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

REVISED FINAL TECHNICAL REPORT

Ralph B. Taylor & Lallen T. Johnson

DATE: 20170321

Acknowledgments. The authors appreciate helpful input from Sharad Goel, Aziz Huq, Jens Ludwig and Justin McCrary on the design of the present analyses. The authors thank Jeff Ward for reviewing a draft version. 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. The authors disclose receipt of the following financial support for the research and authorship of this work: Authors were paid by the City of Chicago as part of the above referenced agreement to provide statistical input to the Hon. Arlander Keys (Ret.).

1 CONTENTS

2 INTRODUCTION TO REVISED VERSION 6
3 FOR THE NON-TECHNICAL READER: FAQ 7
3.1 Outcomes of interest and data sources 7
3.2 How to think about ethnoracial differences 7
3.2.1 Least restrictive: Gross impact of race or ethnicity 7
3.2.2 More restrictive: Net impact of race or ethnicity 8
3.2.3 Even more restrictive: Statistically significant net impact of race or ethnicity 8
3.2.4 Most restrictive: Is the statistically significant net impact causal or just correlational? 8
3.2.5 Clarifying causal 9
3.2.6 Gross impacts versus net impacts and the importance question 9
3.2.7 Gross and net impacts and disparate impact and treatment 10
3.3 Describing ethnoracial differences: Locating gross impacts of race and ethnicity 10
3.4 Selecting the other factors 12
3.5 Describing geographic differences 12
3.5.1 Locating basic geographic differences on outcome variables 12
3.5.2 Geography as an important source of differences on the outcome 13
3.5.3 Important "left over" geographic differences even after taking model factors into account 13
3.5.4 Source of significant geographic discrepancies not currently clear 14
3.6 Takeaway lessons 14
3.6.1 Pat downs 15
3.6.2 Pat downs during a stop in which no enforcement action is delivered 15
3.6.3 Searches and ethnicity 15
3.7 Limitations 15
4 EXECUTIVE SUMMARY 17
5 Scope 19
5.1 Outcomes of interest 19
5.2 Questions addressed 20
5.2.1 Descriptive 20
5.2.2 Involving statistical inference 20
6 Background: Police post stop outcomes 20
6.1 General 20

6.2 Comments on specific outcomes 26
6.2.1 Hit rate outcomes 26
6.2.2 Frisk or pat down and release 26
6.3 Analytic concerns 27
6.3.1 Internal replication across independent samples 27
6.3.2 Internal replication across alternative analytic approaches 27
6.3.3 Clustered data 28
6.3.4 Statistical power 28
6.3.5 Multiple correlated outcomes 28
7 Methodology 28
7.1 Data sources 28
7.2 Terms 28
7.3 Data processing 29
7.4 Sampling 32
7.5 Units of analysis 32
7.6 Clustering 33
7.7 Geographies and implications for analyses 33
7.8 Outcome variables 33
7.8.1 Overall descriptive statistics 33
7.8.2 Pat downs: Across groups and districts 35
7.8.3 If a pat down is conducted, are any weapons/firearms recovered? 38
7.8.4 Is a search conducted or not? 40
7.8.5 If a search is conducted, are any weapons recovered? 43
7.8.6 Quick aside: Search hits on weapons or contraband 46
7.8.7 Is any enforcement action delivered or not? 47
7.8.8 Pat down but no enforcement action 49
7.9 Independent variables 56
7.10 Analytic sequence: Rationale and details 58
7.10.1 Outcomes where there is no necessary selection process 58
7.10.2 Outcomes where there is sequential selection 59
8 A priori statistical power calculations 59
9 Background on analytic choices 61
9.1 diagnostics and rationale 61
9.1.1 Regression Diagnostics 61

9.1.2 Propensity models: Assessing selection on observables ... 62
9.1.3 Propensity models: Assessing selection on unobservables ... 62
9.2 Multicollinearity in regression models ... 62
9.3 Clustered data ... 63
9.4 Geography ... 63
10 Results ... 63
10.1 Did a pat down occur? ... 63
10.1.1 Regression ... 63
10.1.2 Caliper matched propensity score models: Non-Hispanic Black vs. White civilians 73
10.1.3 Caliper matched propensity score models: Hispanic vs. White non-Hispanic civilians 77
10.2 Did the pat down result in a weapon/firearm being discovered? ... 80
10.2.1 Multiple logistic regression models with predicted probabilities of a pat down ... 82
10.2.2 Heckman probit selection models ... 87
10.2.3 Conclusions on weapon recovery from pat downs ... 93
10.3 Was a search conducted? ... 95
10.3.1 Exclusion question ... 95
10.3.2 Search links to other enforcement outcomes ... 96
10.3.3 Mixed effects regression models ... 96
10.3.4 Propensity score models: Black vs. White non-Hispanics only ... 99
10.3.5 Propensity score models: White non-Hispanic vs. Hispanic only ... 100
10.3.6 Summing up on search outcome and race and ethnicity ... 102
10.4 Did a search result in a weapon being discovered? ... 102
10.5 Did the officer engage in enforcement? ... 102
10.5.1 Regression results ... 102
10.5.2 Diagnostics ... 104
10.5.3 Propensity selection model – Black non-Hispanic vs. White non-Hispanic ... 106
10.5.4 Propensity selection model – Hispanic vs. White non-Hispanic ... 107
10.5.5 Overall conclusion on race/ethnicity and enforcement ... 109
10.6 If no enforcement took place, what determined whether a pat down took place? ... 109
10.6.1 Main modeling approach ... 109
10.6.2 Alternative models ... 111
11 Discussion ... 112

11.1 Limitations and strengths........ 112
11.1.1 Limitations........ 112
11.1.2 Potential strengths........ 113
12 Key findings........ 113
13 ADDENDUM 1........ 117
14 References........ 118

2 INTRODUCTION TO REVISED VERSION

Comments by the Parties and their experts on the initial version of this report led to modifications that appear in this version. The major modifications include the following.

1. Pointing out to readers, in section 3, where they can find how the different ethnoracial groups scored on the outcomes of interest here. Those tables also show how scores on each outcome varied by district, and varied within district by ethnoracial category.
2. Clarifying the four levels of scrutiny be applied to ethnoracial differences on each outcome: gross impact, net impact, statistically significant net impact, and statistically significant net impact that may be causal rather than just correlational.
3. Section 3 also explains the specific social science meaning of the term "cause" and "causal impact" as it is used in this report.
4. Section 3 further highlights the series of questions that each analysis is designed to answer.
5. Clarifying how geographic variation in the multivariate analyses was reported and presented.
6. Clarifying that the main model applied to each outcome here **as requested by the Parties experts and as agreed, are multiple regression models.** They have some improvements over garden-variety ordinary least squares single-level multiple regression models, but they are multiple regression models at heart. Further, the improvements they incorporate are in line with current best social science scholarly practices in this area.
7. Clarifying that stops with searches associated with arrests were dropped only in analyses of the search outcome, not when other outcomes were considered, and addressing the under-excluding/over-excluding question when dropping searches associated with arrests.
8. Discussing the partialling fallacy as a potential limitation when interpreting net impacts of race or ethnicity variables.

3 FOR THE NON-TECHNICAL READER: FAQ

This section asks and answers frequently asked questions the non-technical reader might have about this report. It simultaneously guides the non-technical reader to findings and interpretations that might be of most interest to him or her. Even technical readers might benefit from scanning the questions and answers listed here.

3.1 OUTCOMES OF INTEREST AND DATA SOURCES

Q: What is this report about?
A: This report describes what happens to civilians stopped by Chicago Police Department officers during the first six months of 2016. Of special interest is how what happens may depend on the race or ethnicity, that is, the ethnoracial category, of the stopped civilian.

Q: What kinds of things can happen to a stopped civilian during a police encounter?
A: Many things can happen, but only a few of those are considered here. The post stop outcomes investigated include: whether the civilian is patted down or not; whether the pat down resulted in a weapon or firearm being recovered; whether the civilian was searched or not; and, among those civilian stops that resulted in no enforcement action being delivered, if race or ethnicity link to whether or not a pat down took place.

In some jurisdictions a pat down is also known as a frisk.

Q: What data source does this report rely on?
A: The report analyzes records from the Chicago Police Department investigatory stop reports (ISR) database. The database provides a wealth of information, only some of which is used here.

3.2 HOW TO THINK ABOUT ETHNORACIAL DIFFERENCES

Q: How do you decide if an outcome depends on the race or ethnicity of the stopped civilian?
A: This report frames the question of ethnoracial differences on each outcome in multiple ways. From a social science perspective, those ways range from less restrictive to more restrictive.

Records in the CPD stop database record the race and ethnicity of the stopped civilian. The three numerically largest groups stopped, in terms of their race and ethnicity, were: Black non-Hispanic civilians, Hispanic civilians, and White non-Hispanic civilians. It is the differences between Black vs. White non-Hispanics, and Hispanics vs. White non-Hispanics, on each outcome, that we investigate here.

Q: Can you explain what you mean by less restrictive vs. more restrictive view on race or ethnicity impacts?
A: Yes. See below.

3.2.1 Least restrictive: Gross impact of race or ethnicity

The least restrictive way to think about these ethnoracial differences is to look at group mean differences on each outcome across the three groups. Simple differences in the average score of each group on an outcome describe a **gross impact** of ethnoracial category on the outcome. It is called a gross impact because no other factors are taken into account. That is, the difference on the outcome across the ethnoracial groups has nothing removed from it. A gross impact is usually just described rather than tested using statistical analysis. Nevertheless, gross impacts may prove important in the discussion of these results.

Although gross impacts can prove important for many purposes, from a social science perspective we want to know more. We recognize that many other factors link to race, and/or the outcome in question. So we seek an estimate that takes those other factors into account, and then re-examines the connection between race or ethnicity and the outcome after doing that.

3.2.2 More restrictive: Net impact of race or ethnicity

Statistical analyses remove the impacts of these other factors from both the race or ethnicity variable, and the outcome in question. After this removal what remains is a **net impact** of ethnoracial category on the outcome. It is called a net impact because it is the amount of connection that remains between ethnoracial categories and the outcome *after* removing the influences of these other factors.

If the analyses work as they are supposed to, the net impact is made up of the connection between two quantities: the portions of the race and ethnicity variables that are unrelated to any of the other factors that have been considered; and the portion of the outcome variable that is unrelated to any of those other factors as well.

Oftentimes, but not always, net impacts of ethnicity or race will be smaller in size – that is, reflect less of an impact – than the gross impacts. There is a tradeoff. The (e.g.) race impact might be smaller in size after taking other factors into consideration. But, depending on the circumstances and one's perspective, and what statistical model diagnoses reveal, one may be more assured that the link is telling you about more about (e.g.) race per se.

3.2.3 Even more restrictive: Statistically significant net impact of race or ethnicity

A third and even more restrictive way to think about these ethnoracial differences is to gauge whether a net impact of ethnoracial differences on an outcome represents something more than just noise in the data or a chance connection.

We rely on the statistical probability associated with a net impact to decide if it is indeed more than just noise or chance. If, given certain assumptions, the statistical probability associated with the net impact in question is very low -- usually this means we would see a result like this due just to chance alone fewer than five times in 100 -- we are more confident that the net impact in question is meaningful in a statistical sense.

Putting aside the specific type of statistical analyses done, in social science investigations of potential disparities in policing, this guidepost – is the net ethnicity or race difference on the outcome statistically significant? – is what is routinely relied upon by those using such studies as part of their inquiry into potential disparities.

3.2.4 Most restrictive: Is the statistically significant net impact causal or just correlational?

The fourth and most restrictive way to think about these ethnoracial differences is to test the statistical models we have done, to "look under the hood" if you will, and conduct additional statistical models. The hope is to learn whether the net impact examined should be interpreted as causal or correlational.

You may have heard the phrase “correlation does not necessarily imply causation” or more simply “correlation is not causation.”

Even though a statistical model might tell us that a predictor like an ethnoracial difference has a statistically significant net impact on an outcome like whether the stopped civilian is patted down or not, we're not sure that it is the race or ethnicity difference *per se* that is responsible for that net impact. Even though we have tried to control for other factors that we have data on in the database, those other factors could still be playing a role. Further, there might be factors outside the variables used in the statistical models that could be playing a role.

So we put the statistical models we have run under the microscope, and conduct additional statistical models, to try to learn whether other factors in our models, or other factors outside our models, could still be playing a role in generating the statistically significant net impact of ethnoracial category that we have observed.

In almost every instance these additional diagnostics suggest the net connections observed here are not assuredly causal in nature, suggesting a correlational interpretation of the link may be the more prudent interpretation. It is not at all unusual when analyzing data sources that do *not* come from a *true* scientific experiment to have doubts about whether the impacts seen are causal.

3.2.5 Clarifying causal

As social scientists, when we say that an impact could be causal, we are saying that the impact appears to be related to the predictor alone, and is not influenced by other factors inside or outside the model, or by selection dynamics. In social science it is extremely difficult to prove a causal claim unless a very particular type of study is done: a randomized controlled trial. These are often done in in medicine and public health as well as many other areas.

But the data here are from ongoing operations of the Chicago Police Department, not a randomized controlled trial. Police are more or less likely to encounter civilians of particular races and ethnicities at certain times in certain locations with certain surrounding circumstances based on a whole range of factors. Separating race and ethnicity differences from those other factors in a situation like this is extremely challenging.

This challenge is *not* something specific to the outcomes being investigated here or the database used or the location. This is a *general* challenge that crops up in *almost all non-experimental data.*

3.2.6 Gross impacts versus net impacts and the importance question

Q: Because estimates of net impacts attempt to remove influences of other factors, does that mean that these estimates of net impacts of ethnoracial category are more important than the estimates of gross impacts of ethnoracial category?
A: It depends on your point of view.

One could argue, depending upon the policy or practice in question and one's viewpoint, that the simple difference between the experiences of the civilians in the three different ethnoracial groups considered is of primary importance.

Alternatively, one could argue from a social science perspective that the net impacts are more important because they may be more likely to inform the reader about the impacts associated with the key variable in question. The social science goal is to test for the impact of individual factors, like being Black non-Hispanic vs. White non-Hispanic, and attach to that difference the impact that seems to be associated with *just* that difference. That goal is not always met as we see here from additional diagnostics of models, but that is the goal.

Q: But when you control for these other factors, don't you run the risk of *underestimating* the remaining racial and ethnic differences?
A: You could run that risk –social scientists call it the partialling fallacy – depending on how you set up some details in your analyses, what those other factors are in your model, and what your *theory* says about the links between racial or ethnic differences and those other factors.

For a number of technical reasons – multicollinearity assessments, selecting other factors based on comparable other studies, how geography is handled in the models, and the underlying theoretical frames – we would argue that these analyses do not commit the partialling fallacy. But that can be a point for vigorous debate and we recognize that others may disagree with us on this point. Scholars argue as well about the partialling fallacy among themselves.

3.2.7 Gross and net impacts and disparate impact and treatment

Q: I do not see anything in your report about legal standards like disparate impact and disparate treatment. Why not?
A: For two reasons. First, the authors are social scientists, not legal scholars. From a social science perspective, the purpose of the analysis is to gauge gross impacts of race or ethnicity differences, or net impacts of race or ethnicity differences, on stop activity, where net impacts are defined in progressively stricter ways. Second, for the outcomes in question here, the authors are not aware of a widely accepted mapping of gross or net impacts as defined in a social science framework onto disparate impact or disparate treatment standards.

We have described four ways, with varying levels of restrictiveness, for gauging impacts of racial or ethnic differences on outcomes of interest. Should all four of those ways be interpreted as relevant to disparate impact? Should some of those ways be interpreted as relevant to disparate treatment? We don't know, but think that the cross-referencing question, and how the cross-referencing might depend on broader features of context, merits conversation between the legal scholars and social scientists.

3.3 Describing ethnoracial differences: Locating gross impacts of race and ethnicity

Q: I am interested in seeing the average scores of the civilians in each of the three different ethnoracial groups for each post stop outcome you examine. Where do I find that information?
A:

Table 1 below tells you where to find the average score of each of the three ethnoracial groups on each outcome. These differences describe the gross impact of ethnoracial grouping on the outcome in question.

Table 1 Where to find gross impacts of ethnoracial category for each outcome

Post-stop outcome	Ethnoracial group differences found in:
Pat down conducted	Table 10
Weapons/firearms discovered through pat down	Table 12
Search conducted	Table 14, Table 15
Weapons/firearms discovered through search	Table 16, Table 17, Table 18
Weapons/firearms or contraband discovered through search	Table 19
Stopped civilian receives any enforcement action	Table 21
Stopped civilian receives pat down but no enforcement action	Table 22, Table 24

Note. Average scores for each ethnoracial group usually found in the last row of the listed table. All the outcomes listed in this table are scored 0 if the outcome did not happen and 1 if it did. Thus the average score for each group represents the proportion of that group that did experience that outcome.

3.4 Selecting the other factors

Q: How do you decide what the other factors are that you're going to take into account?
A: We looked carefully at other studies where researchers have investigated questions like the ones being considered here. That, along with a general concern about taking into consideration outcome variation that can be due to time of day, time of the week, or season; and outcome variation that can be due to geographic differences, led to the final selection of other factors.

Q: So what are the other specific factors to take into account in your models?
A: You will find them in Table 26. The variables that do not have a star are used in almost all the models that took multiple factors into account. In addition to race and ethnicity, other variables included gender, age categories, time of day categories, whether the stop happened on the weekend, and whether it was a vehicle as opposed to a pedestrian stop.

In addition to these specific features of individual stops, geographic variation also was taken into account. Geography was taken into account by allowing each police district to have its own average score on the outcome variable being considered.

3.5 Describing geographic differences

3.5.1 Locating basic geographic differences on outcome variables

Q: I am interested in seeing how scores on each outcome examined vary across police districts. Where can I find that information?
A: Each of the tables listed in

Table 1 above also show differences on the outcome score by police districts. All these tables have a separate row for each police district. The number that appears at the end of each row indicates the average outcome score in that particular district.

3.5.2 Geography as an important source of differences on the outcome

Q: Is the geographic variation on outcome scores important?

A: Yes. As you can see from the different proportions for each district for each outcome in the tables noted above, each outcome is more likely to happen in some places and less likely to happen in others. All of the statistical models taking multiple factors into account confirm that the geographic variation in each outcome is more than just chance or noise in the data.

3.5.3 Important "left over" geographic differences even after taking model factors into account

But geography matters in a second way as well. For some of the outcomes we examined what was left over, that is, the portion of each outcome that is not predicted by the factors used in the model.

It turns out that for some outcomes that remaining geographic variation suggest "something going on" in some districts. By "something going on" we mean something that is statistically discrepant from the overall picture, and is unrelated to the factors that we used in our models.

For example, take a look at Figure 6. This shows results from analyzing the first random sample. Each district has a filled in circle. If the filled in circle for a district is **below** the **horizontal line** it means that in that district, even after taking into account ethnoracial category and all the other factors used when predicting whether or not a pat down took place, and even after allowing each district to have its own (adjusted) average score on the outcome, the **proportion of stops resulting in a pat down** in that district is **lower than overall**. If the filled in circle for a district is **above** the **horizontal line** it means that in that district, even after taking the same factors into account, the **proportion of stops resulting in a pat down** in that district is **higher than overall**.

Q: Why does each filled in circle have lines coming out of it?

A: Those lines take sampling error into account. After we consider that error, our best guess is that the true mean score for that district on that outcome is somewhere between where the upper line ends and the lower line ends.

Q: Are any of these district differences in Figure 6 meaningful?

A: They may be.

Look at the left-most district mean. This is for District 16. Because the lines coming out of the circle do not cross the horizontal line this means that after taking predictors into account stops in this district are still significantly less likely to result in a pat down compared to the overall average across all the districts.

The line coming out of the fourth circle from the left corresponds to District 2. Here too the proportion of stops resulting in a pat down is significantly lower than the overall average, even after taking all factors into account.

Take a look at the two right most circles with vertical lines coming out of them. These correspond to **Districts 6 and 7**. In these two districts, **even after** taking other factors into account, the proportion of stops here resulting in a pat down in each of these two districts is significantly **higher than the overall average**.

Q: So are you saying there may be something going on in Districts 6 and 7, based on Figure 6, that is unrelated to the factors you used, that is resulting in significantly more stops involving pat downs compared to the overall average across all the districts?
A: We are.

Q: Do Districts 6 and 7 stand out this way when you analyze your second random sample?
A: They do. See Figure 8.

3.5.4 Source of significant geographic discrepancies not currently clear

Q: Do you know what is responsible?
A: We do not. In each case it could be something about the district organization itself, something about the mix of people encountered on the street walking or driving, something about the mix of land uses or public transit in these districts, or some other factor(s). We just don't know for certain.

But we did tentatively explore the connection between these district deviations from expected patterns. See Figure 7 and Figure 9.

In each figure, the vertical axis shows the district mean deviation, after taking model factors into account, on proportion stopped civilians patted down. On the horizontal axis is the percent of stopped civilians who were Black and non-Hispanic.

The curvy shows the locally-weighted relationship between these two factors. For both the first and second random samples, it looks like districts where probabilities of a pat down taking place are higher, even after taking model factors into account, are also districts where the proportion of those stopped was more predominantly Black and non-Hispanic.

But this is just an **exploratory descriptive** examination, **with no tests for statistical significance** and is **not definitive.** We just cannot say anything definitive about what this "something going on" is that results in some districts having higher fractions of stops with pat downs than the model expects.

3.6 Takeaway Lessons

Q: What are your most important findings?
A: "Most important" is in the eye of the beholder. From our social science vantage, however, we would focus most attention on those statistically significant net impacts of a Black vs. White difference or Hispanic vs. non-Hispanic White difference that:

- Appear with both random samples using the primary analysis model;
- Appear with both random samples using an alternative analysis model; and where
- There was a low degree of concern about other observed or unobserved factors interfering with the race or ethnic impact observed.

Table 55 organizes the findings using these considerations. Given these considerations, in our view the ***strongest*** findings were as follows. ***This does not mean we think any of the other***

findings are necessarily unimportant. It is just that these highlighted findings seem the most durable, at least from an analytic perspective.

3.6.1 Pat downs

A: Significant net differences between White non-Hispanic and Black non-Hispanic stopped civilians appeared on the pat down outcome in both random samples using both the main statistical analysis and the alternative statistical analysis.

The **gross** difference between these two groups in both samples was about 11% or 12%; about 34% to 35% of stopped Black non-Hispanic civilians got patted down as compared to about 23% to 24% of stopped White non-Hispanic civilians. The size of the **net** impact can be expressed in the how odds of this happening versus that happening – the odds of [a pat down happening versus not happening] were higher or lower depending on the group in question. These odds, which reflect net impacts, were anywhere from 19% to 32% higher for Black non-Hispanic compared to White non-Hispanic stopped civilians, depending on the sample and the model. This race differential was always **statistically significant** meaning it was not due to chance or noise in the data. Given model diagnostics this link is probably best interpreted as correlational rather than causal.

3.6.2 Pat downs during a stop in which no enforcement action is delivered

A: This outcome is also about pat downs, but only in situations where officers deliver no enforcement action. Procedural justice scholars suggest that getting patted down is intrusive, and if it happens in a stop where no other actions are taken against the civilian he or she may perceive such actions as unwarranted.

In stops where police officers delivered no enforcement action, *both* significant net race and net ethnicity impacts appeared. Black non-Hispanic stopped civilians got patted down significantly more often than White non-Hispanic stopped civilians in these situations, as did Hispanic stopped civilians. The gross difference between White non-Hispanics and the other two groups in these stops was about 12% to 15%; see Table 25. This result is highlighted here because the significant net impact replicated across two random samples and across alternate analytics.

3.6.3 Searches and ethnicity

A: The most stringent analyses conducted found that Hispanic stopped citizens, as compared to non-Hispanic stopped citizens, were more likely to be searched.

3.7 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 taken into account a different set of factors than the ones we used. In addition, there were things we wanted to do either in terms of different types of analytics, or additional diagnostics of the models we used, that we have not yet had time to complete.

4 EXECUTIVE SUMMARY

This report analyzes investigative stop report data from the Chicago Police Department for the period 1/1/2016-6/30/2016. Five different outcomes from these stops are analyzed. Simple differences on *each* of these outcomes across the three ethnoracial groups of interest – stopped Black non-Hispanic civilians, stopped White non-Hispanic civilians, and stopped Hispanic civilians – are also displayed in a series of tables (see Table 1in the FAQ section above). These allow the reader to examine differences by race and ethnicity as well as differences by location, and race and ethnicity differences within locations. These simple differences, which we also call **gross impacts** of different ethnoracial categories, are of interest in their own right.

Beyond these descriptive differences, of key interest is whether, after taking into account other factors, there are differences on a post stop outcome associated with civilian race, civilian ethnicity, or civilian gender. A **net impact** of an ethnoracial category difference refers to these associations observed after controlling for other factors.

Of even more interest are net impacts of ethnoracial category that prove statistically significant. This means the net link observed is likely not due to chance or noise in the data.

Further, if a statistically significant net impact is found, statistical models are diagnosed to learn whether that connection is best interpreted as causal or correlational. If the interpretation is correlational only, that is because other factors, or selection dynamics, may play roles in "driving" the net connection between the ethnicity or race variable and the outcome.

