detainee variable had sizable loadings on the discriminant function separating the four groups considered in these analyses. Given some diagnostic concerns about the main model, and no alternative analytics providing insight into either the selection on observed or unobserved covariates problems, this link is probably best interpreted as a net correlational impact.

This finding replicates the robust link in Period 1 seen between detainee race and stops with pat downs but no enforcement (Taylor & Johnson, 2017: Table 55 p. 116).

The current effort did not propose to learn whether the *size* of each of these two links remained the same from the first to the last half of the year. It is not clear if links are stronger or weaker across the two periods. Nonetheless, they prove consistent.

11.3 Limitations and strengths

The most important limitation of the current analysis, for policy purposes, is that the framework used here does not clearly cross reference with policy concerns about disparate impact and disparate treatment (Ayres, 2002, 2010). That said, *specifically for the context of the types of post stop outcomes examined here,* the authors are not aware of accepted scholarship that does this cross referencing. White and Fradella (2016: 18-35) begin that discussion in a general way around stop and frisk issues. But further development is needed. Lacking that framework, the approach here gauged three types of ethnoracial links with outcomes: gross impacts, net impacts, and discriminating between net impacts that were correlational or causal.

A second important limitation arises from the scope of the effort. These analyses are based on archival reports. How individual reports link to what happens on the street in individual stops is not known. Lacking a systematic and expensive ridealong program with trained observers, and observed connections between on-the-street actions and archival reports, this limitation will remain. On the other hand, if such an effort were mounted it might find strong correspondences between written reports and observed encounters. So, at this juncture this matter is best considered a *potential* limitation. We just don't know if it is a problem, and if so, how big it is. This limitation is a concern for much of the stop and frisk scholarship, which relies heavily on exactly the types of reports considered here (White & Fradella, 2016).

A third limitation is that although alternative analytics are used throughout, each analysis here could be subjected to more extensive diagnostics. For the models run, one could spend more time looking "under the hood" of each model. Further, different varieties of some models could have been run. For example, we could have run propensity score models with Mahalanobis distance matching. Time and limited resources prevented this.

A fourth limitation is that results seen are specific to the predictor sets used. Different predictor sets could result in different observed impacts.

Strengths of the work are twofold. Both key findings are internally replicated across two independent random samples of the data. This means that a particular finding is not dependent just on a particular mix of cases. In addition, each outcome is analyzed with two different statistical approaches. In short, robustness across samples, and robustness across statistical assumptions both are gauged in the current work.

Finally, a strength of the current effort emerges when findings here are compared to those in the first period report (Taylor & Johnson, 2017). *The two key consistent findings observed there replicated here.*

12 References

Aakvik, A. (2001). Bounding a Matching Estimator: The Case of a Norwegian Training Program. *Oxford Bulletin of Economics and Statistics, 63*(1), 115-143. doi:10.1111/1468-0084.00211

Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: The Bonferroni vs Holm methods. *American Journal of Public Health, 86*(5), 726-728.

Apel, R. J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. In A. R. Piquero & D. Weisburd (Eds.), *Handbook of Quantitative Criminology* (pp. 543-562). New York: Springer.

Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research, 46*(3), 399-424. doi:10.1080/00273171.2011.568786

Ayres, I. (2002). Outcome Tests of Racial Disparities in Police Practices. *Justice Research and Policy, 4*(1-2), 131-142. doi:10.3818/jrp.4.1.2002.131

Ayres, I. (2010). Testing for discrimination and the problem of 'included variable bias'. *Yale Law Schol Working Paper [ONLINE: http://islandia.law.yale.edu/ayres/ayresincludedvariablebias.pdf]*. Retrieved from

Baum, C. F. (2006). *An Introduction to Modern Econometrics Using STATA*. College Station, TX: Stata Press.

Berk, R. A. (1983). An Introduction to sample selection bias in sociological data. *American Sociological Review, 48*, 386-398.

Blalock, H. M., Jr. (1979). *Social statistics* (Revised second edition ed.). New York: McGraw Hill.

Bryan, M. L., & Jenkins, S. P. (2016). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review, 32*(1), 3-22. doi:10.1093/esr/jcv059

Bushway, S., Johnson, B. D., & Slocum, L. A. (2007). Is the Magic Still There? The Use of the Heckman Two-Step Correction for Selection Bias in Criminology. *Journal of Quantitative Criminology, 23*(2), 151-178. doi:10.1007/s10940-007-9024-4

Bushway, S., & Reuter, P. (2008). Economists' contribution to the study of crime and the criminal justice system *Crime and Justice: A Review of Research, Vol 37* (Vol. 37, pp. 389-451).

Chainey, S., & Ratcliffe, J. H. (2005). *GIS and Crime Mapping*. London: John Wiley and Sons

Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw Hill.