Finally, we conduct alternative statistical models to learn if statistically significant net impacts of race or ethnicity can be repeated using models that make different assumptions.

This executive summary focuses on key findings for race and ethnicity and suggested interpretations. Much of this section is repeated in the final Key Findings section at the end. One also can find there a summary table (Table 55).

Pat downs. The strongest pattern revealed by these analyses are net connections between race and whether a pat down occurred, and between ethnicity and this outcome. Both analytic approaches yielded statistically significant net connections in both samples.

Diagnostics of both types of pat down models, however, suggested a moderate level of potential concern about observed and unobserved selection biases. Stated differently, there were other things going on that were not handled sufficiently by the analytics. Given that, the net race and ethnicity impacts are probably best interpreted as correlational. Nonetheless, the links were there, after controlling for other factors, and for district context. As compared to White non-Hispanic civilians, Black non-Hispanic and Hispanic civilians were more likely subjected to a pat down.

Pat downs leading to weapons. Previous work on pat down and search hit rates suggested that pat downs of Black and Hispanic civilians would be less likely to lead to recovered weapons. This turned out to be true when examining weapons produced from pat downs, after controlling for other factors and district context. It held for Black as compared to White civilians. Hit rates were significantly lower in both random samples in the regression analyses for Black as compared to White non-Hispanic civilians.

The significant net race effect did not resurface, however, using more stringent analytics. Again, diagnostics suggested some concerns. The conclusion seems to be that there is a net race effect, but it is probably correlational and was just not quite strong enough to be robust across alternate analytics.

Searches. The search outcome results showed no significant net race effects. But significant net ethnicity links appeared, for both samples, using the more stringent alternative analytics. Hispanic civilians were more likely to be searched than non-Hispanic White civilians after controlling for other factors. Diagnostics suggested some level of concern, so the conclusion about ethnicity and the search outcome is that the link is probably correlational, but not robust across different approaches.

Reviewers have raised some worthwhile points with regard to the search outcome. In essence they argue that removing stops where searches and arrests happened may have inappropriately dropped a large number of stops, and were those included a different picture of net race impacts might appear. Future work will address this concern.

Any enforcement action delivered. The enforcement outcome yielded robust net ethnicity links across both samples and both analytic approaches.

Stopped Hispanic civilians, as compared to stopped Black non-Hispanic civilians, were significantly more likely to be subjected to some type of enforcement action during the stop. The gross impact of ethnicity was as follows (Table 21): whereas 28 percent of stopped White non-Hispanic civilians received some type of enforcement action, 31 percent of stopped Hispanic civilians received such an action.

Net race links surfaced only with one analytic approach. The conclusion seems to be, in light of diagnostics, that for both race and ethnicity there is a net connection with this outcome, that for both it is probably best considered correlational, and that for race it is not robust across alternative analytic approaches.

Pat down and no enforcement. The last outcome examined contrasted two types of stops, no enforcement action and no pat down vs. no enforcement action and receiving a pat down. Analyses included both a main and an alternate approach. No diagnostics of either analytic model have yet been completed.

Across both analytic approaches, significant net race and ethnicity effects surfaced. After controlling for other factors and district context, in stops where no enforcement actions were taken by police, Black non-Hispanic and Hispanic stopped civilians had much higher odds of being patted down than did stopped White non-Hispanic civilians. Given the potentially corrosive nature of police interactions such as this, this would seem to be an important pattern to address.

These net race and ethnicity links should be considered correlational only at this time, since no diagnostics have been completed.

The gross impact was as follows. In 38 percent of the Black non-Hispanic stops with no enforcement, and in 36 percent of the Hispanic stops with no enforcement, a pat down was delivered. The corresponding percent for stopped White non-Hispanics in stops with no enforcement was 24 percent.

5 SCOPE

This report analyzes investigatory stop report (ISR) records generated by the Chicago Police Department (CPD) during the period January 1, 2016- June 30, 2016.

Analyses consider multiple post-stop outcomes.

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.

The connections are considered in a number of different ways.

First, the connections are considered on their own, without taking other factors into account. These represent gross impacts of race or ethnicity differences on the outcome.

The connections are also considered using progressively stricter criteria.

So the second examination asks: Does the difference persist after controlling for other factors? We refer to these as net impacts.

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

And finally, after reviewing model diagnostics, and perhaps conducting alternative analytics, the fourth examination asks: is a statistically significant net impact more appropriately interpreted as causal or correlational?

5.1 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 post-stop outcomes. Questions of who is stopped where is 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. Is any enforcement action delivered or not?
F. 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?

5.2 Questions addressed

5.2.1 Descriptive

To provide descriptive context, simple race and ethnicity differences, and district differences, are portrayed for these outcomes.

Although statistical tests are often not applied to these differences, these descriptive differences between ethnoracial categories represent an important part of the examination.

5.2.2 Involving statistical inference

For each outcome, the question is the same:

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

The ethnicity question. Controlling for observed covariates, is there a statistically significant net difference on outcome scores between 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 context.

Potential net gender links with each outcome are of interest as well.

6 Background: Police post stop outcomes

6.1 General

At the time of the current study, researchers have been investigating questions of racially or ethnically biased policing, for civilians on foot and in cars stopped by police, for well over two decades (Banks, 2003; Beckett, Nyrop, & Pfingst, 2006; Brunson & Miller, 2006; Engel, 2008; Engel & Calnon, 2004; Engel, Calnon, & Bernard, 2002; Engel & Tillyer, 2008; Fagan, 2002; Fagan & Braga, 2015; Fagan, Geller, Davies, & West, 2009; Fridell, 2005; Gelman, Fagan, & Kiss, 2007; Grogger & Ridgeway, 2006; Harris, 1997; Jernigan, 2000; Lundman & Kaufman, 2003; MacDonald, Stokes, Ridgeway, & Riley, 2007; Meares, 2014; Ridgeway, 2006, 2007a, 2007b, 2009; Ridgeway & MacDonald, 2010; Ridgeway & MacDonald, 2009; Ridgeway & Riley, 2007; Rojek, Rosenfeld, & Decker, 2012; R. Tillyer, Engel, & Cherkauskas, 2010; Rob Tillyer, Klahm, & Engel, 2012; Tyler, Fagan, & Geller, 2014; Walker, 2001).

At the broadest level, for social scientists investigating potential racial or ethnic disparities for a particular post-stop outcome, there are two broad challenges for the analyst: separation and selection. These are described below.

Separation refers to separating out three different sources that could be contributing to a racial or ethnic differences – or any other group based difference -- in police recorded behaviors (Ridgeway, 2009; Walker, 2001). (1) The race or ethnicity linked police differential could arise from differences across groups or across locations in the amounts of recorded or reported criminal/disorderly behaviors drawing attention from officers. (2) The different groups might experience differential exposure to patrolling officers. If some neighborhoods are more heavily policed because of crime or calls for service differences, and if there are racial and or ethnic differences in who is found walking or driving in those neighborhoods, the racial or ethnic differential in exposure could lead to differences in police stop or post stop outcome rates. (3) The third possibility is that police are treating members of different groups in disparate ways. Research has underscored the many problems with finding indicators that can reliably be used to estimate sources (1) and (2), and remove that variation (Ridgeway & MacDonald, 2010), so that the size of (3) can be gauged.

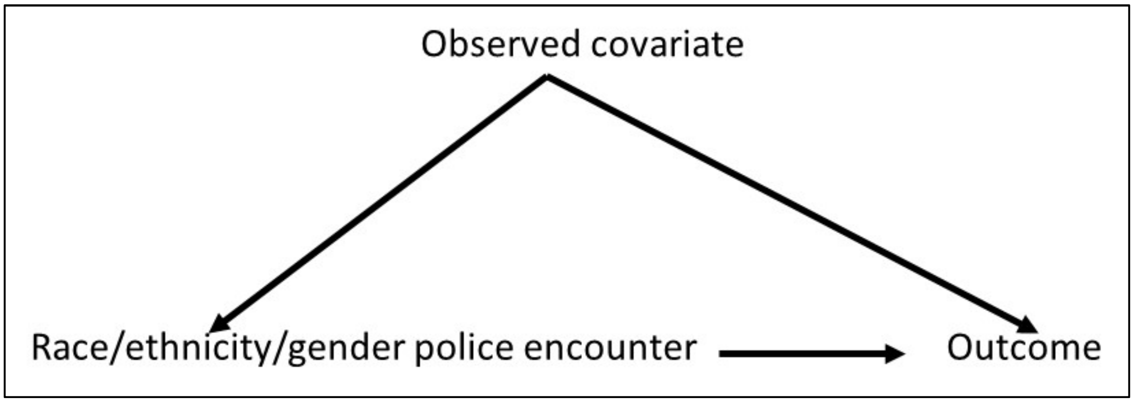
Selection is used here to refer to three distinct but related dynamics. Researchers also seek to gauge the size of these three dynamics. If these three different dynamics can be either estimated or ruled out, then the researcher can make a stronger case that the connection between the race or ethnicity indicator, and the outcome, if such a connection is observed, arises from causal rather than correlational processes. The literature on these matters refers to treatment and control groups. For example, the researcher might be interested in comparing intensive supervision probationers to regular supervision probationers (Petersilia & Turner, 1990).

The data considered here are observational not experimental data. “Observational data generally create challenges in estimating causal effects” (Imbens & Wooldridge, 2009: 7).

Here, selection dynamics refer to differences between stopped Non-Hispanic Black civilians vs. stopped White civilians, or between stopped Hispanic vs. Non-Hispanic White stopped civilians, or between stopped men vs. stopped women, rather than treatment and control groups. This shift in conceptual frame is at some level problematic. The race and ethnicity of civilians encountered by police link to so many aspects of where people live (Peterson & Krivo, 2010). Further, race, ethnicity and gender link to so many features of people’s interactions with others (Delgado & Stefanic, 2012; Reskin, 2012). Consequently, the challenge of disentangling or un-confounding impacts on police recorded behavior of the race or ethnicity or gender of the civilians they encounter, from relevant other attributes, seems Herculean. Nevertheless, attempts to disentangle proceed.

In the situations examined here, the selection problem has three aspects: selection on observables (Figure 1), selection on unobservables (Figure 2), and sequential selection (Figure 3). Each is explained in turn.

Figure 1. The Problem of selection on observed covariates

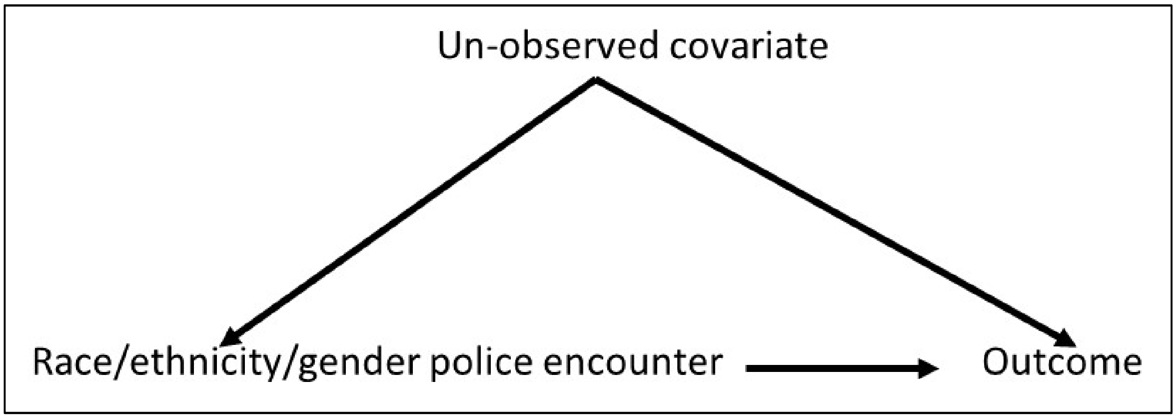


Selection on observables. This type of selection, also known as "unconfoundedness, exogeneity, [and] ignorability," represents an assumption that must be satisfied if one is to support causal interpretations for the impact of a treatment, or, here, the race, or ethnicity or gender variable, on an outcome (Imbens & Wooldridge, 2009: 7). "All these labels refer to some form of the assumption that adjusting treatment and control groups for differences in observed covariates, or pretreatment variables, remove all biases in comparisons between treated and control units … Without unconfoundedness, there is no general approach to estimating treatment effects" (Imbens & Wooldridge, 2009: 7). Therefore, patterns arising from different diagnostics associated with different models merit scrutiny to learn whether the selection on observables can be ruled out. If it can, a causal interpretation of racial or ethnic or gender impacts receives more support. If it cannot, a correlational interpretation receives more support.

For example, in the current work, age of stopped civilians is known, so this is an observed covariate. So its influence can be controlled, and patterning between this covariate and outcome residuals can be examined to see if connections remain.

Selection on unobserved covariates. With observational data, key differences between two groups of interest could be present but not detected. That is there could be "differences due to unobserved covariates" and this "should be addressed … using models for sensitivity analyses" (Rubin, 2001: 173). "Unobserved covariates" refers to factors outside of those used in the model.

Figure 2 The Problem of selection on un-observed covariates



So the challenge here is estimating the potential impacts of *factors not included in the models analyzed* that might be linked to either race or ethnicity.

Different diagnostics for different models help gauge whether selection on unobserved covariates is a sizable concern. If it is a sizable concern, then a causal interpretation of an observed race or ethnicity or gender impact is unwarranted; rather, the interpretation should remain correlational.

Sequential selection. This refers to a well-known problem in economics, sociology, criminology, criminal justice and other social and hard sciences (Babu & Jang, 2006; Berk, 1983; Bushway, Johnson, & Slocum, 2007; Bushway & Reuter, 2008; Fu, Winship, & Mare, 2004; Heckman, 1979). The problem surfaces if the data for an outcome variable gets collected only if something else happens prior to that. In criminal justice, for example, whether or not a defendant found guilty is sentenced to one or more years of prison, or to a less severe sentence, depends on the defendant being found guilty in the first place. A researcher studying the determinants of sentence severity would want to take into account and control for the determinants of the prior outcome, obtaining a guilty verdict. Should the researcher fail to model those prior selection dynamics, answers she obtains to her main question of interest, the determinants of more vs. less severe sentences, could be misleading.

Sequential selection surfaces as a concern in research on race or ethnicity or gender and police activities such as driver stops, pedestrian stops, frisks of stopped civilians, or searches of stopped civilians. Ideally, and this has been done in some of the driving while Black research (Grogger & Ridgeway, 2006), one wants to control for or at least neutralize the factors associated with one group being more likely to be stopped in the first place.

Of most relevance here, and separate from being selected for a stop in the first place, are other sequential selection concerns if a post stop outcome depends upon the prior occurrence of an earlier post stop outcome. Whether a pat down results in weapons being located requires that a pat down occur in the first place. Whether a search results in weapons being located requires that a search occur in the first place.

The sequential selection analytic concern aligns with a broader theoretical assumption about policing behavior prior to and during civilian stops: that officers are making a series of decisions prior to and during a stop. For example, Fallik and Novak (2012: 148) discuss three decision points in the case of automobile stops:

> Racial profiling within automobile stops has focused on three distinct officer-initiated decision-making points that can measure the presence of racial and/or ethnic non neutrality … The first is the officer's decision to initiate a stop. This decision-making point typically considers the propensity of racial and ethnic minorities to be stopped or whether Blacks or Hispanics are stopped at a higher rate than their community representation … The second officer-initiated decision making point considers an officer's application of formal sanctions or the exercise of coercion. Research from this decision-making point considers the propensity of racial and ethnic minorities to be warned, cited, arrested, and have force used against them … The third decision-making point considers minority representation in searches … An extension of this line of inquiry involves analyzing the contraband hit rate or outcome test during searches[.]

Backing up this idea of sequential decision-making points during stops are numerous studies of how police respond to ongoing civilian actions and civilian demeanor during stops (Mastrofski, Reisig, & McCluskey, 2002; Reisig, McCluskey, Mastrofski, & Terrill, 2004; Terrill & Mastrofski, 2002).

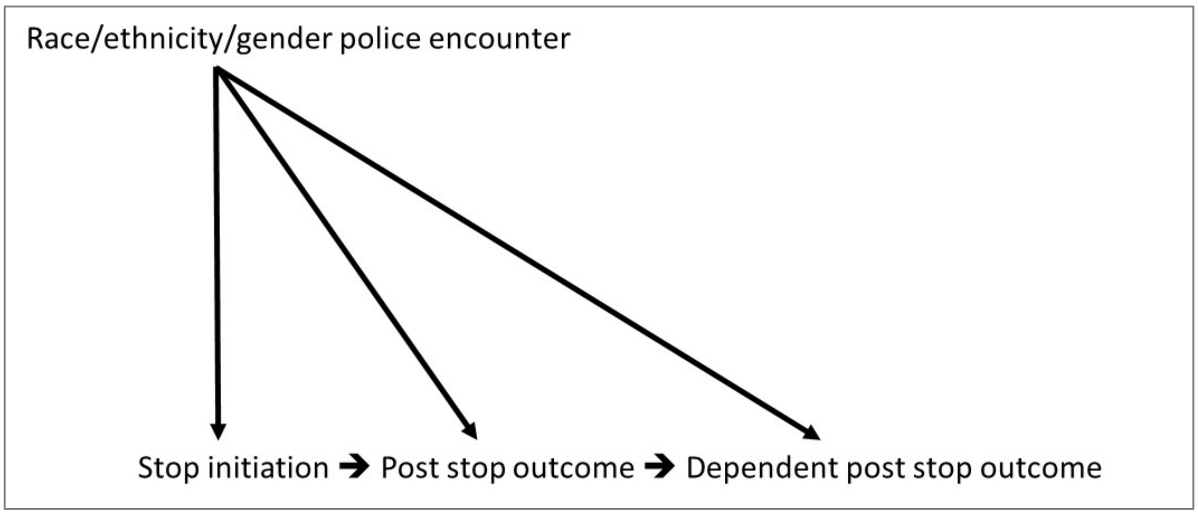
No specific view is promoted at this point about what the specific sets of distinct decision-making or action-selection dynamics are, or about what the temporal relationship might be between the different sets of dynamics, or how one set of dynamics might condition the other. Those are important theoretical questions to be addressed by others. But the general idea of officer sequential decision making prior to and during a stop does lead to the following points that are relevant here.

(1) There are at least two distinct but certainly related sets of dynamics: those leading to the stop, and those involving whether certain actions are initiated after the stop. The degree of relatedness or overlap across these dynamics is not known. The degree to which **each** of these dynamics is racially or ethnically linked represent important and distinct questions.
(2) The fact that information, associated with officers selecting or not selecting a civilian for a stop, is not available here, places an important limitation on this report. It means that **all the outcomes examined in this report fail to report for the first stage of sequential selection.** The extent to which this omission is problematic cannot be estimated with the information available. That said, this omission similarly plagues a large number of other studies of police/civilian post stop outcomes.
(3) Once a stop is underway, if the outcome being examined depends on an action the officer took subsequent to initiating the stop but before this outcome is known, then there is a third set of dynamics which may be racially/ethnically/gender linked. Consider pat down weapon hit rates for example. Race or ethnicity may link to this outcome. Those racial or ethnic links to that outcome could be different than the race or ethnicity links to whether the civilian gets selected for a pat down in the first place.
(4) In each post stop dynamic, race or ethnicity or gender may play distinguishable roles. For example, race or ethnicity could be involved only in pat down selection; race or ethnicity

could be involved only in determining pat down outcomes; or, race or ethnicity could be involved in both pat down selection and in pat down outcomes.

(5) As far as these authors understand, researchers in this field have yet to make a clear case about why race or ethnicity should be involved in some of these post stop dynamics and not others. Therefore, analyses of outcomes that depend on the officer doing something else first, while the stop is underway, need to gauge all possible ways race or ethnicity could contribute to each of these dynamics. That is, for these outcomes sequential selection needs to be modeled. Studies investigating post stop outcomes that *fail* to explicitly also model the selection dynamic (see for example Carroll (2014)) may be generating misleading results.

Figure 3 The Problem of sequential selection



So the race or ethnicity or gender of the person encountered, could have contributed in three separate ways to the chain of events leading to a weapon being produced or not produced from a pat down. Any one of these, or all of these, could have affected the chances that the civilian would be stopped. Any or all of them could have affected the chances that the stopped civilian would be patted down. And, finally, any or all of them could have affected the chances that the pat down would produce a weapon. [1]

In short there are three different ways race or ethnicity or gender could affect officers' discretionary decision making. There are three processes, in sequence, that lead in each case to some person or some action being selected. Consequently, the experts sought to disentangle some of processes. That is, in the case of this outcome, they wanted to estimate the impacts of

[1] The reason for a potential race or ethnicity link to the last outcome arises from what researchers call the subgroup validity problem. In non-technical terms, members of one group may engage more frequently in verbal or nonverbal behaviors that the officers' training suggest are clues to acting suspiciously or having something to hide. But the higher rate of doing those things may just be a group difference, not a clue to something suspicious. So a Black and a White stopped civilian may both be engaging in the same set of behaviors indicating something to hide, but the Black civilian may in fact actually be less likely to be hiding something.

race or ethnicity or gender on the pat down outcome – was a weapon found? -- *separate from* the impacts of each of these on the pat down occurring in the first place. They had no way of separating out these sequential selection dynamics from the selection factors associated with stop initiation in the first place. [2]

These selection dynamics reflect officers' discretion. Figuring out when highly discretionary decision making shades into racially- or ethnically- or gender-*biased* decision making is a tough call. Research on criminal justice decision making does suggest that more highly discretionary decision points have greater chances of being influenced by decision makers' biases (Gottfredson & Gottfredson, 1988). But for now the goal is just to learn about how race or ethnicity or gender link to each of these decision points or outcomes, while taking what happened earlier in the stop into account.

6.2 COMMENTS ON SPECIFIC OUTCOMES

6.2.1 Hit rate outcomes

Search hit rates have drawn particular interest in the driving stop and pedestrian stop literatures. An example hit rate would be: in the case of police stops on a major interstate, what fraction of vehicles searched produced drugs? In the civilian pedestrian stop context, if the purpose of the police stop strategy is to interdict those carrying weapons who are in high crime locations at high crime times, one can ask: is the fraction of searches of Black pedestrians producing a weapon lower than the same fraction for stopped and searched White pedestrians? Economists, making certain assumptions, have provided the conceptual underpinnings for the hit rate analysis (Knowles, Persico, & Todd, 2001; Persico & Todd, 2008).

Other researchers question the assumptions behind this model (Barnes, 2005; Ridgeway & MacDonald, 2010: 22 [online]). The potential subgroup validity problem (Ayres, 2002) seems to be the biggest concern. Simply put, the kinds of verbal and nonverbal factors police are trained to use during a stop to gauge civilian suspiciousness happen at different base rates in different racial/ethnic groups. "The subgroup validity problem remains a concern for the application of the outcome test to police searches … verbal and non-verbal behavioral cues to suspicion and deception are not racially neutral. Thus the accuracy of suspicion cues will likely differ across racial/ethnic groups. Conclusions of racial bias cannot be made using the outcome test" (Engel, 2008: 24). Controversy about the outcome test continues (Engel & Tillyer, 2008; Persico & Todd, 2008).

6.2.2 Frisk or pat down and release

An outcome not previously examined in stop, question and frisk research is introduced here: civilians being patted down and released vs. released without a pat down. Two arguments warrant its examination.

First, situated accounts of police civilian interactions highlight that pat down and release does occur and that it does bother civilians (McArdle & Erzin, 2001; Simon & Burns, 1997). Such

---

[2] This is because only data on stopped individuals were available. There is no information about persons in comparable situations but not stopped by CPD officers.

interactions contribute to tension between inner city Black residents and police (Brunson, 2006, 2007a; Brunson & Gau, 2011; Gau & Brunson, 2010). To be patted down and released may strike many residents of color as simply being hassled by police (McArdle & Erzin, 2001).

Further, this outcome seems particularly relevant given a procedural justice perspective (Sunshine & Tyler, 2003; Tyler, 1988, 1997, 2001, 2003; Tyler & Huo, 2002; Tyler & Lind, 2001). The outcome reflects a component of the construct "degree of police intrusion during… stops" (Tyler et al., 2014: abstract), an outcome recently introduced by procedural justice scholars.

Tyler, Fagan, and Geller (2014: 763) used telephone survey data of young men living in New York City to learn about impacts of their contacts with police on both their views of police legitimacy and their willingness to cooperate with police and courts. In describing "general neighborhood experiences with police" participants reported on the "degree of intrusion during those stops" happening near where they lived. Several survey items contributed to a broader index reflecting intrusion. One of the items in this index was "did the police… 'Frisk or pat you down'"(Tyler et al., 2014: 784).

Would most agree that a stop ending with a pat down and release is more intrusive than a stop and no pat down and release? This certainly seems to be the implication of the work by Tyler, Fagan, and Geller (2014). Those authors observed significant impacts of police intrusiveness on respondents' willingness to cooperate with police (Table 6). This aligns with much of the ethnographic work on urban Black residents and police agrees that unwarranted frisks are intrusive and affects residents' views of police (Brunson, 2005, 2006, 2007b).

That said, no inferences are drawn about the fraction of frisk-and-release stops where police had grounds for a much more intrusive stop such as for example a frisk-and-cite or frisk-and-search stop, or a frisk-search-and-arrest stop. Nor are any inferences made about the fraction of no-frisk-and-release stops where police similarly might have had grounds for more intrusive actions.

6.3 Analytic concerns

6.3.1 Internal replication across independent samples

Two representative random samples of data were available after sampling. Tests of statistical significance were then conducted on both samples. If a key statistically significant finding surfaced 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 particular mix of records found in one sample but not the other.

6.3.2 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 as a result of the consent decree emerging from *Bailey et al. v. City of Philadelphia* use multiple regression models.

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. Second, mixed effects models separate random variation by district on each outcome, and allow for correlated errors within districts. They also make Empirical Bayes adjustments to district-level means.

In one case the outcome is categorical so the model used rather than logistic multiple regression is multinomial multiple regression.

But, in addition to these main multiple regression models, we employed **for every outcome an alternate analytic strategy**. Doing so allows us to learn whether a particular statistically significant net impact of a race or ethnicity difference is robust across different models with may make different assumptions and/or use the data in different ways.

So this allows for a different type of internal replication to see if results are robust across different statistical approaches.

6.3.3 Clustered data

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

6.3.4 Statistical power

A priori power analyses were run (see below) and used to guide selection of the alpha level.

6.3.5 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 was still concerned, he or she could make his/her own internal Bonferroni adjustment by only considering effects that are significant at $p < .01$ rather than $p < .05$.

7 METHODOLOGY

7.1 DATA SOURCES

Chicago Police Department (CPD) personnel made available monthly csv files containing the final version of each Investigatory Stop Report (ISR) available during the period. Each record represented an individual stop report. The files contained data relevant to each field in the ISR form adopted by CPD in January, 2016.

7.2 TERMS

An individual **stop** references one particular stopped civilian whose information was recorded by officers during any type of interaction recorded in the ISR database. These include vehicle stops, pedestrian stops, and gang enforcement. An **event** refers to stops which are grouped together. For over 99 percent of the records here, that grouping of stops was based on unique CPD event numbers [field = **event_no**]. For the remaining less than one percent of the cases stops were grouped if they shared the same date, the same district, the same beat, the same starting hour and minute, and the same first officer star number.

7.3 DATA PROCESSING

CPD sent monthly csv files. Data processing included the following steps. Date and time variables were checked for out of range values and recoded to missing as needed. Numeric variables were created from string variables as needed. Age values below 7 were recoded to 7 and ages above 90 were recoded to missing given the ambiguity in some of the values (was 115 15 or 11?)

Data were de-duplicated so there was only one record with each individual ISR number.

Authors understand from the CPD that in January sometimes different ISRs were generated for the same stop. Those are not removed here. [3]

CPD uses a field for event number (**event_no)** to keep track of different events. This was missing for 424 of 54,701 records (0.78 percent). For these records a proxy event number was generated based on different records taking place in the same district on the same day at the same time and with the same responding first officer. A dummy variable (eventmis) was included in analyses to control for the fact that for some number of records a proxy event number was used.

Using assigned and proxy event numbers permitted gauging the number of stops per event. The distribution appears in Table 2. The number of stops per event ranged from one to 21. Over half of the stops involved three or fewer stops per event.

[3] One way to resolve that matter would have been to randomly sample one ISR number per event number. That was not done, given the importance attached to analyzing all the stops taking place.

Table 2 Number of individual stops per event

N of stops per event (variable = event_n3)	N	Percent	Cumulative Percent
1	10,435	19.08	19.08
2	11,520	21.06	40.14
3	9,936	18.16	58.3
4	7,416	13.56	71.86
5	5,510	10.07	81.93
6	3,408	6.23	88.16
7	2,569	4.7	92.86
8	1,456	2.66	95.52
9	999	1.83	97.35
10	620	1.13	98.48
11	308	0.56	99.04
12	204	0.37	99.42
13	130	0.24	99.65
14	42	0.08	99.73
15	60	0.11	99.84
16	32	0.06	99.9
17	17	0.03	99.93
18	18	0.03	99.96
21	21	0.04	100
-----------	-----------	-----------	-----------
Total	54,701	100	

Source: Jan-June 2016 ISR data, CPD

Indicator (or dummy) variables where 1 = quantity present and 0 = quantity not present were created for gender, race, and ethnicity, various times of day, days of the week, months, and age ranges.

The original distribution of race/ethnicity codes used by CPD personnel in the field **RACE_CODE_CD** appears in Table 3. This report will focus on three racial/ethnic groups: White non-Hispanics, White Hispanics, and Black non-Hispanics. Stops associated with other races or ethnicities are dropped from the analysis. [4] This permits a clean focus on three mutually exclusive racial/ethnic groups most prevalent in Chicago. These three groups represent 54,116 out of 54,701 cases and 98.9 percent of ISR records for the period.

[4] The small number of Black Hispanics in the data are dropped so that the three groups of interest are completely exclusive of one another.

Table 3 Counts and categories for CPD variable for race and ethnicity (RACE_CODE_CD)

Description	Code	N	Percent
Asian Pacific Islander	API	417	0.76
Black	BLK	38,361	70.13
American Indian / Alaskan Native	I	98	0.18
Undocumented code	P	67	0.12
Black Hispanic	WBH	3	0.01
White	WHI	35	0.06
White	WHT	4,163	7.61
White Hispanic	WWH	11,557	21.13
	------------	-----------	-----------
	Total	54,701	100
Three group sub-total: White non-Hispanic, White Hispanic, Black non-Hispanic	Sub-Total	54,116	98.93

Note. Period = January-June, 2016. Source: CPD ISR data

The distribution across districts of the three predominant racial/ethnic groups among stopped civilians appear in Table 4. Among these three groups, Black non-Hispanic civilians are the group most frequently stopped, making up almost 71 percent of the stops of members of these three groups. Hispanic civilians comprised 21 percent of those stopped in these three groups. And White non-Hispanic stopped civilians occurred least frequently, appearing in about eight percent of the stops.

That said, the racial/ethnic mix often varied markedly by district. Stopped Black non-Hispanic civilians contributed 98 percent of the stops in district 3, but only sixteen percent among the stops in district 17. Stopped Hispanic civilians made up 61 percent of those stopped of these three groups in district 14, but less than one percent in district 3.

Table 4 Number of stopped civilians by district and race/ethnicity, and district percent by race/ethnicity

	White NH		Black NH		Hispanic		District total	
District	N	Percent	N	Percent	N	Percent	N	Percent
1	126	15.27	640	77.58	59	7.15	825	100
2	37	1.61	2,237	97.43	22	0.96	2,296	100
3	34	1.11	3,027	98.47	13	0.42	3,074	100
4	85	2.37	2,900	80.85	602	16.78	3,587	100
5	25	1.3	1,867	97.29	27	1.41	1,919	100
6	37	1.49	2,417	97.3	30	1.21	2,484	100
7	59	1.23	4,694	97.65	54	1.12	4,807	100
8	332	9.49	1,552	44.34	1,616	46.17	3,500	100
9	380	8.85	1,572	36.6	2,343	54.55	4,295	100
10	110	2.76	2,683	67.36	1,190	29.88	3,983	100
11	341	5.91	5,113	88.66	313	5.43	5,767	100
12	256	11.07	1,041	45.01	1,016	43.93	2,313	100
14	126	13.98	224	24.86	551	61.15	901	100
15	68	2.13	3,033	95.2	85	2.67	3,186	100
16	612	47.3	330	25.5	352	27.2	1,294	100
17	303	27.03	181	16.15	637	56.82	1,121	100
18	125	13.31	705	75.08	109	11.61	939	100
19	279	21.83	717	56.1	282	22.07	1,278	100
20	179	22.1	288	35.56	343	42.35	810	100
22	65	5.37	1,120	92.56	25	2.07	1,210	100
24	381	18.82	1,081	53.41	562	27.77	2,024	100
25	238	9.51	939	37.51	1,326	52.98	2,503	100
Total	4,198	7.76	38,361	70.89	11,557	21.36	54,116	100

Note. NH = non-Hispanic. Period = January-June, 2016. Source: CPD ISR data. Counts only shown for the three most predominant racial/ethnic combinations among stopped civilians. Percentages shown are the district share associated with each racial/ethnic combination.

7.4 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.

7.5 Units of analysis

The unit of analysis is the individual (person) stopped, that is, each individual stop.

7.6 Clustering

Multiple stops can and do occur within a single event. Further, events are nested within districts. Attempts to model these three levels – stops within events within districts – failed to converge. Therefore models presented control only for the clustering of stops within districts. Mixed effects models with stops at Level 1 and districts at Level 2 are used.

Future models will attempt to simultaneously control for the clustering of stops within events, and events within districts.

7.7 Geographies and Implications for Analyses

The current report uses Chicago Police Department districts as the geographic unit of clustering. Since there were only a small number of these this creates some analytic limitations (Bryan & Jenkins, 2016; Schmidt-Catran & Fairbrother, 2016). Given the limitations associated with only 22 grouping units, these analyses do not incorporate specific district-level predictors. They simply to allow the outcome to differ across districts, and incorporate district-to-district differences as random effects in these models.

There is one instance where a stop feature, the district-level proportion of stopped Black civilians, is used as part of diagnostic routines. It is probably advisable in the future to move to smaller within-district units of analysis such as beats within districts.

7.8 Outcome Variables

7.8.1 Overall descriptive statistics

Descriptive statistics on the binary outcome variables appear in Table 5 Details on the levels and patterns for each variable are described further below. [5]

[5] Earlier circulated analysis plans, and discussions with the City's and ACLU's experts referenced additional outcomes beyond those mentioned here. Those additional outcomes included any pat down hits where the latter were defined as either weapons or contraband, a pat down hits based only on contraband, any search hits resulting in resulting in either weapons or contraband, and any search hits resulting in contraband. Time did not permit including models of those other outcomes. Given the policy salience of weapons and weapons recovery, hit rate analyses here focused only on weapons.

Table 5 Descriptive statistics, binary outcome variables

Variable	Variable name	N	Min.	Max.	Mean	SD	Median	Sum
pat down conducted	dpat	54,116	0	1	0.339	0.473	0	18,364
pat down➔weapon (*)	pathit_w2	18,364	0	1	0.025	0.157	0	465
Search conducted	dsearch	54,116	0	1	0.177	0.382	0	9,595
Search ➔weapon v. 1 (*) [a]	se_hit_w	54,116	0	1	0.006	0.076	0	313
Search➔weapon v. 2 (*) [b]	se_hit_w2	9,595	0	1	0.027	0.163	0	263
Search➔weapon v. 3 (*) [c]	se_hit_w3	2,640	0	1	0.009	0.095	0	24
Enforcement action taken	denforce2	54,116	0	1	0.322	0.467	0	17,425

Note. For all binary outcomes, 1 = outcome occurred, 0 = did not occur
Note. For each outcome variable 1 = yes; 0 = no.
Source: Data from January-June 2016 ISR reports, CPD. MED. = median
(*) = dependent variable depends on selection through another dependent variable.
[a] On this version of the "search hits on weapon" variable, a hit counted as recovering either a firearm or another type of weapon or, as happened in ten instances, both.
[b] On this version of the "search hits on weapon" variable, a hit was as defined above in v. 1, but 0 was recoded to missing if officers did not check the search box on the ISR form. The discrepancy between v. 1 and 2 of the search weapon hit variable summary count indicates 50 instances where officers indicated a weapon or firearm was recovered from a search but the search box was checked "N". This was verified directly.
[c] On this version of the "search hits on weapon" variable, a hit was as defined above in v. 2, but 0 was recoded to missing if an arrest took place during the stop. This removed from the variable all weapons found incident to custodial searches conducted while taking a civilian into custody.

Some outcomes are dependent upon another particular post stop outcome taking place and are marked accordingly in the table (*). The pat down weapon hit variable, and three versions of the search weapon hit variable are all in this group. This means that models capturing sequential selection, as described above, are preferred.

The search weapons hit variable was constructed three different ways, resulting in three different totals for numbers of weapons recovered (Table 5).

With no restrictions, searches surfaced 313 weapons (version 1). If this variable is considered valid only if officers also checked the search box, then searches surfaced 50 fewer weapons, a total of 263 (version 2). If weapons found during searches are removed from consideration if the stop resulted in an arrest, then only 24 weapons surfaced (version 3). **The searches removed with this version could be searches incident to taking the civilian into custody. They also could be searches that *led* to discovering something that in turn led to an arrest. Because narrative fields were not analyzed for all records, we do not know how many of the search/arrest stops were searches incident to taking into custody vs. searches leading to an arrest.** We comment later on this exclusion when we get to the search outcome.

Descriptive statistics for the one categorical outcome analyzed appear in Table 6. The analyses of this outcome will consider all four possible combinations of outcomes when enforcement and pat down actions are jointly considered, but attention will center on the determinants outcome category 2 vs. outcome category 1. Among those experiencing no enforcement action during the stop, what was associated with receiving a pat down or not receiving a pat down?

Table 6 Descriptive statistics: Categorical outcome variable, pat down and enforcement combination

	Category	N	Percent
No pat down delivered, no enforcement action	1	23,236	42.94
pat down delivered, no enforcement action taken	2	13,444	24.84
No pat down delivered, enforcement action taken	3	12,508	23.11
pat down delivered and enforcement action taken	4	4,917	9.09
Missing	.	11	0.02
	Total	54,116	100.00

Note. 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.

7.8.2 Pat downs: Across groups and districts

In about a third of the stops – 18,364/54,116 or 33.9 percent – the officer delivered a pat down to the stopped civilian.

The number of pat downs in each district, for each of the three racial/ethnic groups, appears in Table 9. The number of pat downs ranged from a high of 2,377 in District 7, to a low of 162 in District 1 (the Loop).

Within each district, the proportion of each racial/ethnic group receiving a pat down appears in Table 10.

Looking at the overall numbers in the bottom of the table, the chances that a stopped civilian would be patted down does appear to depend on the race/ethnicity of the stopped civilian. Whereas about a third of stopped non-Hispanic Black civilians (34.9 percent) or stopped Hispanic civilians (34.7 percent) received a pat down, only about a quarter of stopped non-Hispanic White civilians received the same (23.3 percent).

To give the reader a sense of odds ratios that get 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. This is shown below in Table 7.

Table 7. Patted down vs. not patted down: Proportions and odds

Group	Proportion patted down vs. not patted down	Odds of being patted down vs. not patted down
White NH	0.233 /(1-0.233)	0.304
Black NH	.349/(1-.349)	0.536
Hispanic	.347/(1-.347)	0.531

For example, White non-Hispanics **odds** of being [patted down vs. not patted down] are derived by taking the proportion patted down and dividing it by the proportion not patted down. That creates odds of [pat down vs. no pat down] of .34. One could say: White non-Hispanics chances of getting a pat down versus not getting one were about 3 out of 10.

Odds are always about the chances of [this versus that]. **Odds are different from proportions** because proportions are just about the chances of this.

The reader can see that Black non-Hispanics' odds of [getting vs. not getting a pat down] were higher: their odds were .536. One could say: Black non-Hispanics' chances of getting a pat down vs. not getting one were around 5 in 10.

So Black non-Hispanics' odds of [getting vs. not getting a pat down] were higher than White non-Hispanics' odds. How much higher.

To find out one takes the ratio of the two odds, making an **odds ratio.** The odds ratio tells you how much higher or lower one group's odds were relative to the odds of the other group.

So to find the odds ratio of White NH/Black NH – the difference in the odds between the two groups – one divides the two odds.

Odds of Black NH [getting vs. not getting pat down]
--- = Odds Ratio of [Black NH vs. White NH] [getting vs. not getting pat down]
Odds of White NH [getting vs. not getting pat down]

So for

Black NH /White NH OR = .536/.304 = 1.765

That is, Black non-Hispanics' odds of [getting vs. not getting patted down] were **76 percent higher** than the odds for White non-Hispanics of [getting vs. not getting patted down].

The odds ratio for being Hispanic vs. White non-Hispanic = .531/.304 = 1.749

When you have an odds ratio close to 1 it means the two groups have about equal chances of [this vs. that] happening. Take the odds ratio for [getting vs. not getting patted down] for

Hispanic vs. Black non-Hispanic = .531/.536 = 0.991

Table 8 Odds ratios depicting ethnoracial differences in odds of getting vs. not getting patted down

Comparison of odds	OR
Black NH vs. White NH	1.765
Hispanic vs. White NH	1.749
Hispanic vs. Black NH	0.991

Odds ratios will be the main metric used to describe net impacts of racial or ethnic differences in analyses gauging net impacts.

Going back to Table 10, the last column in the table demonstrates that the chances of receiving a pat down depended on district context. In several districts (16, 18) police patted down around one out of six or one out of seven stopped civilians. In some districts that proportion was around one out of three (e.g., 3, 4, and 9). In a small number of districts that proportion hovered around one out of two (6, 7).

Table 9 Counts of pat downs by district and race/ethnicity

District	White NH	Black NH	Hispanic	Total
1	15	129	18	162
2	4	503	9	516
3	12	1,074	5	1,091
4	29	1,101	220	1,350
5	5	792	16	813
6	19	1,197	10	1,226
7	25	2,327	25	2,377
8	78	427	465	970
9	117	557	823	1,497
10	45	754	513	1,312
11	60	1,339	78	1,477
12	68	263	267	598
14	30	82	227	339
15	13	1,074	33	1,120
16	79	34	86	199
17	70	44	205	319
18	19	115	43	177
19	63	267	65	395
20	36	77	100	213
22	13	451	8	472
24	111	384	213	708
25	66	386	581	1,033
Total	977	13,377	4,010	18,364

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

Table 10 Proportion of stopped civilians patted down, by district and race/ethnicity

District	White NH	Black NH	Hispanic	Total
1	0.119	0.202	0.305	0.196
2	0.108	0.225	0.409	0.225
3	0.353	0.355	0.385	0.355
4	0.341	0.38	0.365	0.376
5	0.2	0.424	0.593	0.424
6	0.514	0.495	0.333	0.494
7	0.424	0.496	0.463	0.494
8	0.235	0.275	0.288	0.277
9	0.308	0.354	0.351	0.349
10	0.409	0.281	0.431	0.329
11	0.176	0.262	0.249	0.256
12	0.266	0.253	0.263	0.259
14	0.238	0.366	0.412	0.376
15	0.191	0.354	0.388	0.352
16	0.129	0.103	0.244	0.154
17	0.231	0.243	0.322	0.285
18	0.152	0.163	0.394	0.188
19	0.226	0.372	0.23	0.309
20	0.201	0.267	0.292	0.263
22	0.2	0.403	0.32	0.39
24	0.291	0.355	0.379	0.35
25	0.277	0.411	0.438	0.413
Total	0.233	0.349	0.347	0.339

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

7.8.3 If a pat down is conducted, are any weapons/firearms recovered?

How many actual weapons or firearms were recovered as a result of officers patting down stopped civilians? The counts appear in Table 11. For the period, the recovered weapons totaled 465. The number of recovered firearms/weapons varies from a low of 2 in District 20 to a high of 59 in District 7.

Table 11 Counts of weapons/firearms recovered from pat downs, by district, race/ethnicity

District	White NH	Black NH	Hispanic	Total
1	0	4	1	5
2	0	13	0	13
3	0	25	2	27
4	0	19	7	26
5	0	18	0	18
6	0	28	1	29
7	5	54	0	59
8	2	17	11	30
9	4	15	22	41
10	0	17	13	30
11	3	33	4	40
12	2	6	6	14
14	3	0	8	11
15	2	21	2	25
16	3	0	6	9
17	3	2	5	10
18	1	3	0	4
19	3	7	3	13
20	2	0	0	2
22	0	16	0	16
24	5	1	8	14
25	2	9	18	29
Total	40	308	117	465

Note. NH = non-Hispanic. Only weapons and firearms recovered in course of a pat down listed. Source: January-June 2016 ISRs, CPD.

The corresponding proportions appear in Table 12. Over all groups and over all districts about 2 1/2 percent of the pat downs yielded a weapon or a firearm. The weapon/firearm yield appeared somewhat higher for White non-Hispanics – around four percent – as compared to Black non-Hispanics – a little over two percent. The yield for Hispanic stopped and patted down civilians was between these two.

Table 12 Proportion of pat downs yielding a weapon/firearm by district and race/ethnicity

District	White NH	Black NH	Hispanic	Total
1	0	0.031	0.056	0.031
2	0	0.026	0	0.025
3	0	0.023	0.4	0.025
4	0	0.017	0.032	0.019
5	0	0.023	0	0.022
6	0	0.023	0.1	0.024
7	0.2	0.023	0	0.025
8	0.026	0.04	0.024	0.031
9	0.034	0.027	0.027	0.027
10	0	0.023	0.025	0.023
11	0.05	0.025	0.051	0.027
12	0.029	0.023	0.022	0.023
14	0.1	0	0.035	0.032
15	0.154	0.02	0.061	0.022
16	0.038	0	0.07	0.045
17	0.043	0.045	0.024	0.031
18	0.053	0.026	0	0.023
19	0.048	0.026	0.046	0.033
20	0.056	0	0	0.009
22	0	0.035	0	0.034
24	0.045	0.003	0.038	0.02
25	0.03	0.023	0.031	0.028
Total	0.041	0.023	0.029	0.025

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

7.8.4 Is a search conducted or not?

During the period, officers conducted 9,595 searches of stopped civilians who were in these three racial/ethnic groups. [6] This amounted to one search for every five to six stops. The numbers of searches by racial/ethnic group, and district, appear in Table 13. The largest number of searches of stopped Black non-Hispanic civilians took place in District 11, where there were over 1,000 searches during the first six months of 2016. The largest number of searches of stopped Hispanic civilians took place in District 9, where there were 318. In many districts the number of searches for a specific racial/ethnic group were quite low. This means the ethnoracial proportions of stopped civilians who were searched should be interpreted with caution in these instances.

[6] There were 173 cases where the search checkbox completed by police indicated that no search took place, but police also indicated that some type of contraband was recovered as part of a search. Regardless of search hit variables, if no search check box was checked no search was coded.

Table 13 Number of searches by district, by racial/ethnic group

District	White NH	Black NH	Hispanic	Total
1	18	90	10	118
2	3	268	5	276
3	8	464	4	476
4	16	447	81	544
5	4	400	3	407
6	13	513	9	535
7	9	896	15	920
8	26	168	148	342
9	63	226	318	607
10	21	472	201	694
11	76	1145	92	1313
12	39	207	133	379
14	25	34	96	155
15	12	598	13	623
16	101	65	100	266
17	57	30	134	221
18	10	71	14	95
19	55	147	65	267
20	24	44	60	128
22	8	235	5	248
24	86	219	137	442
25	53	200	286	539
Total	727	6939	1929	9595

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

The proportions appear in Table 14. Those proportions across all districts are roughly the same for all three different racial/ethnic groups. Across the entire city for each of the three groups of stopped civilians about one in five or one in six were searched.

Table 14 Proportion of stopped civilians who were searched, by racial/ethnic group and by district

District	White NH	Black NH	Hispanic	Total
1	0.143	0.141	0.169	0.143
2	0.081	0.12	0.227	0.12
3	0.235	0.153	0.308	0.155
4	0.188	0.154	0.135	0.152
5	0.16	0.214	0.111	0.212
6	0.351	0.212	0.3	0.215
7	0.153	0.191	0.278	0.191
8	0.078	0.108	0.092	0.098
9	0.166	0.144	0.136	0.141
10	0.191	0.176	0.169	0.174
11	0.223	0.224	0.294	0.228
12	0.152	0.199	0.131	0.164
14	0.198	0.152	0.174	0.172
15	0.176	0.197	0.153	0.196
16	0.165	0.197	0.284	0.206
17	0.188	0.166	0.21	0.197
18	0.08	0.101	0.128	0.101
19	0.197	0.205	0.23	0.209
20	0.134	0.153	0.175	0.158
22	0.123	0.21	0.2	0.205
24	0.226	0.203	0.244	0.218
25	0.223	0.213	0.216	0.215
Total	0.173	0.181	0.167	0.177

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

Because officers arresting or transporting a civilian are required to conduct custodial search before taking the stop civilian into custody, the numbers and proportions searched were re-run after excluding stops that resulted in an arrest. As discussed further below, we do not know if all these searches were custodial searches, or if some of them were searches *which led to* an arrest. Nonetheless, the numbers of those searched after excluding stops resulting in an arrest, and the proportion of non-arrested civilians in each racial/ethnic group, in each district, who were searched, appear in Table 15. Focusing only on those not arrested, these figures suggest that searches were conducted in about one out of 20 stops, and this proportion looked roughly comparable across the three different racial/ethnic groupings.