Fagan, J. (2012). *Report of Jeffrey Fagan, Ph.D. United States District Court, Southern District of New York, David Floye et al. v. City of New York et al., 08 Civ. 01034 (SAS)*. Retrieved from

Golledge, R. G., & Stimson, R. J. (1997). *Spatial Behavior: A Geographic Perspective*. New York: Guilford Press.

Gordon, R. A. (1968). Issues in multiple regression. *American Journal of Sociology, 73*, 592-616.

Guo, S., & Fraser, M. W. (2015). *Propensity Score Analysis: Statistical Methods and Applications* (Second ed.). Thousand Oaks: Sage.

Hannon, L. An Exploratory Multilevel Analysis of Pedestrian Frisks in Philadelphia. *Race and Justice, 0*(0), 2153368717730106. doi:10.1177/2153368717730106

Hox, J., J. (2010). *Multilevel analysis (2nd Edition).* East Sussex: Routledge.

Hunter, A. (1971). The ecology of Chicago: Persistence and change, 1930 - 1960. *American Journal of Sociology, 77*, 425-443.

Kane, R. J. (2002). The social ecology of police misconduct. *Criminology, 40*(4), 867-896.

Klinger, D. A. (1997). Negotiating order in patrol work: An Ecological theory of police response to deviance. *Criminology, 35*(2), 277-306.

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Long, J. S., & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata* (Second ed.). College Station, Texas: Stata Press.

Mantel, N., & Haenszel, W. (1959). Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease. *Journal of the National Cancer Institute, 22*(4), 719-748. doi:10.1093/jnci/22.4.719

Peterson, R. D., & Krivo, L. J. (2010). *Divergent Social Worlds: Neighborhood Crime and the Racial-Spatial Divide*. New York: Russell Sage

Rabe-Hesketh, S., & Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata Volume II: Categorical responses, counts and survival. Third edition (ISBN 9781597181044)*. College Station, TX.: Stata Press.

Rudovsky, D., Messing, P., & Lin, S. (2017). *Plaintiffs' seventh report to court and monitor on stop and frisk practices: Fourteenth Amendment racial disparity issues. Mahari Bailey et. al, v. City of Philadelphia, et al., in the United States District Court for the Eastern District of Pennsylcania, C.A. No.10-5952*. Retrieved from

Schmidt-Catran, A. W., & Fairbrother, M. (2016). The Random Effects in Multilevel Models: Getting Them Wrong and Getting Them Right. *European Sociological Review, 32*(1), 23-38. doi:10.1093/esr/jcv090

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (2nd Edition)*. Los Angeles, CA: Sage.

Taniguchi, T. (2010). *Policing a negotiated world: An Empirical assessment of the ecological theory of policing. Unpublished doctoral dissertation.* (Ph.D.), Temple University, Philadelphia, PA.

Taylor, R. B. (2015). *Community Criminology: Fundamentals of Spatial and Temporal Scaling, Ecological Indicators, and Selectivity Bias*. New York: New York University Press.

Taylor, R. B. (2017). *Conceptual framing and racial disparities in frisk and search hit rates*. Paper presented at the Annual meetings of the American Society of Criminology, November, Philadelphia, PA.

Taylor, R. B., & Johnson, L. (2017). *RE: ACLU Matter vs. - REF# 1340012232: Analysis of Chicago Police Department Post-stop Outcomes during Investigatory Stops January through June 2016: Input to Hon. Arlander Keys' (Ret.) First Year Report. Unpublished Report. Available online at: https://www.aclu-il.org/en/publications/march-2017-stop-frisk-report ; listed as Appendix B: Analysis of CPD Post-stop Outcomes during Investigatory Stops. Also available online at: http://www.rbtaylor.net/pubs.htm* Retrieved from

Taylor, R. B., Ratcliffe, J. H., & Perenzin, A. (2015). Can we predict long term community crime problems?: The estimation of ecological continuity to model risk heterogeneity. *Journal of Research in Crime and Delinquency, 52*(5), 635-657. doi:DOI: 10.1177/0022427815586062

The United States Census Bureau. (2014). American Community Survey (ACS): Design and methodology report. [ONLINE: https://www.census.gov/programs-surveys/acs/methodology/design-and-methodology.html ; accessed November 11, 2017]. Retrieved from

Tyler, T., Fagan, J., & Geller, A. (2014). Street stops and police legitimacy: Teachable moments in young urgban men's legal socialization. *Journal of Empirical Legal Studies, 11*(4), 751-785. doi:doi:10.1111/jels.12055

White, M. D., & Fradella, H. F. (2016). *Stop and Frisk: The Use and Abuse of a Controversial Policing Tactic*. New York, NY: New York University Press.