Table 15 Count and Proportion searched, by racial/ethnic group, by district: Stops leading to arrest excluded

District	Count: Searched				Proportion within Each Group Searched			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	11	22	4	37	0.096	0.039	0.075	0.051
2	1	107	4	112	0.029	0.052	0.19	0.053
3	3	167	3	173	0.103	0.062	0.25	0.063
4	5	116	20	141	0.068	0.046	0.038	0.045
5	0	120	1	121	0	0.078	0.042	0.076
6	5	191	5	201	0.179	0.093	0.192	0.095
7	1	239	6	246	0.02	0.06	0.136	0.061
8	7	54	64	125	0.023	0.038	0.043	0.039
9	22	94	131	247	0.066	0.067	0.062	0.064
10	2	92	51	145	0.023	0.041	0.05	0.043
11	7	206	19	232	0.026	0.051	0.083	0.051
12	3	30	50	83	0.014	0.036	0.054	0.042
14	6	17	28	51	0.058	0.085	0.06	0.066
15	1	118	6	125	0.018	0.047	0.079	0.048
16	21	5	20	46	0.041	0.019	0.078	0.045
17	19	6	35	60	0.074	0.039	0.068	0.065
18	1	15	7	23	0.009	0.024	0.071	0.027
19	11	31	5	47	0.051	0.054	0.026	0.047
20	11	13	19	43	0.068	0.051	0.065	0.061
22	0	67	2	69	0	0.074	0.095	0.07
24	39	90	52	181	0.117	0.096	0.11	0.104
25	8	58	66	132	0.042	0.074	0.061	0.064
Total	184	1858	598	2640	0.052	0.057	0.06	0.057

7.8.5 If a search is conducted, are any weapons recovered?

The counts and proportions of each racial group within each district producing a weapon or firearm or both as a result of a search appear in the following three tables.

In Table 16 counts and proportions, by district and by racial/ethnic group, are shown for all records for these three racial/ethnic groups. Note that the variable equals 0 if no weapons or firearms are found as a result of the search, and 1 if a weapon, or a firearm, or both, are found as a result of the search. Because some searches (10) resulted in both a weapon and a firearm, the number of weapons recovered is greater than the number of search “hits” for weapons or firearms.

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

Table 16 Searches resulting in weapons or firearms or both: No exclusions

District	Count				Proportion within Each Group Yielding Weapons Hit			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	0	3	0	3	0	0.005	0	0.004
2	0	13	0	13	0	0.006	0	0.006
3	0	13	1	14	0	0.004	0.077	0.005
4	2	15	5	22	0.024	0.005	0.008	0.006
5	0	25	0	25	0	0.013	0	0.013
6	0	14	1	15	0	0.006	0.033	0.006
7	3	45	0	48	0.051	0.01	0	0.01
8	0	5	3	8	0	0.003	0.002	0.002
9	2	6	12	20	0.005	0.004	0.005	0.005
10	1	7	3	11	0.009	0.003	0.003	0.003
11	0	29	2	31	0	0.006	0.006	0.005
12	1	5	2	8	0.004	0.005	0.002	0.003
14	0	1	7	8	0	0.004	0.013	0.009
15	0	18	0	18	0	0.006	0	0.006
16	4	1	4	9	0.007	0.003	0.011	0.007
17	0	1	4	5	0	0.006	0.006	0.004
18	1	2	0	3	0.008	0.003	0	0.003
19	1	4	1	6	0.004	0.006	0.004	0.005
20	1	0	1	2	0.006	0	0.003	0.002
22	0	13	0	13	0	0.012	0	0.011
24	3	6	3	12	0.008	0.006	0.005	0.006
25	2	4	13	19	0.008	0.004	0.01	0.008
Total	21	230	62	313	0.005	0.006	0.005	0.006

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

The total number of times a search resulted in a weapons "hit", as shown in Table 16, was 313. This translated to searches generating weapons, a weapons "hit rate," of 6/10ths of a percent. Descriptively speaking, that hit rate seemed closely comparable across the three racial/ethnic groups: 5/10ths of a percent for White Non-Hispanic stopped civilians and Hispanic stopped civilians, and 6/10ths of a percent for Black Non-Hispanic civilians.

Table 17 Searches resulting in weapons or firearms or both: Records included only if search check box also checked

District	Count				Proportion within Each Group Yielding Weapons Hit			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	0	2	0	2	0	0.022	0	0.017
2	0	10	0	10	0	0.037	0	0.036
3	0	11	1	12	0	0.024	0.25	0.025
4	2	14	5	21	0.125	0.031	0.062	0.039
5	0	21	0	21	0	0.052	0	0.052
6	0	10	1	11	0	0.019	0.111	0.021
7	2	41	0	43	0.222	0.046	0	0.047
8	0	4	3	7	0	0.024	0.02	0.02
9	1	6	8	15	0.016	0.027	0.025	0.025
10	1	5	3	9	0.048	0.011	0.015	0.013
11	0	23	2	25	0	0.02	0.022	0.019
12	1	5	2	8	0.026	0.024	0.015	0.021
14	0	1	6	7	0	0.029	0.063	0.045
15	0	16	0	16	0	0.027	0	0.026
16	4	1	4	9	0.04	0.015	0.04	0.034
17	0	1	3	4	0	0.033	0.022	0.018
18	1	2	0	3	0.1	0.028	0	0.032
19	1	2	1	4	0.018	0.014	0.015	0.015
20	1	0	1	2	0.042	0	0.017	0.016
22	0	10	0	10	0	0.043	0	0.04
24	2	5	2	9	0.023	0.023	0.015	0.02
25	2	3	10	15	0.038	0.015	0.035	0.028
Total	18	193	52	263	0.025	0.028	0.027	0.027

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

The picture shifts if search weapons hits are calculated only on records where officers also recorded that a search had taken place. As seen in Table 17, this reduced the number of searches generating a weapons hit to 263.

It also increased the search weapons hit rate to between two and three percent: 2.7 percent overall. Further, the weapons hit rate for the three different ethnic/racial groups, speaking descriptively, looked similar: non-Hispanic Black civilians generated a search weapons hit rate of 2.8 percent, slightly above the overall average, while White non-Hispanic civilians generated a search weapons hit rate slightly below the overall average, at 2.5 percent.

*Table 18 Searches resulting in weapons or firearms or both: Records included only if search check box also checked **and** no arrest associated with the stop*

District	Count				Proportion within Each Group Yielding Weapons Hit			
	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	1	1	0	0	0.333	0.006
4	0	2	0	2	0	0.017	0	0.014
5	0	0	0	0		0	0	0
6	0	2	1	3	0	0.01	0.2	0.015
7	0	5	0	5	0	0.021	0	0.02
8	0	1	0	1	0	0.019	0	0.008
9	0	1	0	1	0	0.011	0	0.004
10	0	0	0	0	0	0	0	0
11	0	1	0	1	0	0.005	0	0.004
12	0	0	0	0	0	0	0	0
14	0	1	0	1	0	0.059	0	0.02
15	0	2	0	2	0	0.017	0	0.016
16	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0
18	0	1	0	1	0	0.067	0	0.043
19	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0
22	0	0	0	0		0	0	0
24	0	0	0	0	0	0	0	0
25	1	2	3	6	0.125	0.034	0.045	0.045
Total	1	18	5	24	0.005	0.01	0.008	0.009

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

The picture shifts again if stops resulting in an arrest are removed, as shown in Table 18. Again, as in Table 17, only records where officers also checked the search box are considered. The removal of stops associated with an arrest is undertaken because officers are required to conduct a search prior to taking the arrested civilian into custody. Of course, this also may inappropriately remove some searches that *led to* an arrest.

Now the overall search weapons hit rate is slightly below one percent: 9/10^{ths} of a percent. (The reader can find the number of searches taking place for stops with no arrest in Table 15.) This is based on 24 searches generating a weapons hit out of 2,640 searches for stops with no arrests.

7.8.6 Quick aside: Search hits on weapons or contraband

Although this outcome is not analyzed statistically, for further descriptive context Table 19 shows search hit rates if a hit is widened to include **either** weapons **or** drug contraband. The numbers below are for all searches, with no restrictions.

Table 19 Search hit rates: Any weapons or contraband, by district and race/ethnicity

District	White NH	Black NH	Hispanic	Total
1	0.111	0.167	0.1	0.153
2	0	0.235	0	0.228
3	0.5	0.157	0.5	0.166
4	0.375	0.221	0.247	0.23
5	0.5	0.22	0.333	0.224
6	0	0.216	0.111	0.209
7	0.222	0.234	0.267	0.235
8	0.269	0.107	0.162	0.143
9	0.238	0.186	0.255	0.227
10	0.286	0.208	0.149	0.193
11	0.303	0.244	0.25	0.248
12	0.077	0.256	0.15	0.201
14	0.16	0.118	0.146	0.142
15	0.417	0.281	0.385	0.286
16	0.158	0.062	0.25	0.169
17	0.263	0.233	0.306	0.285
18	0.7	0.141	0.143	0.2
19	0.091	0.218	0.2	0.187
20	0.167	0.114	0.133	0.133
22	0.25	0.26	0	0.254
24	0.267	0.215	0.19	0.217
25	0.264	0.185	0.178	0.189
Total	0.227	0.22	0.203	0.217

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

For all three racial/ethnic groups, in roughly about one out of four or one out of five searches, weapons or contraband were discovered. The hit rates varied by district from about one out of four (district 11, district 22) to around one out of seven (district 8).

7.8.7 Is any enforcement action delivered or not?

CPD recorded four types of enforcement actions.

The numbers of each type appear in Table 20. Some type of enforcement action was delivered in 17,436 stops; out of 54,116 stops this means an enforcement action was delivered in 32.2 percent of these stops. [7]

[7] In seven instances, the recording of a specific enforcement action conflicted with the overall indicator completed by officers indicating whether any enforcement action was taken. In the analyses of any enforcement action taken (see section 10.5) the outcome analyzed aligned with the overall indicator completed by officers, not the recording of a specific enforcement action.

Table 20 Frequencies of different enforcement actions

Types of enforcement actions	N	Percent
ANOV (administrative notice of violation)	5,141	29.48
ARR (arrest)	8,037	46.09
OTH (other)	3,386	19.43
PSC (personal service citation)	861	4.94
Total	17,425	100.00

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD. 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 OTH and in one instance it was PSC. In statistical models using this outcome, or this outcome combined with a pat down, these 11 cases were set to missing on the outcome.

Counts of enforcement action of any type appear by district and race/ethnicity combination in Table 21. Police engaged in fewest enforcement actions in district 20, and the most in district 11.

Proportions of stops receiving any enforcement action, by district and race/ethnicity combination, also appear in Table 21. Stopped civilians in district 1 (the Loop) were the most likely to be targeted for enforcement; about 42 percent of stops in that district resulted in some kind of enforcement action by police. Overall, slightly over a quarter of stopped White non-Hispanic civilians received some type of enforcement action by the stopping officer. The corresponding proportion for stopped Black non-Hispanic civilians was around a third. The proportion for stopped Hispanic civilians was between these two.

Table 21 Counts and proportions of stopped civilians receiving any enforcement action, by district and race/ethnicity

	Count: Any enforcement action				Proportion: Any enforcement action			
District	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	56	271	25	352	0.444	0.423	0.424	0.427
2	7	586	7	600	0.189	0.262	0.318	0.261
3	13	911	2	926	0.382	0.301	0.154	0.301
4	25	731	174	930	0.294	0.252	0.289	0.259
5	6	579	7	592	0.24	0.31	0.259	0.308
6	20	904	13	937	0.541	0.374	0.433	0.377
7	17	1655	13	1685	0.288	0.353	0.241	0.351
8	81	521	610	1212	0.244	0.336	0.377	0.346
9	86	418	600	1104	0.226	0.266	0.256	0.257
10	35	939	347	1321	0.318	0.35	0.292	0.332
11	94	2021	123	2238	0.276	0.395	0.393	0.388
12	63	351	229	643	0.246	0.337	0.225	0.278
14	34	48	153	235	0.27	0.214	0.278	0.261
15	19	1113	16	1148	0.279	0.367	0.188	0.36
16	183	89	137	409	0.299	0.27	0.389	0.316
17	87	63	277	427	0.287	0.348	0.435	0.381
18	30	194	37	261	0.24	0.275	0.339	0.278
19	104	207	126	437	0.373	0.289	0.447	0.342
20	58	83	98	239	0.324	0.288	0.286	0.295
22	24	463	6	493	0.369	0.413	0.24	0.407
24	84	223	135	442	0.22	0.206	0.24	0.218
25	63	310	421	794	0.265	0.33	0.317	0.317
Total	1189	12680	3556	17425	0.283	0.331	0.308	0.322

Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

7.8.8 Pat down but no enforcement action

As described above, this outcome emerges from the procedural justice literature, and considers the relative likelihood of two joint outcomes.

In simultaneously considering whether the stopped civilian is patted down, and whether the stopped civilian receives any enforcement action, there are four possible sets of outcomes

1. Citizen is not patted down, nor does he/she receive any enforcement action.
2. Citizen *is* patted down, but no enforcement action taken.
3. No pat down, but enforcement action taken.
4. Pat down and enforcement action both taken.

The analyses reported here simultaneously contrasted option 1 above with each of the other three in a multinomial model. But the reporting of results focuses only on the contrast of 1 vs. 2. [8] Stated differently, does race or ethnicity affect the stopped civilians' odds of experiencing:

[A pat down but no enforcement action (2) vs. no pat down and no enforcement action (1)]?

Counts of stops where during a stop civilians were patted down by police but did not receive any enforcement action from the officer appear in

[8] The focus on this contrast of 1 vs. 2 emerges from the procedural justice literature. **Of course the other contrasts are important, and race or ethnicity differences can and do prove important in those other contrasts**. For example, impacts of ethnoracial differences on 1 vs. 3 are worthy of exploration. Those impacts, however, do not align with the procedural justice frame which is our conceptual starting point when considering this outcome.

Table 22, organized both by district and by race/ethnicity. This happened a total of 13,444 times during the timeframe. It occurred over thousand times each in districts 4, 7, 9, and 11. The number of times this occurred with stopped Black non-Hispanic civilians – 9,828 – was more than 10 times the corresponding number for stopped White non-Hispanic civilians.

Recall that the unit of analysis here is the stop. Therefore, there is no way of knowing how many times the **same** civilian was in a stop with a pat down but no enforcement.

Focusing just on stops where no enforcement action occurred, Table 24 indicates the fraction of *those* stops where a pat down occurred. So in essence, if no enforcement action took place what were the chances that a pat down simultaneously occurred?

Table 22 Counts and proportions of stops where civilians receiving pat down but no enforcement action, by district and race/ethnicity

DISTRICT	Count				Proportion			
District	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	10	88	6	104	0.079	0.138	0.102	0.126
2	4	352	3	359	0.108	0.157	0.136	0.156
3	8	798	3	809	0.235	0.264	0.231	0.263
4	22	882	169	1073	0.259	0.304	0.281	0.299
5	5	598	12	615	0.2	0.32	0.444	0.32
6	8	837	8	853	0.216	0.346	0.267	0.343
7	18	1762	20	1800	0.305	0.375	0.37	0.374
8	53	312	345	710	0.16	0.201	0.213	0.203
9	93	441	642	1176	0.245	0.281	0.274	0.274
10	34	543	361	938	0.309	0.202	0.303	0.236
11	51	930	52	1033	0.15	0.182	0.166	0.179
12	51	190	204	445	0.199	0.183	0.201	0.192
14	16	62	160	238	0.127	0.277	0.29	0.264
15	11	800	26	837	0.162	0.264	0.306	0.263
16	54	19	47	120	0.088	0.058	0.134	0.093
17	52	27	122	201	0.172	0.149	0.192	0.179
18	11	67	28	106	0.088	0.095	0.257	0.113
19	40	186	45	271	0.143	0.259	0.16	0.212
20	20	52	72	144	0.112	0.181	0.21	0.178
22	6	282	4	292	0.092	0.252	0.16	0.241
24	88	303	161	552	0.231	0.28	0.286	0.273
25	51	297	420	768	0.214	0.316	0.317	0.307
Total	706	9828	2910	13444	0.168	0.256	0.252	0.248

Note: The counts in the columns at left reflect the total number of stops where **both** of the following took place: 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. Source: January-June 2016 ISRs, CPD.

Overall, the proportion of non-enforcement delivered stops where a pat down occurred appears larger for stopped Black non-Hispanic civilians (.383) than for stopped White non-Hispanic civilians (. 235). The corresponding proportion for stops with Hispanic civilians (.364) seems quite close to the stops with Black civilians' proportion.

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? Table 23 compares the proportions of each ethnoracial group when overall representation in all stops is contrasted with representation in stops with pat downs and no enforcement. That comparison appears in the last column of the table. If the three groups were represented, proportionally, the same way in all stops, and in 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.

Results show that White Non-Hispanic civilians are under-represented in stops with pat downs but no enforcement, given their overall share of all stops. Whereas this group contributed 7.76 percent of all stops they contributed only 5.25 percent of stops with pat downs but no enforcement. Their chances of being in this type of stop were about 28 percent less than their overall stop share.

By contrast, Black Non-Hispanic civilians were somewhat over-represented in stops with pat downs but no enforcement (73 percent), given their overall stop share (71 percent). Their chances of being in this type of stop were about three percent higher than their overall stop share.

Table 23. Proportional representation, three ethnoracial groups: All stops vs. stops with (pat down and no enforcement action)

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	73.10	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.

These descriptive results suggest that proportional representation in stops with pat downs and no enforcement may not be comparable across the three ethnoracial groups considered. Statistical models presented later seek to learn whether that disproportionality can be linked exclusively to race or ethnicity.

These suggested disproportionalities between White Non-Hispanic and Black Non-Hispanics appear to loom larger if the focus drills down to consider just stops where no enforcement actions took place. In Table 24 the left hand columns are the same numbers as seen in

Table 22. But the proportions in the right hand columns differ because the (pat down + no enforcement) stop count is now being divided by *only* the total number of stops where no enforcement action took place. So the numbers on the left translate to higher proportions.

Table 24 Focusing ONLY on stops where no enforcement actions occurred: Counts and proportions of stops where civilians receiving pat down but no enforcement action, by district and race/ethnicity

	Count				Proportion			
District	White NH	Black NH	Hispanic	Total	White NH	Black NH	Hispanic	Total
1	10	88	6	104	0.143	0.238	0.176	0.22
2	4	352	3	359	0.133	0.213	0.2	0.212
3	8	798	3	809	0.381	0.377	0.273	0.377
4	22	882	169	1073	0.367	0.407	0.395	0.404
5	5	598	12	615	0.263	0.464	0.6	0.463
6	8	837	8	853	0.471	0.553	0.471	0.551
7	18	1762	20	1800	0.429	0.58	0.488	0.577
8	53	312	345	710	0.211	0.303	0.343	0.31
9	93	441	642	1176	0.316	0.382	0.368	0.369
10	34	543	361	938	0.453	0.311	0.428	0.352
11	51	930	52	1033	0.206	0.301	0.274	0.293
12	51	190	204	445	0.264	0.275	0.259	0.266
14	16	62	160	238	0.174	0.352	0.402	0.357
15	11	800	26	837	0.224	0.417	0.377	0.411
16	54	19	47	120	0.126	0.079	0.219	0.136
17	52	27	122	201	0.241	0.229	0.339	0.29
18	11	67	28	106	0.116	0.131	0.389	0.156
19	40	186	45	271	0.229	0.365	0.288	0.322
20	20	52	72	144	0.165	0.254	0.294	0.252
22	6	282	4	292	0.146	0.429	0.211	0.407
24	88	303	161	552	0.296	0.353	0.377	0.349
25	51	297	420	768	0.291	0.472	0.464	0.449
Total	706	9828	2910	13444	0.235	0.383	0.364	0.366

Note: The counts in the columns at left reflect the total number of stops where **both** of the following took place: the civilian was patted down **and** no enforcement action was recorded. The proportions in the right most columns express those counts as fractions of JUST stops where no enforcement actions occurred. Stops where any enforcement actions occurred are dropped from the entire table.
Note. NH = non-Hispanic. Source: January-June 2016 ISRs, CPD.

The fraction of non-enforcement action stops where a pat down is delivered varies across districts. The proportion is over half in Districts 6 and 7. It is around the fifth in Districts 1 and 2.

The discussion can be further specified if the focus shifts to proportional representation, for each ethnoracial group, in two sets of stops: all stops with no enforcement actions (NEA), and, of the latter, the subset that also included pat downs (PD + NEA). When each group's relative contribution to the latter (PD + NEA) is contrasted with its contribution to the former (NEA), one can learn whether, among stops with no enforcement action, certain groups of civilians were more or less likely to be patted down. See Table 25. The ratio of the two proportions, for each group, is shown in the right most column. As was seen before when all stops were considered in Table 23, White Non-Hispanic civilians were under-represented in the pat down stops, and Black Non-Hispanic civilians were somewhat over-represented. White Non-Hispanics' representation in the subset of no action stops with pat downs is one third less than their proportional representation in the set of all no enforcement action stops (ratio = .64). Black Non-Hispanics' representation was about four percent higher (ratio = 1.04) in the non enforcement stops with pat downs than it was in the set of all non enforcement stops.

Table 25 Focusing ONLY on stops where no enforcement actions occurred: Proportional representation, three ethnoracial groups: All non enforcement stops vs. non enforcement stops with pat down

	N: NEA	PR: NEA	N: PD + NEA	PR: PD + NEA	Ratio
White NH	3,009	8.2	706	5.25	0.64
Black NH	25,672	69.99	9828	73.10	1.04
Hispanic	7,999	21.81	2910	21.65	0.99
Total	36,680		13,444	100.00	

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. Ratio in right most column compares, for each group: [(proportional representation in PD + NEA stops) / (proportional representation in NEA stops)]

7.9 INDEPENDENT VARIABLES

Descriptive statistics for independent variables appear in Table 26. Some variables listed there are not used in the analyses but provide more detail about features of the data being examined.

Table 26 Descriptive statistics: Independent variables

	Variable	N	MIN	MAX	MEAN	SD	MED.
Black non-Hispanic civilian	dblack	54116	0	1	0.709	0.454	1
Hispanic civilian (d)	dhisp	54116	0	1	0.214	0.410	0
White civilian (*) (d)	dwhite	54116	0	1	0.078	0.268	0
Male civilian (d)	dmale	54116	0	1	0.866	0.341	1
Age in years (*)	age2	54112	7	89	29.568	13.355	25
Age in years (centered) (*)	c_age2	54112	-22.568	59.432	0.000	13.355	-4.568
Age 10-17 (*) (d)	age1017	54116	0	1	0.168	0.374	0
Age 18-25 (d)	age1825	54116	0	1	0.345	0.475	0
Age 25-35 (d)	age2635	54116	0	1	0.210	0.408	0
Age 36-45 (d)	age3645	54116	0	1	0.116	0.320	0
Age 46 and up (d)	age46pl	54116	0	1	0.161	0.368	0
District 1 (*) (d)	dist01	54116	0	1	0.015	0.123	0
District 2 (*) (d)	dist02	54116	0	1	0.042	0.202	0
District 3 (*) (d)	dist03	54116	0	1	0.057	0.231	0
District 4 (*) (d)	dist04	54116	0	1	0.066	0.249	0
District 5 (*) (d)	dist05	54116	0	1	0.035	0.185	0
District 6 (*) (d)	dist06	54116	0	1	0.046	0.209	0
District 7 (*) (d)	dist07	54116	0	1	0.089	0.284	0
District 8 (*) (d)	dist08	54116	0	1	0.065	0.246	0
District 9 (*) (d)	dist09	54116	0	1	0.079	0.270	0
District 10 (*) (d)	dist10	54116	0	1	0.074	0.261	0
District 11 (*) (d)	dist11	54116	0	1	0.107	0.309	0
District 12 (*) (d)	dist12	54116	0	1	0.043	0.202	0
District 14 (*) (d)	dist14	54116	0	1	0.017	0.128	0
District 15 (*) (d)	dist15	54116	0	1	0.059	0.235	0
District 16 (*) (d)	dist16	54116	0	1	0.024	0.153	0
District 17 (*) (d)	dist17	54116	0	1	0.021	0.142	0
District 18 (*) (d)	dist18	54116	0	1	0.017	0.131	0
District 19 (*) (d)	dist19	54116	0	1	0.024	0.152	0
District 20 (*) (d)	dist20	54116	0	1	0.015	0.121	0
District 22 (*) (d)	dist22	54116	0	1	0.022	0.148	0
District 24 (*) (d)	dist24	54116	0	1	0.037	0.190	0
District 25 (*) (d)	dist25	54116	0	1	0.046	0.210	0
Weekend (Sat, Sun) (d)	wknddum	54116	0	1	0.266	0.442	0
Midnight to 3 AM (*) (d)	dhr0003	54116	0	1	0.080	0.271	0
3 AM – 6 AM (d)	dhr0306	54116	0	1	0.018	0.133	0
6 AM – 9 AM (d)	dhr0609	54116	0	1	0.042	0.200	0
9 AM – noon (d)	dhr0912	54116	0	1	0.141	0.348	0
Noon – 3 PM (d)	dhr1215	54116	0	1	0.164	0.370	0
3 PM – 6 PM (d)	dhr1518	54116	0	1	0.132	0.339	0
6 PM – 9 PM (d)	dhr1821	54116	0	1	0.231	0.421	0
9 PM – 11:59 (d)	dhr2123	54116	0	1	0.193	0.394	0
Vehicle stop (d)	dvehstop	54116	0	1	0.074	0.261	0
ISR missing event no. (d)	eventmis	54116	0	1	0.008	0.088	0

Note. (d) = binary variable; 1 corresponds to variable name, 0 otherwise. (*) = variable not used in multivariate analyses. MED = median. Source: January-June 2016 ISRs, CPD.

7.10 Analytic sequence: Rationale and details

The specific analytic sequence depends in part on the specific outcome being examined. Nonetheless, the following broad outlines may be helpful.

Each random sample was analyzed separately. As mentioned above, this allowed learning whether crucial statistically significant impacts could be internally replicated across the two samples. If they could, that would suggest more confidence in findings.

7.10.1 Outcomes where there is no necessary selection process

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 any enforcement action was delivered; and
- whether a pat down occurred in a stop in which no enforcement action took place

For the first three of these outcomes the analytic sequence is as follows.

(1) A series of mixed effects logit models determine (a) whether there is significant variation in the outcome across districts; (b) the gross impacts of race and ethnicity on the outcome in question; (c) the net impacts of race and ethnicity after controlling for other covariates. All these mixed effects models control for the district context as a random effect.

As noted earlier **these are at heart multiple regression models**, incorporating necessary improvements to align with the clustered nature of the data and the binary or categorical nature of the outcomes.

(2) Results from the net impact model are subjected to diagnostics. These seek to gauge the extent to which observed or unobserved selection is potentially problematic. These diagnostics shape whether the interpretation of any observed net race or ethnicity effects should be along correlational or causal lines.

(3) Propensity score matching models are built separately for two contrasts: White non-Hispanics versus Black non-Hispanics; Hispanics versus White non-Hispanics.

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

The steps of the propensity score matching models were as follows.

(a) For each contrast a mixed effects logit model using the same covariates that appeared in the regression models predict the race or ethnicity contrast. These models generate a propensity score for each stopped civilian included in that contrast – for example the propensity of the stopped civilian to be Black and non-Hispanic instead of White and non-Hispanic.

(b) One-to-one propensity score matching is carried out, and nonmatched cases are dropped. The matching is done with various caliper restrictions: within .10 or .07 or .06 of a standard deviation on the propensity score. Most models use just the most stringent matching threshold, .06.

(c) "Treatment effects", that is the impacts of being Black and non-Hispanic versus White and non-Hispanic, or being Hispanic versus White and non-Hispanic, are estimated for each outcome using just the matched cases. Again, mixed effects logit models with random effects for districts are used for this estimation.

(4) Results from the propensity score matching models are subjected to diagnostics to learn whether selection on observed covariates is potentially problematic. If there is an observed selection problem then the interpretation of any effects seen in the propensity models should be along correlational rather than causal lines.

(5) Results from the propensity score matching models are subjected to a sensitivity analysis to estimate the extent to which unobserved selection is potentially problematic. If the results of the sensitivity analysis indicate that this could be a concern, interpretation of any effects seen in these models should be along correlational rather than causal lines.

For the last of these outcomes, a series of multinomial mixed effects logit models indicate whether the outcome of interest varies across districts, the size of the gross race and ethnicity impacts, and net race and ethnicity impacts after controlling for other covariates.

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

7.10.2 Outcomes where there is sequential selection

Several outcomes are observed only if something prior takes place. This brings up the problem of sequential selection mentioned earlier.

Two complementary approaches get applied to these outcomes. The first approach employs mixed effects logit models with districts as random effects as was done previously. But these models will include as an additional predictor the probability of being selected for the outcome. For example if the outcome is whether or not a pat down resulted in a weapon being discovered, the predicted probability that a pat down would take place is included as a predictor.

The second approach uses a single level model with error terms clustered by district: a Heckman selection model for a binary (probit) outcome (Baum, 2006). [9]

8 A PRIORI STATISTICAL POWER CALCULATIONS

Statistical power calculations were carried out using GPower software (Faul, Erdfelder, Buchner, & Lang, 2009). [10]

Although preferences differ depending upon the discipline in question, in psychology an acceptable level of statistical power is usually considered to be .80 or higher (Cohen, 1988,

[9] In Stata this is heckprobit.

[10] These calculations ignore the clustered nature of the data. Power calculations will be replicated at a later time using simulation software that recognizes such clustering in the data (Browne, Lahi, & Parker, 2009). The OD power estimation program for hierarchical models is inappropriate here because it explicitly assumes an experimental rather than a nonexperimental set up (Spybrook, Raudenbush, Congdong, & Martinez, 2009).

1992). One minus the level of statistical power represents the Type II error rate, that is, the chances that a significant difference will be overlooked.

Detailed power calculations were conducted for the first outcome, whether or not the stopped civilian received a pat down. This outcome was selected for detailed power analysis because it is relevant to all stopped civilians. Power analysis considers whether a more stringent alpha level, for example .01 or .001 instead of .05, was desirable given the large number of stops being examined.

Power, with a focus on the impact of the binary variable for Black vs. non-black stopped civilian was estimated for a multiple logistic regression model with 26,000 cases, roughly the number of stops in each 50 percent random sample. The power analyses were further tuned to reflect the mean on the outcome, and the overlap between being patted down and being a Black civilian. Power curves were estimated for an odds ratio associated with the Black variable that ranged from 1.05 to 1.30 in .05 increments. For each specific odds ratio, different power curves were run assuming either 10 percent, 20 percent, or 30 percent of the outcome variation was explained by other predictors. For each specific combination of the above, power curves were run for two tailed alpha levels of .05, .01, and .001. Results from these power curves are summarized in Table 27. Entries at or exceeding the recommended power level of .80 appear in **bold**. A sample power curve appears in Figure 4.

Table 27 A Priori statistical power estimates for pat down outcome

		Alpha level (two tailed)								
		0.05			0.01			0.001		
		R squared other			R squared other			R squared other		
		0.1	0.2	0.3	0.1	0.2	0.3	0.1	0.2	0.3
	OR									
Race impact expressed as an odds ratio (OR)	1.05	0.37	0.33	0.3	0.17	0.15	0.13	0.05	0.04	0.03
	1.1	**0.89**	**0.85**	**0.8**	0.73	0.66	0.59	0.46	0.38	0.31
	1.15	**1**	**0.99**	**0.99**	**0.98**	**0.97**	**0.94**	**0.92**	**0.87**	**0.8**
	1.2	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0.99**	**0.98**
	1.25	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	1.3	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**

The summary table suggests that with an a priori alpha level of .05 an odds ratio associated with the race variable of 1.1 or higher has an 80 percent chance or better of being detected, regardless of how much of the outcome is explained by the other variables in the model.

If an a priori alpha level of .01 or .001 is adopted, power is estimated to be acceptable if the odds ratio associated with the race variable is 1.15 or higher, regardless of how much of the outcome is explained by the other variables.

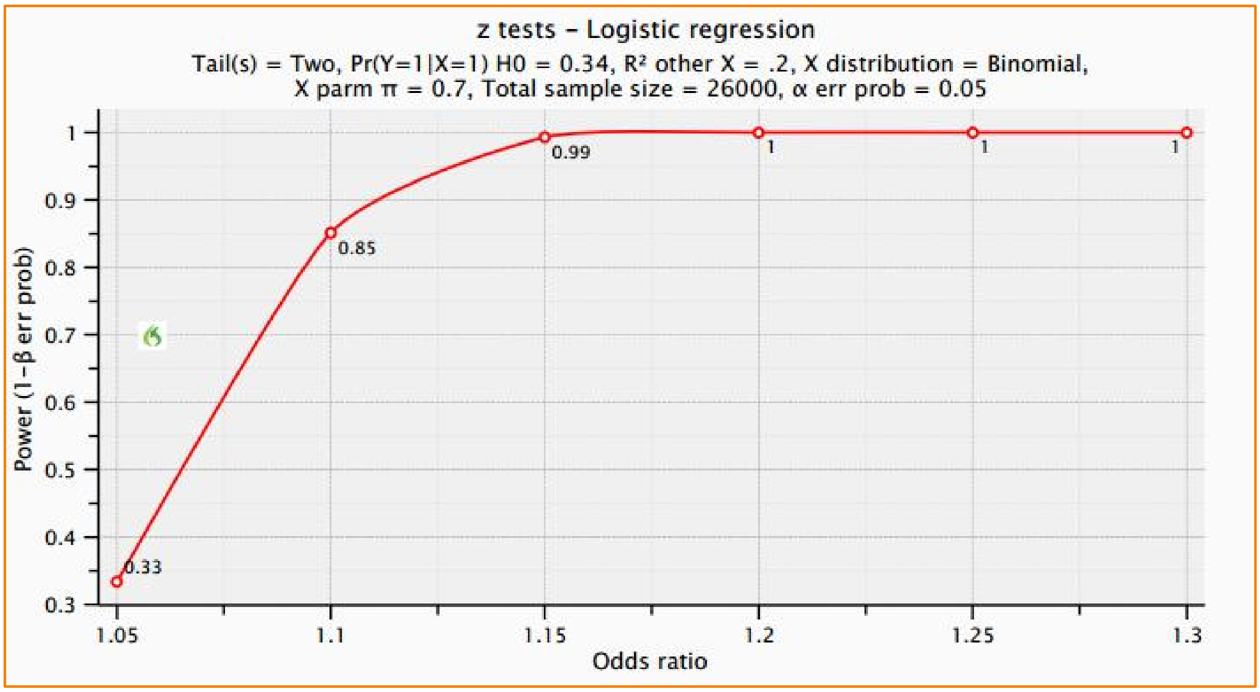
It bears repeating that what is in question here is the odds ratio associated with the race variable, not a percentage difference on the outcome. The race- or ethnicity-linked percentage difference associated with a specific odds ratio depends upon the mean score on the outcome for the group against which stopped Black civilians are being contrasted.

In light of these power estimations, despite the large number of cases being analyzed, the authors decided to use a conventional two tailed alpha level of .05.

These power analyses conducted here are just for single specific outcomes. If multiple outcomes correlate strongly with one another then the experiment-wise alpha level could inflate to something higher than .05.

In fact, save for one exception, correlations across outcomes are below |.04|. The one exception is getting or not getting an enforcement action, and being searched (Kendall's tau = .18 in a randomly sampled 50 percent of the records).

Figure 4 Sample power curve for race impact from power analysis of pat down outcome



9 BACKGROUND ON ANALYTIC CHOICES

9.1 DIAGNOSTICS AND RATIONALE

9.1.1 Regression Diagnostics

In the regression models several types of diagnostics are undertaken. Sequence includes the following:

(a) 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).

Several features of residuals are examined. Throughout, the Anscombe residuals are used. These are "usually close to the standardized deviance" (Hilbe, 2009: 279).

(b) The distribution of residuals is examined for normality and outliers.

(c) Predicted scores on the X axis are plotted against residuals on the Y axis using a LOWESS smoothed scatterplot line (Charpentier, 2013; Cleveland, 1979).

(d) Geographic residuals at the district level and their 95 percent confidence intervals indicate whether there is a geographically patterned lack of fit for the model.

(e) The relationship between residuals and a non-randomly selected covariate is examined to learn whether residuals appear correlated with this covariate.

9.1.2 Propensity models: Assessing selection on observables

Following the matched propensity score models, the sequence of balance diagnostics suggested by Austin (2009) are undertaken. Overall balance statistics suggested by Rubin are considered as well (Rubin, 2001).

9.1.3 Propensity models: Assessing selection on unobservables

Sensitivity analyses are undertaken to address the potential problem of selection on unobserved factors. Sensitivity tests of propensity score models gauge the impact of unobserved factors that might be simultaneously influencing both race and the outcome, or ethnicity and the outcome (Aakvik, 2001; Becker & Calaiendo, 2007; Guo & Fraser, 2015: 358-359; Mantel & Haenszel, 1959). "A sensitivity analysis in an observational study addresses this possibility: it asks what the unmeasured covariate would have to be like to alter the conclusions of the study" (Rosenbaum, 2005: 1809). "It is not possible to estimate the magnitude of selection bias with nonexperimental data. We rather calculate upper and lower bounds on the test-statistics to test the null hypothesis" of no race or ethnicity impact on the outcome (Aakvik, 2001: 129).

The sensitivity test starts by assuming "no unobserved selection bias" and setting $e^{\gamma} = 1$ (Aakvik, 2001: 130). This parameter, e^{γ} or Γ (gamma) in the program, is then adjusted upward, in increments of .05, to 2. "If eγ close to 1 changes the inference about the training effect, then estimated training affects are said to be sensitive to unobserved selection bias. However, if a large value of e^{γ} does not alter inferences about the training effect the study is not sensitive to selection bias" (Aakvik, 2001: 130). For the situation here, substitute race or ethnicity effect for "training effect."

9.2 MULTICOLLINEARITY IN REGRESSION MODELS

The degree of multicollinearity in regression models was gauged by examining variance inflation factors. Inclusion of dummy variables capturing specific districts created significant multicollinearity; that is, VIFs were **substantially** above 4.0. To avoid this problem districts were treated as random effects in a mixed effects model. All predictors had VIFs below 4.0.

Multicollinearity was reassessed when a predicted scores' inclusion was mandated given a conditioned outcome.

9.3 CLUSTERED DATA

Events were clustered within districts, and, if there were multiple stops per event, stops were clustered within events. Such clustered data require mixed effects models for a number of reasons (Snijder & Bosker, 2012). Modeling efforts recognizing both levels of clustering fail to converge. Therefore, the regression models reported here are two level mixed effects models recognizing only the clustering of stops within districts, and ignoring the clustering of stops within events. The implications of ignoring the clustering of stops within events is not known at this time. Nevertheless, given the considerable community criminology research on neighborhood effects (Sampson, Morenoff, & Gannon-Rowley, 2002) and policing work on the ecology of policing behavior across districts (Klinger, 1997), the geographic clustering was judged the more important of the two clustering sources.

9.4 GEOGRAPHY

Crime and delinquency patterns, that is the levels of each and the mix within each, vary geographically. Over a century of work establishes this point (Bursik & Grasmick, 1993; Taylor, 2015). At the same time, ecological models predicting crime and delinquency rates can never completely explain all of this variation (Pratt & Cullen, 2005). Excluding geography results in a theoretically under-specified model. Stated more simply, such a model leaves out causes of the outcome that we already know are important.

Geography is also important from a police perspective. Recent ecological theorizing on policing suggests (Klinger, 1997), and research supports the idea (Taniguchi, 2010), that within a single police department, police district-level norms exist about how to respond to crimes and calls for service of varying seriousness.

In the mixed effects models district context is always included as a random effect. This means that the mean score on the outcome varies across districts. As noted above, due to multicollinearity concerns it was not possible to include dummies for district variables. So geography was modeled as a random effect.

What is left over geographically proves interesting and potentially important. On the pat down outcome we observe **significant district level discrepancies from predicted pat down outcomes for a small number of districts**. These discrepancies may be important and may warrant further investigation.

10 RESULTS

10.1 DID A PAT DOWN OCCUR?

10.1.1 Regression

10.1.1.1 Results

Results appear in Table 29. In both random samples stopped Black non-Hispanic civilians were about 24 to 25 percent more likely to be [patted down versus not patted down] compared to stopped White non-Hispanic civilians ($p < .001$). This suggests a net race impact on the outcome controlling for the other covariates and for district context.

A similarly sized and similarly significant ($p < .001$) net ethnicity impact appeared as well. Stopped Hispanic civilians were also about 23 to 28 percent more likely to be [patted down versus not patted down] after controlling for the other covariates and for district context.

Gender proved significant ($p < .001$) as well. Males were about three times more likely to be [patted down versus not patted down].

For age, the reference group was those younger than 18. Compared to that reference age group, those aged 18 to 25 were significantly more likely ($p < .001$) to be patted down. Stopped civilians older than 36 were significantly less likely to be patted down compared to the youngest reference group ($p < .001$).

The odds of being [patted down versus not patted down] seemed to wane in the later months in the series. Compared to the reference month of January, those odds were about 15 percent lower in April ($p < .05$ or $p < .001$, depending on sample), about 22 percent lower in May ($p < .001$), and about 30 percent lower in June ($p < .001$).

If a civilian was stopped on the weekend, his or her chances of being patted down were about 10 to 13 percent higher ($p < .001$).

The reference time used was stops between midnight and 3 AM. Compared to that timeframe, pat downs were significantly more likely between 3 and 6 AM ($p < .05$ or ($P < .001$, depending on sample), but significantly less likely at all other times ($p < .01$ or $p < .001$, depending on sample and specific time block).

If the stop was flagged as a vehicle stop, the odds of a pat down were significantly higher, anywhere from 37 to 51 percent higher depending on the sample ($p < .001$).

In the first random sample but not the second random sample those stops missing an event number were significantly more likely to include a pat down ($p < .05$).

These results suggest a significant impact of both race and ethnicity on the likelihood of a pat down taking place during the stop. This appears as a net impact because it persists after controlling for other covariates and for district context.

Diagnostics suggest, however, that it might be unwise to interpret this net connection as anything more than correlational. Details appear below in the next section.

Table 28. Gross ethnoracial impacts on predicted probabilities of receiving a pat down

First random sample

Ethnoracial category	Predicted proportion patted down	Standard error of proportion
White NH	0.232	0.003
Black NH	0.347	0.001
Hispanic NH	0.35	0.002

Second random sample

Ethnoracial category	Predicted proportion patted down	Standard error of proportion
White NH	0.233	0.003
Black NH	0.35	0.001
Hispanic NH	0.344	0.002

Note. 2016, Jan.-June ISR data. White NH predicted proportion is significantly lower

The modeled results can be used to describe **gross ethnoracial impacts** on chances of getting a pat down as well. See

Table 28. It shows the predicted probability of receiving a pat down, based on all the factors used in the model, for the first and second random samples. The standard error around each proportion is shown as well. If two proportions are farther apart than two standard errors from each other then they are significantly different in statistical terms. The table shows that, in both random samples stopped White non-Hispanic civilians are predicted to be significantly ($p < .001$) less likely to receive a pat down compared to the other two groups, Black non-Hispanic and Hispanic stopped civilians. In both samples the predicted probabilities for a pat down are at least 10 percent lower for White non-Hispanic stopped civilians.

Table 29. Predicting pat down occurrence: Mixed effects logit models

		First random sample		Second random sample	
Fixed effects		b	OR	b	OR
Black civilian	dblack	0.227***	1.255***	0.214***	1.239***
Hispanic civilian	dhisp	0.244***	1.277***	0.210***	1.234***
Male	dmale	1.136***	3.114***	1.175***	3.240***
Age 18-25	age1825	0.196***	1.217***	0.136***	1.146***
Age 26-35	age2635	-0.0661	0.936	-0.0722	0.93
Age 36-45	age3645	-0.662***	0.516***	-0.524***	0.592***
Age 46 and up	age46pl	-1.166***	0.311***	-1.191***	0.304***
February	dfeb	0.0178	1.018	0.0418	1.043
March	dmar	-0.00820	0.992	0.0474	1.049
April	dapr	-0.185***	0.831***	-0.122*	0.885*
May	dmay	-0.284***	0.753***	-0.218***	0.804***
June	djun	-0.407***	0.666***	-0.329***	0.720***
Weekend	wknddum	0.128***	1.136***	0.0977**	1.103**
3 - 6 AM	dhr0306	0.246*	1.278*	0.418***	1.519***
6 - 9 AM	dhr0609	-0.932***	0.394***	-0.792***	0.453***
9 - 12 AM	dhr0912	-0.647***	0.523***	-0.697***	0.498***
12 - PM	dhr1215	-0.474***	0.622***	-0.447***	0.639***
3 - 6 PM	dhr1518	-0.311***	0.733***	-0.244***	0.784***
6 - 9 PM	dhr1821	-0.346***	0.708***	-0.312***	0.732***
9 - 12 PM	dhr2123	-0.268***	0.765***	-0.180**	0.835**
Vehicle stop	dvehstop	0.319***	1.376***	0.412***	1.510***
Missing event no.	eventmis	0.316*	1.371*	0.241	1.273
Constant		-1.347	0.26	-1.443	0.236
Random effects					
	District variance	0.134**		0.140**	
Observations		27,058		27,058	
BIC		31699		31817	
Number of groups		22		22	
Note.	*** p<0.001, ** p<0.01, * p<0.05				

Source: January – June 2016 ISR data, CPD

Note. For sample 1: Null model: LR $\chi 2$ test vs. logistic model = 823.76; $p < .001$; BIC = 33,848.69

For sample 2: Null model: LR $\chi 2$ test vs. logistic model = 810.58; $p < .001$; BIC = 33,891.18

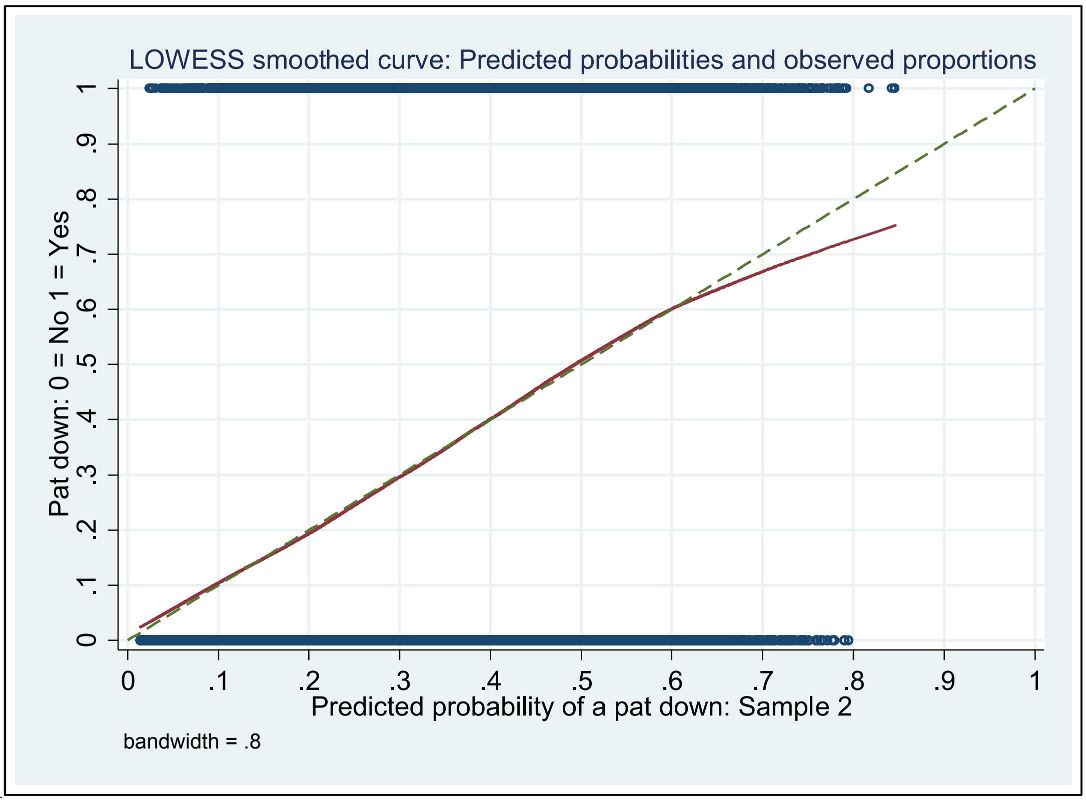
10.1.1.2 Model diagnostics

10.1.1.2.1 Patterns

Diagnostics of both the predicted scores and residuals revealed areas of concern.

Starting with predicted scores, LOWESS smooth curves linking predicted probabilities with observed outcome scores showed a significant lack of fit above predicted probabilities of around .7. This occurred in both random samples. The relationship for the second sample appears in Figure 5. It shows that predicted probabilities that a pat down would occur started to be markedly lower than the observed probabilities as the observed proportion patted down climbed above .70.

Figure 5 Predicted probabilities fit to observed proportions: pat down outcome, sample 2



R

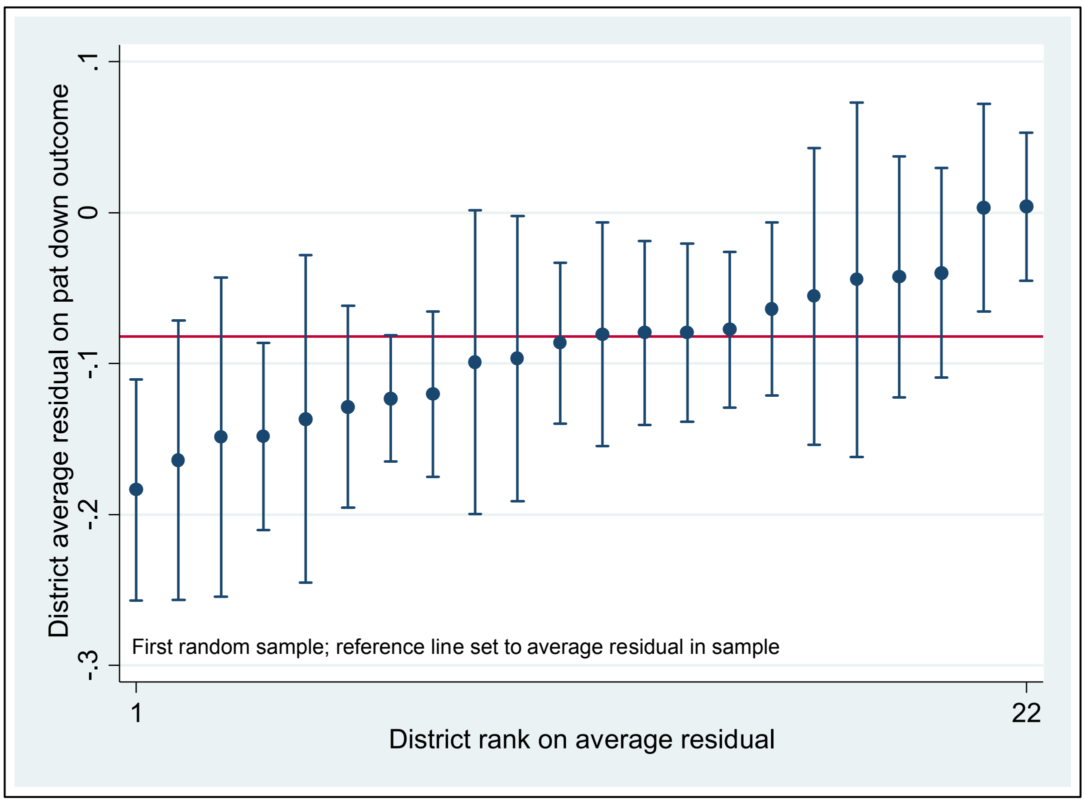
Plots of predicted scores against residuals with a superimposed LOWESS smoothed curve showed no relationship between the two (results not shown).

Residuals appeared to be potentially problematic in two ways: geographically, and in relationship to at least one covariate.

The average district-level residuals for the first sample appear in Figure 6. The reference line shown corresponds to the overall residual. [11] Starting on the left-hand side of the figure, the first district (district 16) and fourth district (district 2) had residuals significantly ($p < .05$) below the average. This means that after taking the predictors into account, stops in these districts were predicted to be significantly less likely to result in a pat down. In district 16, 26 percent of the stopped pedestrians were non-Hispanic Black civilians, and in district 2 97 percent were in the same group.

In districts, 6 and 7, the average residual was significantly above the average. In both districts approximately 98 percent of the stopped civilians were Black non-Hispanics. Because this is a positive average residual, it suggests that a significantly higher fraction of stopped civilians were patted down than factors in the model led us to expect.

Figure 6. Average pat down residuals, by district, first random sample



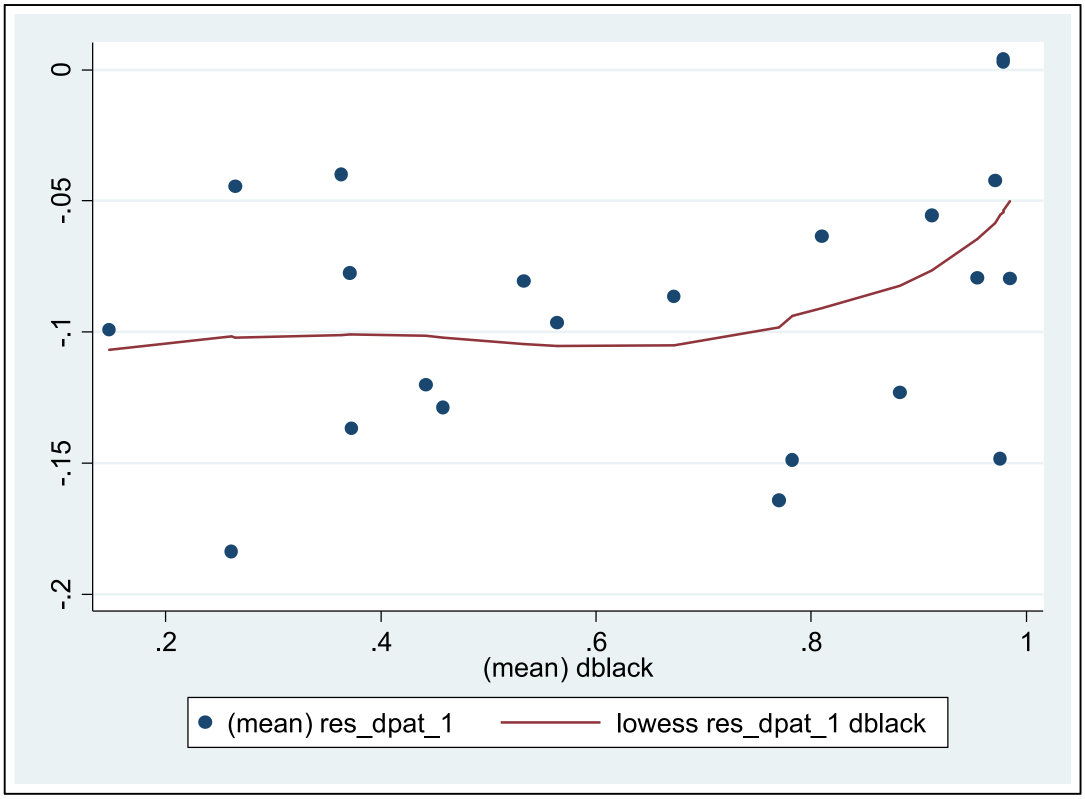
Note. January – June 2016 CPD ISR stop data. 95 percent confidence intervals shown

The relationship between the district residuals and the percent of stopped civilians who were Black non-Hispanic in that district appears in Figure 7. The smoothed LOWESS curve suggests that district level residuals trended upward if more than about 80 percent of the stopped civilians in the district were Black non-Hispanic. This suggests that non-modeled factors associated with the racial mix of stopped civilians in these districts were contributing to higher fractions of stops

[11] The overall residual is not zero because more stopped civilians were not patted down than were. If the outcome was not patted down, scored a zero, the residual is automatically a negative number.

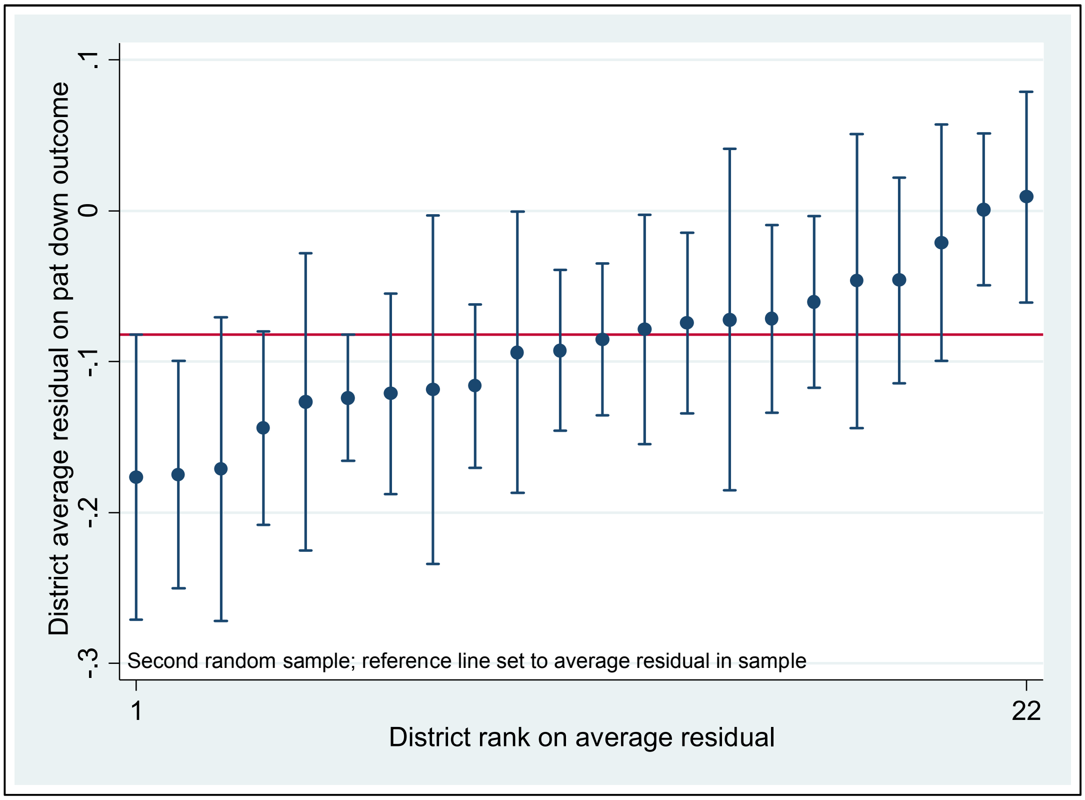
resulting in pat downs. This depicted relationship uses a district level covariate, and as pointed out previously, there are problems with using district factors given the low number of districts. **So this pattern should be considered exploratory only.**

Figure 7 First random sample: District pat down residual and percent stopped civilians who are Black



The caterpillar plot of district residuals for the second sample appears in Figure 8. Starting again on the left hand side of the plot, the district second from left, district 16, had an average residual significantly below the mean for the sample. This meant that fewer stopped civilians were patted down in this district than expected given the features in the model and the behavior of the other districts. This same departure from normality was noted with the results from the first random sample. In this second random sample, 25 percent of stopped civilians in this location were non-Hispanic Black.

Figure 8 Second random sample: District pat down residual and percent stopped civilians who are Black

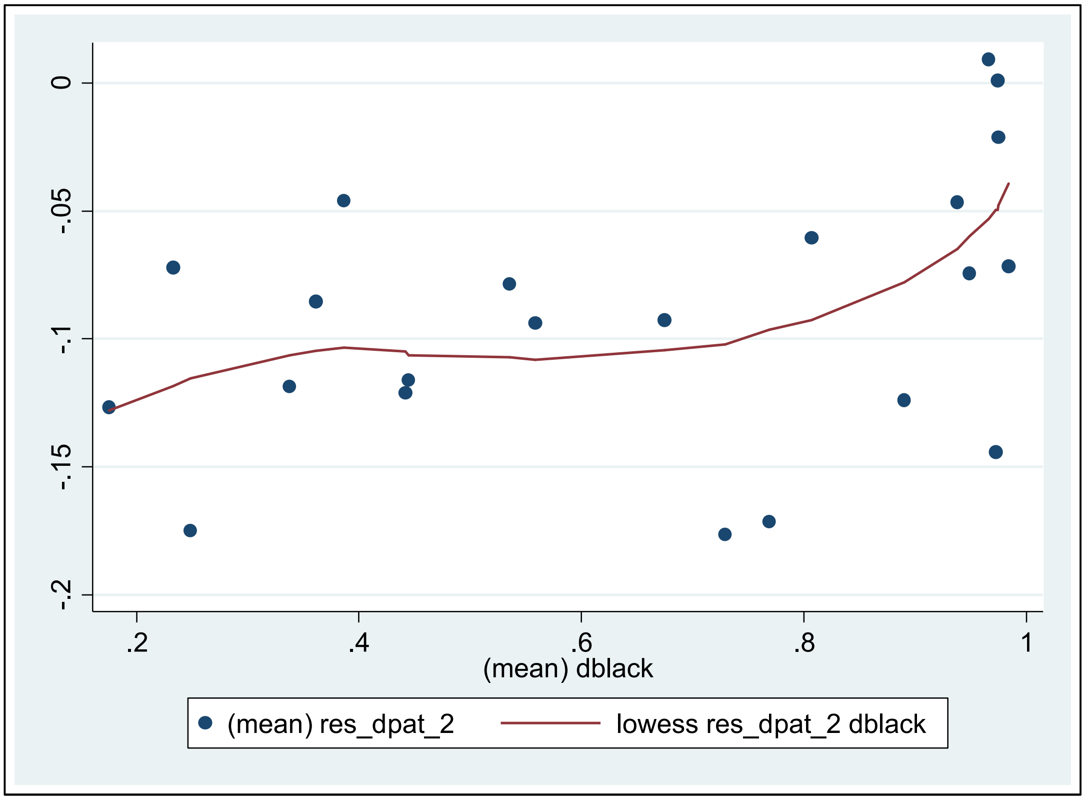


Note. January – June 2016 CPD ISR stop data. 95 percent confidence intervals shown

On the right-hand side of Figure 8, two districts have an average residual significantly above the overall average. These are Districts (left to right) 7 and 6. In both these districts approximately 97 percent of stopped civilians were Black non-Hispanic. These two districts also surfaced in the results from the first random sample as locations with significantly higher than average residuals. Again, the implication is that more pat downs were occurring in these locations than were anticipated by the features included in the model.

The connection for the second random sample between these district residuals and the percentage of stopped civilians in the district who were Black is displayed in Figure 9.

Figure 9 Second random sample: District pat down residual and percent stopped civilians who are Black



Again, as was seen with the first random sample, the smoothed LOWESS curve suggests that residuals were trending upward when more than about 80 percent of stopped civilians in a district were Black. **This depicted relationship should be considered exploratory only.**

Results from both samples suggest there is one district, 16, where significantly fewer persons are patted down than expected, and two districts, 6 and 7, where more stopped civilians are patted down than the model expects.

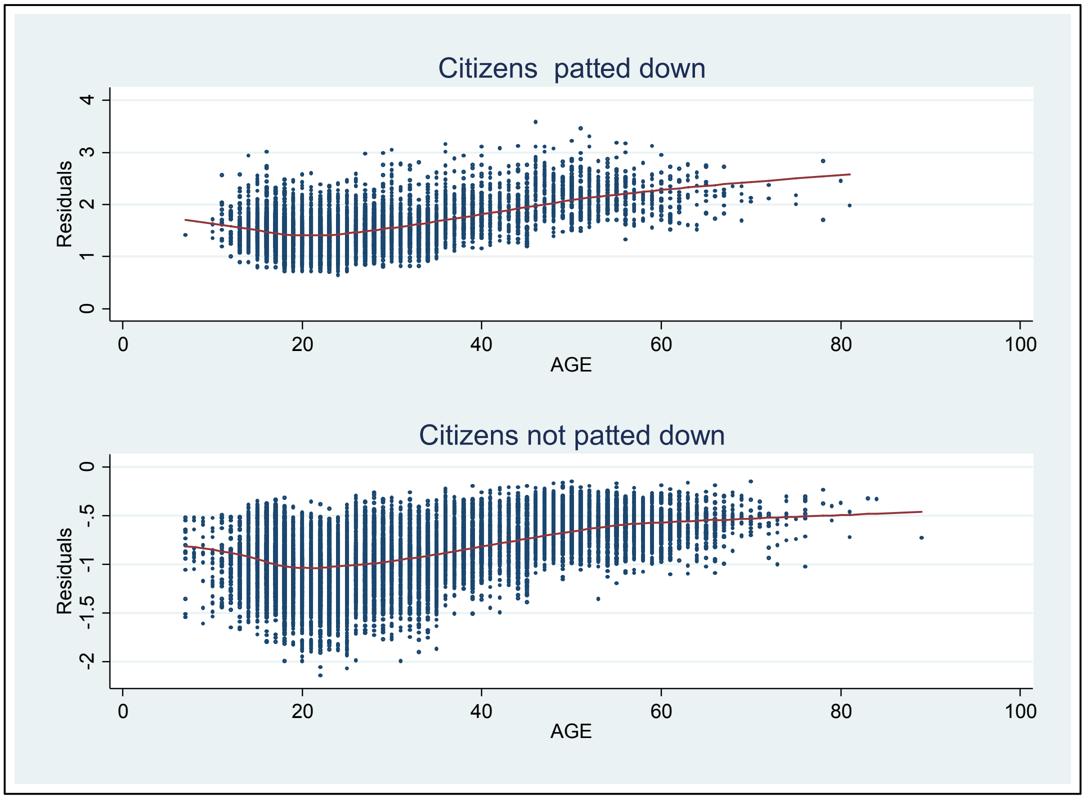
For the first sample, the relationship between pat down residuals and civilian age was examined, separately for each outcome group (Figure 10). If residuals are well patterned there should be no relationship between scores on the predictor and residual scores. That does not appear to be the case here. In both groups average residuals appear somewhat dependent on age.

This analysis uses continuous age, whereas the model used categorical age. Despite this limitation, a pattern seen suggests that one or more unobserved covariates linked with age were perhaps affecting the outcome.

10.1.1.2.2 Conclusion

These diagnostics, considered in total, argue against a causal interpretation of the impacts of race and ethnicity from these regression models. Some of the diagnostics suggest that selection on unobserved covariates may be affecting the outcome. The safest conclusion at this juncture is that civilian race and ethnicity correlate with the outcome examined, but race or ethnicity *per se* may not be playing causal roles. Rather or ethnicity linked factors, factors not modeled here, cannot be ruled out.

Figure 10 First random sample: Relationship between stopped civilian age and pat down residual



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

10.1.2.1 Steps

Separately for each sample, all the covariates used in the regression model, save race or ethnicity, were used to predict the stopped civilian being Black non-Hispanic versus White non-Hispanic (Hispanics excluded from these models). As with the regression models, these also were mixed effects logit models with random effects for districts. An initial model first confirmed that the race or ethnicity variable being predicted varied significantly across districts.

Following the prediction of race or ethnicity using observed covariates and district context, the predicted score on the race or ethnicity outcome was saved. [12] This predicted score was treated as the propensity score in the matching program.[13] Caliper matching within 1/10th of a standard deviation was specified. Models were run again specifying an even tighter caliper match,

[12] The Stata option mu was used here; this incorporated both the fixed and random effects in the prediction model.
[13] psmatch2 in Stata.

within .07 of a standard deviation on the propensity scores. (Models for later outcomes used even tighter caliper matching requirements, .06 of a standard deviation.)

10.1.2.2 Results

Results from propensity score caliper matched models appear in Table 30. These analyses use only matched pairs of Black and White non-Hispanic stopped civilians. In each pair, the propensity scores of members of the pair are the most closely matching propensity scores of the non-matched cases remaining.

In the first random sample using the caliper match of $1/10^{th}$ of a standard deviation, 1,875 Black stopped civilians were matched with 2,087 stopped White civilians (total = 3,962). If the caliper match is tightened to within .07 hundredths of the standard deviation, the corresponding numbers are 1,873 matched Black civilians and 2,087 matched White civilians (total = 3,960). The numbers of White and Black civilians are not exactly equal because multiple civilians might have exactly the same score on the propensity-to-be-black variable.

Impacts of the contrast between stopped Black non-Hispanic and stopped White non-Hispanic civilians appear for both random samples and for caliper matches within $1/10^{th}$ of a standard deviation on the propensity score and again within 7 hundredths of a standard deviation on the propensity scores. In all analyses non-matched cases are dropped. The model shown, as recommended (Guo & Fraser, 2015: 384), are mixed effects models with random effects for districts.

Regardless of which sampled is examined, and regardless of the restrictions set on caliper matching, in all instances stopped Black non-Hispanic civilians appear more likely to be patted down then matched stopped White non-Hispanic civilians. Black civilians' odds of being [patted down versus not patted down] were anywhere from 19 to 31 percent higher depending on the caliper match specified in the sample. The statistical significance associated with the race variable ranged from $p < .05$ to $p < .001$ depending on the caliper match and the sample. These results confirm a net association, seen in the mixed effects logit models, between race and the likelihood of receiving a pat down.

That said, diagnostics suggested this link should be interpreted as correlational only and not causal. Details appear in the next sections.

Table 30 Propensity score model estimates of impact of Black vs. White civilian race on pat down outcome (non-Hispanics only)

	B	SE	Z	P =	LCL	UCL	OR	OR-LCL	OR-UCL
Caliper = .10									
Sample 1									
Black non-Hispanic	0.277***	0.0743	3.722	0.0002	0.131	0.422	1.319***	1.140	1.526
Constant	-1.106						0.331		
District variance (se)	0.200	0.076							
Observations	3,962								
Number of groups	22								
LR chi square test	83.8 (p < .001)								
Sample 2									
Black non-Hispanic	0.177*	0.0744	2.380	0.0173	0.0312	0.323	1.194*	1.032	1.381
Constant	-1.045						0.352		
	0.193	0.071							
Observations	4,034								
Number of groups	22								
LR chi square test	91.66 (p < .001)								
Caliper = .07									
Sample 1									
Sample 1									
Black non-Hispanic	0.218**	0.0747	2.915	0.0036	0.0713	0.364	1.243**	1.074	1.439
Constant	-1.129						0.324		
District variance (se)	0.135	0.055							
Observations	3,960								
Number of groups	22								
LR chi square test	57.14 (p < .001)								
Sample 2									
Black non-Hispanic	0.231**	0.0737	3.136	0.00171	0.0867	0.375	1.260**	1.091	1.456
Constant	-1.109						0.33		
District variance (se)	0.170	0.066							
Observations	4,026								
Number of groups	22								
LR chi square test	80.03 (p < .001)								

Note. January – June 2016 CPD ISR stop data. 95 percent confidence intervals shown.
* = p < .05; ** = p < .01; *** = p < .001

10.1.2.3 Diagnostics: Covariate balancing and observed selection

Austin (2009) recommends balance diagnostics between the treatment (black) and control (white) groups that examine each covariate, and consider mean differences as well as variance differences. If these balance diagnostics fail, then observed selection cannot be ruled out. This means that even after matching, Black and White stopped civilians still differ on these other factors, and these other factors could be simultaneously affecting both race and the outcome. There is an observed selection problem.

For both samples, regardless of the caliper matching level used, balance diagnostics failed (results not shown). In all instances, there were multiple mean differences on covariates, and the treatment to control variance ratios were outside acceptable limits.

Rubin (2001: 177) has suggested some overall balance statistics that simultaneously take all covariates into account. Rubin's *B* is "the standardized difference in the means of the propensity scores" between the two groups being compared. The suggested limits are within a half a standard deviation (p. 174), which translates in the program used to a value lower than 25. Rubin's *R is* "the ratio of the variances of the propensity scores" (p. 177) of the two groups. The suggested limits are between .75 and 1.25 (p. 177).

Rubin's summary statistics appear in Table 31. These show that Black and White non-Hispanic civilians were *not* sufficiently balanced on covariates after matching, as shown by the Rubin's B values above 25, in sample 1. The two groups, on this summary measure, do appear sufficiently balanced on covariates in sample 2. Ratios of variances (Rubin's *R*) appear acceptable in both samples at both matching levels.

Table 31 Summary covariate balancing statistics after matching Black and White respondents for propensity score model of pat downs

		Rubin's *B*	Rubin's *R*
Caliper match = .10			
	Sample 1	26.6*	1.11
	Sample 2	21.8	1.07
Caliper match = .07			
	Sample 1	25.1*	1.04
	Sample 2	20.7	1.06

10.1.2.4 Diagnostics: Sensitivity to unobserved selection

The results of the sensitivity analysis to gauge the impacts unobserved selection might have on the race impact appear in Table 32. In three out of the four scenarios, if two individuals who are similar on the observed covariates differ in their odds of being Black and non-Hispanic versus White and non-Hispanic by only about 15 percent, then there is 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)

Table 32 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 = .10		
	Sample 1	1.25
	Sample 2	1.15
Caliper match = .07		
	Sample 1	1.15
	Sample 2	1.15

10.1.2.5 Limitations and conclusion

Propensity score matching models can run afoul of a wide variety of problems (Guo & Fraser, 2015: 381-386). The resulting propensity score matching models here should be considered preliminary until additional analyses using a different matching protocol such as the Mahalanobis nearest neighbor approach can be completed. That said, results seen here are robust in the following ways: they replicate across two independent random samples, and they replicate using different caliper matching restrictions. But bear in mind, as noted below, the covariates are **not** balanced.

Excluding all Hispanic stopped civilians, results showed that stopped Black civilians experience significantly higher chances of being patted down compared to matched White civilians. Blacks' odds were anywhere from 19 to 31 percent higher for being [patted down versus not patted down].

Two features of diagnostics suggest, however, that this link should be interpreted as correlational and not causal. Diagnostics suggest that observed selection bias – that is differences between Blacks and matched Whites on the covariates used – was a problem in one of the samples if we just look at summary statistics on covariate balancing, and in both of the samples if we look at the covariate-by-covariate results. Although matching dramatically reduces differences between Blacks and Whites on the observed covariates, troubling discrepancies remained.

Sensitivity analyses also suggested that selection bias on unobserved factors was potentially problematic. If civilians similarly situated on the covariates differ in their odds of being [black and non-Hispanic versus White and non-Hispanic] by as little as 15 percent, then the significant impact of race on the outcome would probably disappear.

10.1.3 Caliper matched propensity score models: Hispanic vs. White non-Hispanic civilians

10.1.3.1 Steps

The procedures paralleled what was done to learn about different outcomes between White non-Hispanic civilians versus Black non-Hispanic civilians. Whereas that analysis dropped all stopped Hispanics, this analysis dropped all Black non-Hispanic stopped civilians, leaving the focus on contrasting Hispanics versus White non-Hispanics who were stopped.

As with the Black versus White contrast, models were done with two levels of caliper matching. The first level of matching created a match if the non-Hispanic stopped civilian was within 1/10th of a standard deviation of the corresponding Hispanic member of the pair on the propensity score. The second level of matching was slightly tighter here, using .06 of a standard deviation on the propensity score rather than .07 as was done with a Black versus White contrast.

10.1.3.2 Results

Propensity score model results for impacts of ethnicity appear in Table 33. Stopped Hispanic civilians, compared to matched White non-Hispanic stopped civilians, had odds of [being patted down versus not patted down] that were anywhere from 35 to 45 percent higher, depending on the sample and the caliper restriction. All of these differences were highly significant statistically ($p < .001$).

To help better understand the results, we use the sample 1 predicted probabilities from the model requiring a caliper match of .06 of a standard deviation or better on the propensity score. These predicted probabilities show that whereas a matched stopped White non-Hispanic civilian in a typical district had predicted chances of being patted down that were about 23.9 percent, the corresponding predicted chances for a Hispanic stopped civilian of being patted down were 31.3 percent.

10.1.3.3 Diagnostics

Diagnostics on whether the covariates are balanced between the Hispanic and the White non-Hispanic groups suggest that observed selection was not a problem (Table 34). Both summary balancing statistics were within acceptable ranges. Although the ratios of the variances contrasting the two groups on individual covariates routinely seem quite different, in most cases means on covariates were not significantly different.

Sensitivity analyses indicated, however, that selection on unobserved covariates probably cannot be dismissed as an important potential confound. Gamma (Γ) values as little as 1.15 rendered the ethnicity impact on the outcome nonsignificant. The interpretation of this ethnicity impact probably should remain correlational rather than causal.

Table 33 Propensity score model estimates of impact of Hispanic vs. White civilian ethnicity on pat down outcome (black non-Hispanics excluded))

	B	SE	Z	p	LCL	UCL	OR	OR-UCL	OR-LCL
Caliper = .10									
Sample 1									
Hispanic	0.300***	0.0743	4.036	< .0001	0.154	0.446	1.350***	1.167	1.562
Constant	-1.157						0.314		
District variance (se)	0.0831	0.0385							
Observations	3,879								
Number of groups	22								
LR chi square test	33.28 (p < .001)								
Sample 2									
Hispanic	0.317***	0.0744	4.262	< .0001	0.171	0.463	1.373***	1.187	1.589
Constant	-1.116						0.328		
District variance (se)	0.122	0.0582							
Observations	3,881								
Number of groups	22								
LR chi square test	42.19 (p < .001)								
Caliper = .06									
Sample 1									
Hispanic	0.371***	0.0740	5.014	< .0001	0.226	0.516	1.449***	1.253	1.675
Constant	-1.156						0.315		
District variance (se)	0.0901	0.0401							
Observations	3,853								
Number of groups	22								
LR chi square test	39.41 (p < .001)								
Sample 2									
Hispanic	0.330***	0.0745	4.424	< .0001	0.184	0.476	1.390***	1.202	1.609
Constant	-1.107***						0.33		
District variance (se)	0.134	0.0628							
Observations	3,859								
Number of groups	22								
LR chi square test	49.08 (p < .001)								

Note. January – June 2016 CPD ISR stop data. 95 percent confidence intervals shown.
*** = p < .001

Table 34 Summary covariate balancing statistics after matching Hispanic and White non-Hispanic respondents for propensity score model of pat downs

		Rubin's B	Rubin's R
Caliper match = .10			
	Sample 1	14.7	1.08
	Sample 2	20.8	1.08
Caliper match = .06			
	Sample 1	15.8	1.12
	Sample 2	17.2	0.99

Table 35 Sensitivity analysis, propensity score models, pat down outcome, Hispanic vs. White non-Hispanic civilians

		Gamma (Γ) value where ethnicity impact becomes non-significant
Caliper match = .10		
	Sample 1	1.15
	Sample 2	1.15
Caliper match = .06		
	Sample 1	1.2
	Sample 2	1.15

10.2 Did the pat down result in a weapon/firearm being discovered?

Officers sometimes recovered weapons including firearms when they conducted pat downs of stopped civilians. Table 36 reports the number of pat downs and the number of recovered weapons for each of the two random samples of data. In each random sample around 8,900 pat downs took place resulting in roughly 240 recovered weapons. If a recovered weapon and only a recovered weapon counts as a hit, then the hit rate for pat downs in each sample was quite close to 2.5 percent.

This outcome depends on a prior officer action during the stop. Recovering a weapon or failing to recover a weapon requires that the officers initiate a pat down. Therefore, this outcome depends on an officer selection process. Statistical modeling must consider that process.

Put another way, and as described above in discussing sequential selection, race or ethnicity or gender can matter twice once a stop is underway. With this particular outcome, race or ethnicity can affect the likelihood of the civilian being selected for a pat down after controlling for other covariates. In addition, race or ethnicity can affect the likelihood that the pat down leads to a recovered weapon. Modeling seeks to separately estimate these two race/ethnicity post-stop dynamics.

Regrettably, theories of officer behavior initiating stops and officer behavior during stops provide no clear theoretical guidance on which civilian, stop, or context features, i.e., covariates, are associated with which specific set of dynamics. Therefore, the models used here make some untested assumptions which will be explained as we go along.

In the first set of analyses, multiple regression mixed effects logit models were run that included an additional predictor intended to take into account dynamics leading to a stopped civilian being selected for a pat down. That additional predictor was the predicted probability from the mixed effects logit models that a stop would result in a pat down. See 10.1.1.1.

Table 36 Frequency of pat downs resulting in firearm/weapon recovered: Samples 1 and 2

Sample 1			
	Pat down?		
	No (0)	Yes (1)	Total
Weapon found?			
No (0)	0	8,942	8,942
Yes (1)	0	229	229
(not applicable)	17,887	0	17,887
Total	17,887	9,171	27,058
Sample 2			
	Pat down?		
	No (0)	Yes (1)	Total
Weapon found?			
No (0)	0	8,957	8,957
Yes (1)	0	236	236
(not applicable)	17,865	0	17,865
Total	17,865	9,193	27,058

Source: January-June 2016 ISR data from CPD.

Following those models, a different type of model formulated specifically to address the selection issue, was run. [14] This is the Heckman selection model (Heckman, 1979) for a binary outcome (Baum, 2006).

For the multiple logistic regression models, an initial mixed effects null model with each sample confirmed a **lack** of significant district-to-district variation in this outcome. A second single level logit model with all the dummy variables for districts, save the Loop, also revealed no significant

[14] This is heckprobit in Stata.

impacts of any district on the weapon recovered outcome. Therefore all the models run were single level models logit and did not include dummy variables for districts. Whether this lack of geographic variation on this pat down “hit” outcome was a function merely of the low base rate of weapon recovery, or something else, is not clear.

Although these are single level models they took clustering within district into account by allowing for clustered errors at the district level using the Huber-White sandwich estimator (White, 1982), despite some recent concerns about these adjustments (Freedman, 2006).

10.2.1 Multiple logistic regression models with predicted probabilities of a pat down

10.2.1.1 Results

Multiple logistic regression models for sample 1 appear in Table 37, and results for sample 2 appear in Table 38. Each table shows the results for three models.

- Model A included just race and ethnicity.
- Model B added in the predicted probability that the stop would result in a pat down. This predicted probability was intended to control for the pat down selection process.
- Model C then added in all the other covariates used in previous models. Tables report both the coefficient (B) and the associated odds ratio (OR) for each predictor.

Hosmer-Lemeshow (2000) tests with both 10 groups and 50 groups generated nonsignificant results, suggesting some degree of overall fit (results not shown).

In short, for both samples Model C with all factors entered yielded a significant impact of race, in the expected negative direction, on the likelihood of a weapon being recovered as a result of the pat down ($p < .05$). In the first sample patted down Black non-Hispanic civilians as compared to patted down White civilians had odds of [the pat down producing a weapon versus not producing a weapon] that were about 31 percent lower (1-.689). The corresponding figure in the second sample was odds that were about 38 percent lower (1-.617). Both of these demonstrated a negative net impact of race on the likelihood of recovering the weapon from a pat down.

Table 37 Multiple logistic regression models of pat down weapon recovered: Sample 1

		Model A		Model B		Model C	
	Reporting:	B	OR	B	OR	B	OR
Predictors	Variable name						
Black civilian	dblack	-0.439	0.645	-0.246	0.782	-0.372*	0.689*
Hispanic civilian	dhisp	-0.179	0.836	-0.00136	0.999	-0.0523	0.949
Predicted: pat down	pre_dpat_1			-1.562***	0.210***	-0.698	0.498
Male	dmale					0.768**	2.156**
Age 18-25	age1825					-0.0172	0.983
Age 26-35	age2635					0.126	1.134
Age 36-45	age3645					0.525**	1.690**
Age 46 and up	age46pl					0.607**	1.835**
February	dfeb					0.613**	1.846**
March	dmar					0.0250	1.025
April	dapr					0.457*	1.580*
May	dmay					0.634**	1.885**
June	djun					0.549*	1.732*
Weekend	wknddum					0.00404	1.004
3 - 6 AM	dhr0306					-0.431	0.650
6 - 9 AM	dhr0609					-0.257	0.773
9 - 12 AM	dhr0912					0.0565	1.058
12 - PM	dhr1215					0.0516	1.053
3 - 6 PM	dhr1518					0.438	1.550
6 - 9 PM	dhr1821					0.108	1.114
9 - 12 PM	dhr2123					0.335	1.398
Vehicle stop	dvehstop					-0.309	0.734
Missing event no.	eventmis					0.371	1.449
Constant		-3.315***	0.0363***	-2.871***	0.0566***	-4.527***	0.0108***
N		9,171		9,171		9,171	

*** p<0.001, ** p<0.01, * p<0.05

Source: January-June 2016 ISR data from CPD.

Table 38 Multiple logistic regression models of pat down weapon recovered: Sample 2

		Model A		Model B		Model C	
	Reporting:	B	OR	B	OR	B	OR
Predictors	Variable name						
Black civilian	dblack	-0.727**	0.484**	-0.518*	0.596*	-0.660*	0.517*
Hispanic civilian	dhisp	-0.501*	0.606*	-0.33	0.719	-0.326	0.722
Predicted: pat down	pre_dpat_2			-1.758**	0.172**	0.0148	1.015
Male	dmale					-0.151	0.86
Age 18-25	age1825					0.105	1.11
Age 26-35	age2635					0.305	1.357
Age 36-45	age3645					0.984***	2.674***
Age 46 and up	age46pl					1.172***	3.228***
February	dfeb					0.46	1.584
March	dmar					0.605	1.831
April	dapr					0.887***	2.427***
May	dmay					0.784***	2.190***
June	djun					0.726**	2.066**
Weekend	wknddum					0.343	1.409
3 - 6 AM	dhr0306					1.086**	2.961**
6 - 9 AM	dhr0609					0.137	1.147
9 - 12 AM	dhr0912					-0.0782	0.925
12 - PM	dhr1215					0.427	1.533
3 - 6 PM	dhr1518					0.504	1.655
6 - 9 PM	dhr1821					0.444*	1.558*
9 - 12 PM	dhr2123					0.297	1.345
Vehicle stop	dvehstop					-0.59	0.554
Missing event no.	eventmis					-	-
Constant		-3.015***	0.0490***	-2.508***	0.0814***	-4.360***	0.0128***
N		9,193		9,193		9,109	

*** p<0.001, ** p<0.01, * p<0.05
Source: January-June 2016 ISR data from CPD.

The predicted probabilities that a stop would result in a recovered weapon, presented separately depending on race, ethnicity and gender, appear in Table 39. In both samples, predicted recovery rates for stopped Black non-Hispanic civilians were in the two and a half percent range, in contrast to predicted probabilities in the three and a half percent to four and a half percent range, depending on the sample, for stopped White non-Hispanic civilians. Gender discrepancies in predicted weapon recovery rates via pat down within race/ethnicity groups proved sample dependent. For each of the three race/ethnicity groups predicted recovery rates were lower for women than men in sample 1. But in sample 2, pat downs of women linked to higher predicted probabilities of recovery for Black non-Hispanic and Hispanic women. The differences described here are descriptive only, and not statistically significant. [15]

[15] Models adding a gender x race interaction resulted in a BIC value that was almost equal to the BIC value of the model with no interaction (results not shown).

Table 39 Predicted probabilities of weapon recovery as a result of a pat down: Single level logit model

Sample		Racial/ethnic Group			Total
1					
		White NH	Black NH	Hispanic	
	Female	0.0202	0.0122	0.0151	0.0136
	Male	0.0371	0.0236	0.0304	0.0258
	Total	0.0351	0.0229	0.0295	0.025
2					
	Female	0.0426	0.0263	0.0335	0.0296
	Male	0.0482	0.0232	0.0289	0.0257
	Total	0.0474	0.0234	0.0291	0.0259

Note NH= non-Hispanic. Results from single level logit model with covariates and predicted probability that a pat down would take place.
Source: Jan-June 2016 ISR reports from CPD.

10.2.1.2 Diagnostics

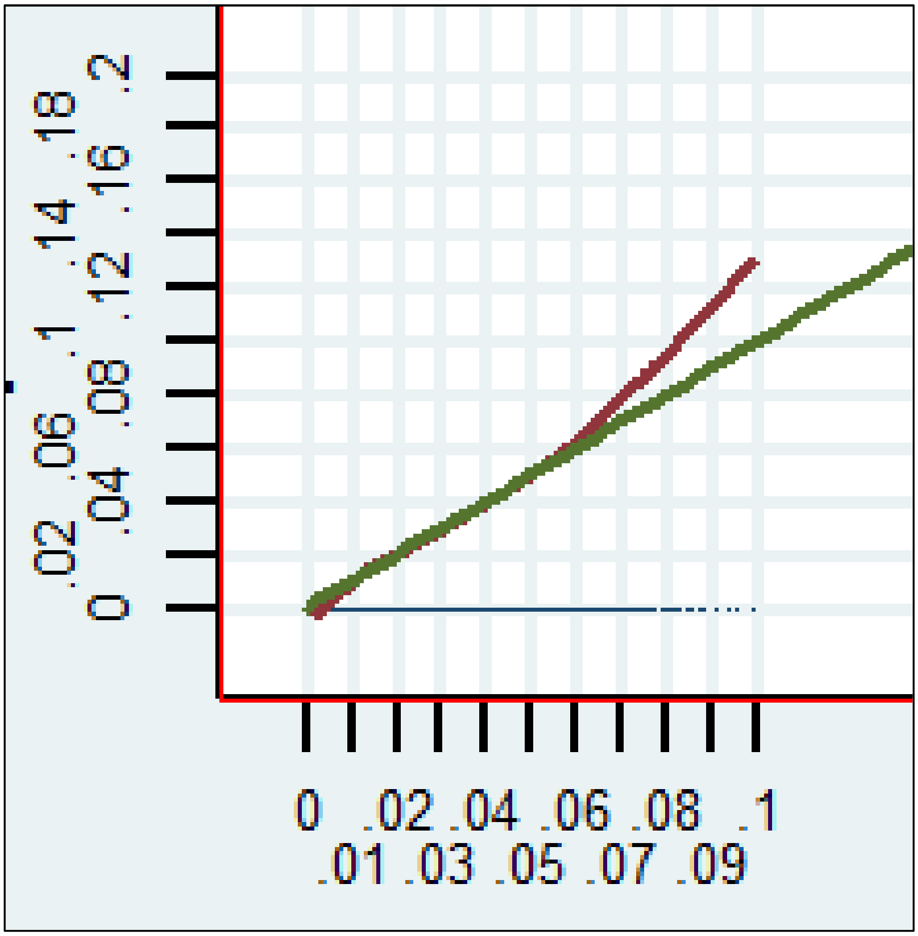
Despite the acceptable Hosmer-Lemeshow fit statistics, other diagnostics suggested some potential concerns about these models. For sample 1, Figure 11 plots predicted probabilities against the moving average of the proportion of pat downs yielding a weapon.

- The green [straight] line = the "moving average of the proportion of cases that equal one [weapon discovered];"
- The red [curved] line = the "fraction of observed cases that equal 1 [weapon discovered] at each level of the model's predicted probability of observing a 1 [weapon discovered]" (Long & Freese, 2006: 156-157).

The red line uses local LOWESS smoothing. The X axis stops at .10 because that was the maximum predicted probability. Figure 12 shows the same information for sample 2.

For sample 1, results suggested the "model fail[ed] in predicting" the higher probabilities of a weapon being discovered where "the fraction of observed cases exceed[ed] the predicted probabilities" (Long & Freese, 2006: 157). This is because at values of predicted probabilities greater than about .06, the predicted red line began to diverge upward from the green line. This divergence suggested a lack of fit in this range of predicted probabilities.

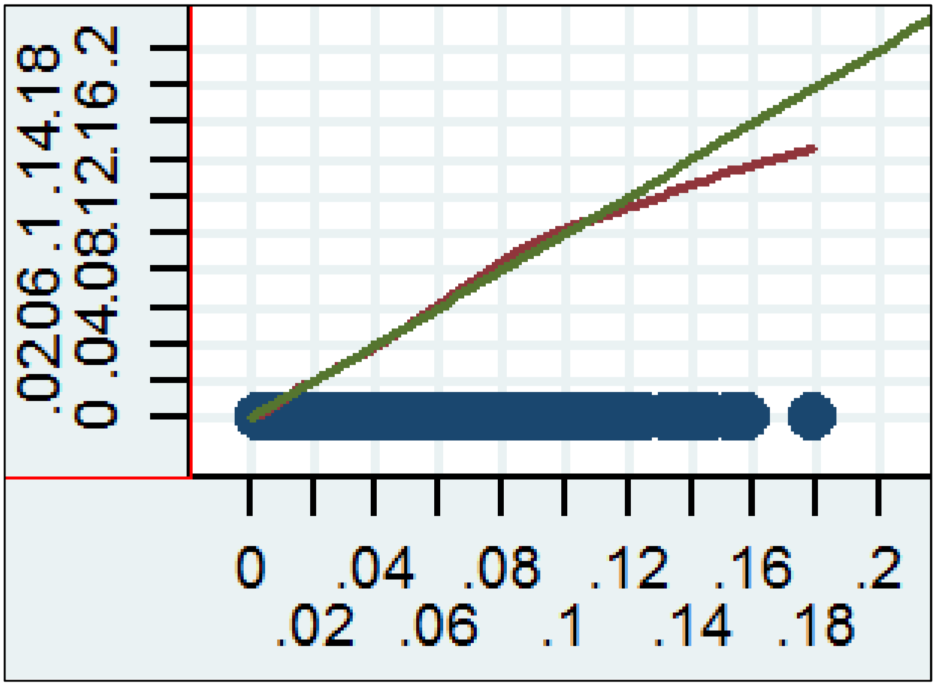
Figure 11 pat down results in a weapon, sample 1: LOWESS smoothed curve of observed (Y axis) and predicted (X axis) outcome



Note. Sample 1. Y (vertical) axis = observed outcome; X (horizontal) axis = predicted outcome. Red line shows local relationship between these two features. Divergence from straight green line suggests lack of fit.

In sample 2, shown in Figure 12, predicted probabilities ranged roughly twice as far as they did in sample 1. Here, predicted probabilities went up to almost .18. Again, as in sample 1, the diagnostics suggested a lack of fit at higher predicted probability values. The divergence here, however, went in the opposite direction. Starting at predicted probabilities of around .14 and going to higher values, the fraction of observed cases was lower than would be predicted from the model results. For example, at a predicted probability of weapons recovery of around .18, the observed proportion of weapons recovered was only .14.

Figure 12 pat down results in a weapon, sample 1: LOWESS smoothed curve of observed (Y axis) and predicted (X axis) outcome



Note. Sample 2. Y (vertical) axis = observed outcome; X (horizontal) axis = predicted outcome. Red line shows local relationship between these two features. Divergence from straight green line suggests lack of fit.

The suggested lack of fit between data and model at higher predicted probabilities, however, should be contextualized. It was in these observed and predicted probability ranges that observations were, in relative terms, somewhat sparse.

At this juncture, further diagnostics to determine the source of the model lack of fit in each sample were not undertaken given time constraints. All that can be said at this point is that there appears to be potential concern that the model, for reasons not yet known, appears to get off-track at higher predicted probabilities, and the way it gets off track depends on the specific random sample. Future diagnostics could more closely examine case level diagnostics such as influence and leverage.

10.2.2 Heckman probit selection models

10.2.2.1 Results

Attention turns now to models specifically designed to incorporate selection dynamics. These models worked as follows. They simultaneously estimated two independent equations: factors affecting the initial outcome, whether or not a pat down occurred; and factors affecting whether a weapon was recovered from a pat down. Each equation simultaneously took into account the other, including the degree of relatedness between the different dynamics represented by the two equations.

Because these analytics were specifically designed for this particular type of selection problem, they provided a more stringent test. Compared to the foregoing multiple logistic regression models, the results of this type of model will present a more "conservative" estimate of the impacts of race or ethnicity on weapon recovery following a pat down.

Although these are single level models they took clustering within district into account by allowing for clustered errors at the district level using the Huber-White sandwich estimator (White, 1982), despite some recent concerns about these adjustments (Freedman, 2006).

In these models the full set of predictors used in the multiple logistic regressions just reported, minus the predicted probability variable, predicted whether a pat down took place. Model A employed three variables to predict weapons recovered: whether the civilian was non-Hispanic Black, whether the civilian was Hispanic, and whether the civilian was male. Model B added age variables to the above to predict weapons recovery. Two different sets of models with different predictors of weapons recovery reflected the theoretical uncertainty mentioned earlier about the factors relevant to different elements of the post stop process. Table 40 presents results from sample 1 and Table 41 for sample 2. Although these tables present the selection equations, results will discuss just the outcome equation, that is, the determinants of whether a weapon was recovered during a pat down. Select features from the outcome equations appear in Table 42. Predicted probabilities of weapons recovered appear in Table 43.

Table 40. Sample 1: Determinants of pat down recovery of weapon controlling for pat down selection (heckprobit model)

SAMPLE 1	MODEL A								MODEL B							
	Outcome equation: Weapon recovered				Selection equation: pat down occurs				Outcome equation: Weapon recovered				Selection equation: pat down occurs			
variable name	b	se	t	OR	b	se	t	OR	b	se	t	OR	b	se	t	OR
dblack	-0.0835	(0.0592)	-1.410	0.920	0.282***	(0.0724)	3.901	1.326	-0.0965	(0.0650)	-1.484	0.908	0.282***	(0.0724)	3.904	1.326
dhisp	0.0177	(0.0703)	0.252	1.018	0.190**	(0.0665)	2.864	1.210	0.0121	(0.0698)	0.174	1.012	0.191**	(0.0664)	2.868	1.210
dmale	0.465***	(0.0925)	5.028	1.592	0.647***	(0.0344)	18.84	1.910	0.439***	(0.0892)	4.926	1.551	0.647***	(0.0344)	18.82	1.910
age1825					0.125*	(0.0524)	2.395	1.134	-0.0026	(0.0806)	-0.0329	0.997	0.125*	(0.0525)	2.389	1.134
age2635					-0.0320	(0.0419)	-0.762	0.969	0.0212	(0.0946)	0.224	1.021	-0.0315	(0.0413)	-0.764	0.969
age3645					-0.385***	(0.0562)	-6.861	0.680	0.113	(0.114)	0.990	1.119	-0.383***	(0.0572)	-6.685	0.682
age46pl					-0.690***	(0.0589)	-11.72	0.502	0.0584	(0.131)	0.446	1.060	-0.690***	(0.0595)	-11.60	0.502
dfeb					0.000114	(0.0346)	0.00331	1.000					0.00105	(0.0347)	0.0302	1.001
dmar					0.00982	(0.0495)	0.198	1.010					0.00988	(0.0495)	0.200	1.010
dapr					-0.0872	(0.0586)	-1.487	0.917					-0.0866	(0.0590)	-1.469	0.917
dmay					-0.160*	(0.0678)	-2.364	0.852					-0.159*	(0.0679)	-2.348	0.853
djun					-0.240***	(0.0660)	-3.634	0.787					-0.239***	(0.0660)	-3.625	0.787
wknddum					0.0779**	(0.0242)	3.224	1.081					0.0780**	(0.0242)	3.215	1.081
dhr0306					0.133	(0.101)	1.317	1.142					0.133	(0.101)	1.315	1.142
dhr0609					-0.604***	(0.150)	-4.028	0.547					-0.606***	(0.151)	-4.002	0.546
dhr0912					-0.436***	(0.0652)	-6.687	0.647					-0.437***	(0.0654)	-6.677	0.646
dhr1215					-0.323***	(0.0663)	-4.862	0.724					-0.323***	(0.0664)	-4.867	0.724
dhr1518					-0.194***	(0.0532)	-3.641	0.824					-0.193***	(0.0531)	-3.631	0.824
dhr1821					-0.219***	(0.0527)	-4.153	0.803					-0.219***	(0.0529)	-4.137	0.803
dhr2123					-0.173***	(0.0414)	-4.174	0.841					-0.172***	(0.0414)	-4.165	0.842
dvehstop					0.248***	(0.0533)	4.664	1.282					0.249***	(0.0533)	4.665	1.282
eventmis					0.187	(0.275)	0.682	1.206					0.189	(0.275)	0.686	1.208
Constant	-2.712			0.0664	-0.845***			0.430	-2.658***			0.0701	-0.846***			0.429
athrho	0.630***	(0.124)	5.096						0.485	(0.250)	1.94					
rho	0.558								0.4505							
Wald test		chi squared (df=1) = 25.97; p < .001							chi squared (df=1) = 3.76; p = .052							
BIC	34,350								34,359							
Observations	27,058															

Robust standard errors in parentheses. *** p<0.001, ** p<0.01, * p<0.05

Table 41 Sample 2: Determinants of pat down recovery of weapon controlling for pat down selection (heckprobit model)

SAMPLE 2	MODEL A								MODEL B							
	Outcome equation: Weapon recovered				Selection equation: pat down occurs				Outcome equation: Weapon recovered				Selection equation: pat down occurs			
variable name	b	se	t	OR	b	se	t	OR	b	se	t	OR	b	se	t	OR
dblack	-0.147	(0.0892)	-1.653	0.863	0.280***	(0.074)	3.812	1.323	-0.223	(0.117)	-1.902	0.8	0.280***	(0.073)	3.839	1.324
dhisp	-0.0678	(0.104)	-0.653	0.934	0.163**	(0.056)	2.9	1.177	-0.099	(0.122)	-0.815	0.905	0.164**	(0.056)	2.926	1.178
dmale	0.231*	(0.0961)	2.398	1.259	0.660***	(0.034)	19.57	1.934	0.0647	(0.190)	0.341	1.067	0.659***	(0.034)	19.58	1.933
age1825					0.0954**	(0.034)	2.787	1.1	0.0431	(0.109)	0.395	1.044	0.0962**	(0.034)	2.87	1.101
age2635					-0.0373	(0.039)	-0.947	0.963	0.1	(0.087)	1.154	1.105	-0.0357	(0.039)	-0.926	0.965
age3645					-0.314***	(0.044)	-7.1	0.73	0.338**	(0.107)	3.155	1.402	-0.304***	(0.044)	-6.94	0.738
age46pl					-0.702***	(0.042)	-16.84	0.496	0.354*	(0.154)	2.294	1.424	-0.698***	(0.042)	-16.61	0.497
dfeb					0.0183	(0.036)	0.511	1.018					0.0219	(0.037)	0.599	1.022
dmar					0.0298	(0.056)	0.53	1.03					0.0346	(0.057)	0.611	1.035
dapr					-0.0699	(0.060)	-1.159	0.932					-0.0623	(0.062)	-1.003	0.94
dmay					-0.138	(0.073)	-1.897	0.871					-0.133	(0.075)	-1.778	0.876
djun					-0.196**	(0.071)	-2.751	0.822					-0.193**	(0.073)	-2.656	0.824
wknddum					0.0595*	(0.024)	2.45	1.061					0.0624**	(0.024)	2.628	1.064
dhr0306					0.218*	(0.086)	2.54	1.244					0.233*	(0.092)	2.528	1.262
dhr0609					-0.533***	(0.142)	-3.748	0.587					-0.539***	(0.142)	-3.794	0.584
dhr0912					-0.449***	(0.055)	-8.178	0.638					-0.455***	(0.055)	-8.312	0.634
dhr1215					-0.298***	(0.051)	-5.84	0.743					-0.299***	(0.052)	-5.753	0.742
dhr1518					-0.150*	(0.069)	-2.195	0.86					-0.148*	(0.069)	-2.15	0.863
dhr1821					-0.209***	(0.051)	-4.082	0.811					-0.208***	(0.052)	-4.016	0.812
dhr2123					-0.117***	(0.032)	-3.65	0.889					-0.117***	(0.032)	-3.629	0.89
dvehstop					0.293***	(0.057)	5.132	1.34					0.292***	(0.057)	5.12	1.34
eventmis					0.174	(0.159)	1.092	1.189					0.166	(0.162)	1.022	1.18
Constant	-2.447			0.0866	-0.886			0.412	-2.13			0.119	-0.893***			0.409
athrho	0.825*	0.383	2.154						0.21	0.298	0.7					
rho	0.678	0.207							0.2065							
Wald test	chi squared (df=1) = 4.64; p < .05								chi squared (df=1)=.49							
BIC	34,523								34,503							
Observations	27,058								27,058							

Robust standard errors in parentheses; *** p<0.001, ** p<0.01, * p<0.05

Table 42 Select model features from equations predicting weapons recovery from pat downs

		Sample	1		2	
		Model	A	B	A	B
	Variable	Name				
	Black	dblack	0.159	0.132	0.099	0.057
Significance level of impacts of:	Hispanic	dhisp	0.801	0.862	0.514	0.415
	Male	dmale	0.000001	0.000001	0.016	0.733
Best fitting/most parsimonious?			Yes			Yes
Significant link between selection and outcome?			Yes	No	Yes	No

Note. Two tailed significance levels associated with t statistics for race, ethnicity and gender from Heckman probit selection models. Model A included just race, ethnicity and gender in the weapons recovery equation (reference group = White non-Hispanic females). Model B also included age categories (reference group = White non-Hispanic females younger than 18) in the same equation. The best fitting/most parsimonious selection was based on sizable differences in BIC values. Source: Jan-June 2016 ISR data from CPD

Three out of the four models showed male pat downs significantly more likely to lead to a recovered weapon compared to female pat downs (Models A & B, sample 1; Model A, sample 2). For example, looking at the predicted probabilities for sample 1 (Table 43), the average predicted probability for males was in the 1.1-1.2 percent range whereas the corresponding figure for females was in the 0.3-0.4 percent range.

In both samples, both models, ethnicity was not associated with the probabilities of weapons recovery from a pat down. Table 42 shows the probabilities associated with the Hispanic variable; all of these values are highly *non*significant. Hispanics and non-Hispanics clearly did not differ, according to these models.

Race results were a closer call. In sample 1, the two tailed statistical significance probabilities associated with being Black were in the .13 to .16 range. In sample 2, they ranged from .06 - .10. Using a standard two-tailed test of statistical significance, as has been done throughout, these were only marginally significant impacts of race.

10.2.2.2 Diagnostics

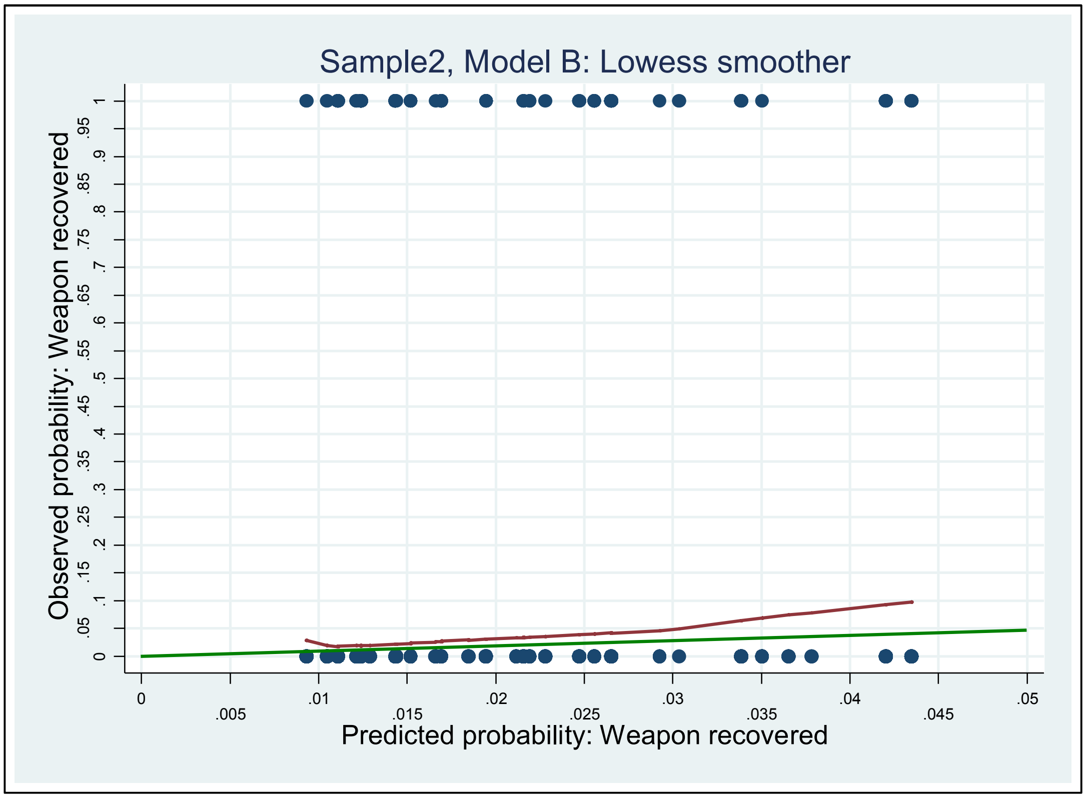
Using the model that came closest to yielding a significant race impact on weapon recovery following a pat down, the smoothed LOWESS curve capturing the relationship between predicted probabilities and observed probabilities is the red line that appears in Figure 13.

Table 43 Predicted probabilities that pat down leads to weapons recovery, by race/ethnicity/gender

Sample	Model		Racial/ethnic Group			Total
1	A					
			White NH	Black NH	Hispanic	
		Female	0.0033	0.0026	0.0035	0.0029
		Male	0.0123	0.0099	0.0129	0.0107
		Total	0.0104	0.009	0.0117	0.0096
	B					
			White NH	Black NH	Hispanic	Total
		Female	0.0044	0.0032	0.0043	0.0036
		Male	0.0145	0.0111	0.0144	0.012
		Total	0.0123	0.0101	0.0131	0.0109
2	A					
		Female	0.0072	0.0047	0.006	0.0053
		Male	0.0133	0.0091	0.0112	0.0098
		Total	0.012	0.0085	0.0105	0.0092
	B					
		Female	0.0254	0.0144	0.0158	0.0161
		Male	0.0304	0.0167	0.0201	0.0184
		Total	0.0293	0.0164	0.0195	0.0181

Note. Predicted probabilities from Heckman probit models. Predicted probabilities generated using the default pmargin which means these represent the success (weapon found) probability
Source: Jan-June 2016 ISR data from CPD.

Figure 13 Predicted and observed probabilities of weapon recovery following pat down



Starting at predicted probability values of .035 and higher, the predicted probabilities begin to demonstrate a lack of fit. The predicted probabilities are substantially higher than the observed probabilities in that range. This lack of fit, however, is happening in a range of predicted probabilities where there are relatively few predicted scores (see Figure 14).

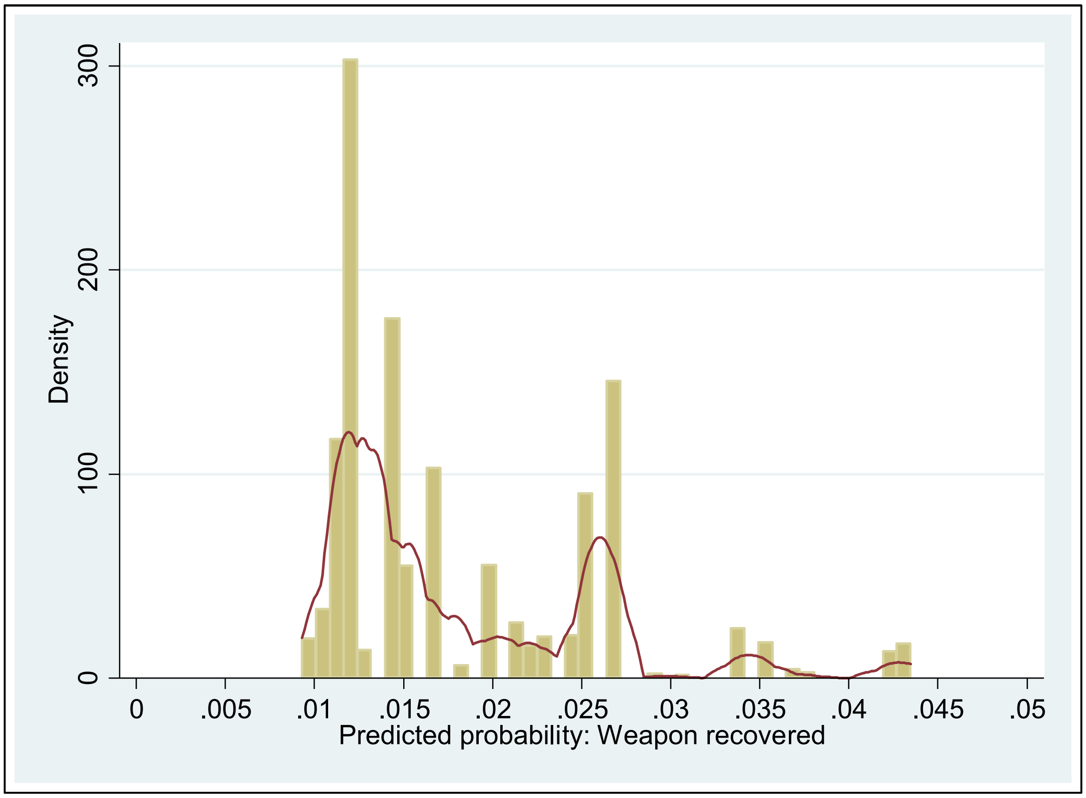
10.2.3 Conclusions on weapon recovery from pat downs

The clearest civilian correlate of whether a pat down yields a recovered weapon was gender. In the regression model results from one sample, and the selection model B results from both samples, male pat downs proved substantially more likely to yield weapons.

The second clearest civilian correlate was age. In Model B regression results from both samples, pat downs of civilians 36 and older proved more likely to reveal weapons. In the Heckman selection models, in sample 2 Model B results, but not in sample 1 Model B results, older civilians pat downs similarly were more likely to yield weapons. The one caution with this surprising finding that older stopped civilians who were patted down proved more likely to be armed was the failure to replicate the significant age impacts across **both** random samples in the Heckman selection models.

In sample 2, the *opposite-to-expected* age effects provided by the Heckman selection model underscored the importance of separating being patted down from a weapon being recovered. Older stopped civilians were significantly *less* likely to be patted down. But, at least in sample 2, if they were patted down they were *more* likely to have a weapon.

Figure 14 Distribution of predicted weapon recovery probabilities: Histogram and kernel density estimates



Note. Results from Model B, sample 2, Heckman probit selection model. Y axis is *density* not frequency. Source: Jan-June 2016 CPD ISR data.

The overall pattern for ethnicity link clear-cut. In the regression results ethnicity showed no significant connection with pat down weapon recovery. The Heckman selection models demonstrated the same lack of impact. But, again, this variable emphasized the importance of separating pat down and recovery dynamics. In the Heckman selection models, the null impacts of ethnicity on weapon recovery occurred in the context of a significant positive impact of ethnicity on being selected for a pat down in the first place. So two different ethnicity impacts appear to be operating once a stop is underway.

The race results prove hardest to summarize because here the results from the regression models and the Heckman selection models diverged most markedly. Regression models for both samples showed significantly lower pat down weapon yield for Black non-Hispanic as contrasted to White non-Hispanic civilians, controlling for the predicted probability that a pat down would take place.

In contrast, Heckman selection models failed to produce significant negative impacts of race on weapon recovery, although in Models A and B in sample 2 the race link had marginal significance. Underscoring yet again the importance of separating pat down and weapon recovery dynamics, the non significant or only marginally significant impacts of race on weapon recovery occurred in the context of stopped Black non-Hispanic civilians being significantly ($p < .001$) more likely to be selected for a pat down in the first place.

In short, race links to significantly lower hit rates if we follow the regression model results, but does not do so if we follow the Heckman selection model results, even though the latter come close to showing a significant race impact on pat down weapon recoveries.

If one argues that the hypothesis tests of race impacts on yield should be one tailed rather than two tailed, that is, the only possible interpretable outcome is that Black pat downs compared to non-black pat downs will be less likely to surface weapons, then the takeaway points are different. Results from both models for sample 2 using the Heckman selection model *do* yield a significant race impact: Black non-Hispanic searches failed to yield as many weapons, proportionally, as White non-Hispanic searches. But, this result *failed* to replicate across both samples, which is a remaining cause of concern.

All of these takeaway points must be contextualized in light of model diagnostics, which proved mixed as well. The regression models revealed acceptable overall fit using standard Hosmer-Lemeshow statistics. But plots of predicted and observed scores suggested divergences indicative of *lack* of model fit at the high end of the predicted probabilities of weapon recovery. Examining the Heckman selection model that came closest to producing a significant race impact on weapon recovery similarly suggested a lack of fit at higher predicted probability levels. In both cases, however, the divergence happened in a range where there were relatively few data points.

10.3 Was a Search Conducted?

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

10.3.1 Exclusion question

***All analyses of the search outcome were conducted on records after removing those stops where an arrest took place*.**

This dramatically reduced the volume of searches examined by roughly two thirds. See Table 45. In sample 1, 4,788 searches were reduced to 1,315. In sample 2, 4,807 searches were reduced to 1,325.

The reason for dropping these searches was this. We know that some fraction of these searches were *incident* to taking the arrestee into custody. Removing these is appropriate. The officer did not decide whether or not to search, rather he/she just followed department procedures in these cases. But there are 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. 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, and this will be addressed in future work.

10.3.2 Search links to other enforcement outcomes

Aside from searches linked to arrests, how often did searches link to other enforcement outcomes? Results, combining both random samples for ease of presentation, appear in Table 44. Slightly over one third of the non arrest linked searches, 37 percent, occurred in stops where another type of enforcement action took place (n=976). But almost two times as many non-arrest-linked searches took place in stops where no enforcement action was recorded (n=1,644). This latter group made up slightly less than two thirds of the non-arrest searches (63 percent).

Table 44. Among non-arrest linked searches: N and proportion linked or not linked with other enforcement actions

N searches linked to non-arrest enforcement actions	
All non-arrest linked searches	2,640
N non-arrest searches linked to other specific enforcement actions	
ANOV	403
PSC	352
Other	221
Sum: N searches linked to non arrest enforcement actions	976
Proportion of non-arrest searches	0.37
N searches linked to no enforcement action	1,664
Proportion of non-arrest searches	0.63

Source: Jan-June 2016 CPD ISR data. Both samples combined.

10.3.3 Mixed effects regression models

10.3.3.1 Results

Initial mixed effects logit models confirmed significant ($p < .001$) variation across districts in the probability that a stop would involve a search (results not shown).

Results from the model with all covariates entered appear in Table 46. Neither civilian race nor civilian ethnicity significantly affected the odds that a search would be conducted. This was true for both samples. So for this outcome there appeared to be no net effects of either race or ethnicity.

Gender, however, did elevate the chances that officers would search civilians. But this impact appeared only for sample 2. Males' odds of [being searched versus not searched] were about 40 percent higher in that sample.

Table 45 Numbers of stops, with and without searches, before and after removing custodial searches incident to arrest

Sample				N	Percent
1					
	Initial				
		Search	No	22,270	82.3
			Yes	4,788	17.7
		Total		27,058	
	After removing custodial searches incident to arrest				
		Search	No	22,270	94.42
			Yes	1,315	5.58
		Total		23,585	
2					
	Initial				
		Search	No	22,251	82.23
			Yes	4,807	17.77
		Total		27,058	
	After removing custodial searches incident to arrest				
		Search	No	22,251	94.38
			Yes	1,325	5.62
		Total		23,576	

Source: Jan-Jun 2016 ISR data from CPD

Turning to other covariates, age mattered. In both samples stopped civilians between the ages of 18 and 35 were more likely to be searched than those below the age of 18. In addition, when the stop took place proved relevant in both samples. Compared to the reference time frame between midnight and 3 AM, searches were less likely, in both samples, between 6 AM and 9 PM. Finally, in both samples vehicle as compared to pedestrian stops had a much higher likelihood of resulting in a search. The expected odds of a search taking place were the least 200 percent higher in both samples if it was a vehicle rather than pedestrian stop.

Table 46 Predicting search occurrence: Mixed effects logit model

		Sample 1		Sample 2	
Fixed effects		b	OR	b	OR
Black civilian	dblack	0.0924	1.097	0.109	1.115
Hispanic civilian	dhisp	0.0977	1.103	0.108	1.114
Male	dmale	0.164	1.178	0.343***	1.409***
Age 18-25	age1825	0.462***	1.587***	0.573***	1.774***
Age 26-35	age2635	0.409***	1.506***	0.604***	1.829***
Age 36-45	age3645	0.112	1.119	0.261*	1.298*
Age 46 and up	age46pl	-0.395**	0.674**	-0.213	0.808
February	dfeb	-0.445***	0.641***	-0.155	0.857
March	dmar	-0.193	0.825	-0.00366	0.996
April	dapr	-0.236*	0.790*	0.0849	1.089
May	dmay	-0.461***	0.631***	-0.163	0.85
June	djun	-0.399***	0.671***	-0.301**	0.740**
Weekend	wknddum	-0.0412	0.960	0.0912	1.096
3 - 6 AM	dhr0306	-0.0300	0.970	0.0532	1.055
6 - 9 AM	dhr0609	-1.326***	0.266***	-0.791***	0.453***
9 - 12 AM	dhr0912	-0.524***	0.592***	-0.281*	0.755*
12 - PM	dhr1215	-0.554***	0.575***	-0.578***	0.561***
3 - 6 PM	dhr1518	-0.403***	0.668***	-0.313*	0.731*
6 - 9 PM	dhr1821	-0.408***	0.665***	-0.259*	0.772*
9 - 12 PM	dhr2123	-0.274**	0.760**	-0.107	0.898
Vehicle stop	dvehstop	1.127***	3.087***	1.329***	3.777***
Missing event no.	eventmis	-0.333	0.717	-0.936*	0.392*
Constant		-2.739	0.0646	-3.409	0.0331
Random effects	District variance	0.0452*		0.0906**	
	Observations	23,585		23,576	
	Number of groups	22		22	
	BIC	9,874		9,841	

*** p<0.001, ** p<0.01, * p<0.05

Note. Jan.-June 2016 ISR data, CPD

For sample 1: Null model: LR $\chi 2$ test vs. logistic model (df = 1) = 45.25; p < .001; BIC = 10,122

For sample 2: Null model: LR $\chi 2$ test vs. logistic model (df = 1) = 80.85; p < .001; BIC = 10,142

Source: Jan-Jun 2016 ISR data from CPD

10.3.3.2 Diagnostics
Diagnostics generally suggested just a few concerns with these models. The LOWESS smoothed curves showing the relationship between predicted probabilities and observed proportions indicated relatively close fit of the predictions at all ranges of predicted probabilities for both samples (results not shown). Plots of residuals against predicted probabilities showed no relationship in both samples (results not shown). Only two features were potentially concerning. As seen earlier in the models predicting whether a pat down took place, here too residuals appeared correlated with age, trending lower for younger age civilians (results not shown). Finally, normal probability plots showed a higher density than expected of residuals in the second quartile for cases where a search took place (results not shown).

In contrast to the results with the pat down outcome, geographic residual variation for the search outcome showed no significant discrepancies across districts (results not shown). For each district, the confidence interval around its average residual always included the overall average residual (-.16).

In sum, these diagnostics suggested a low to moderate level of concern about observed and unobserved selection.

10.3.4 Propensity score models: Black vs. White non-Hispanics only
As with the pat down outcome, propensity score models with caliper matching on the propensity score were conducted. Here, only one level of caliper matching, within .06 of a standard deviation, was examined.

10.3.4.1 Results
Table 47 shows impacts of race on whether a search was conducted, using a propensity score matching model. The model for sample 1 included 1,622 Black non-Hispanic civilians and 1,824 matched White non-Hispanic civilians. The model for sample 2 included 1,646 Black non-Hispanic stopped civilians and 1,831 matched White non-Hispanic civilians. Preliminary models confirmed that the chances of a stopped civilian being Black varied significantly across districts in each sample, and that the outcome varied significantly across districts in both samples when considering just the propensity matched cases (results not shown).

The table tells a simple story. No significant predicted differences on the chances of being searched appeared when contrasting matched White and Black non-Hispanic stopped civilians. For members of both groups, chances of being searched, in both samples, were right around 4.7 percent.

10.3.4.2 Diagnostics
Results do not appear susceptible to selection on observed covariates based on summary measures. In both samples values for Rubin's *B* and Rubin's *R* were well within the acceptable range. In sample 1, B = 16.4, R = 1.00. In sample 2, B = 15.5 and R = 1.04.

Individual covariates, however, did suggest some slight causes for concern. There were both mean differences and variance differences. Looking at individual covariate mean differences after matching in each sample showed one significant difference (proportion of weekend stops in sample 1, proportion of males stopped in sample 2). Other covariates were mean balanced. But the ratios of treated (black) vs. control (white) variances were all outside the acceptable range in both samples.

All of this suggests a small to moderate level of concern about selection on observed covariates.

Table 47 Propensity score model estimates of impact of Black vs. White civilians on search outcome

	B	SE	Z	p =	LCL	UCL	OR	OR-LCL	OR-UCL
Propensity score model race impact, Caliper=.06									
Sample 1									
Black non-Hispanic	-0.00437	0.160	-0.0273	0.978	-0.318	0.309	0.996	0.728	1.362
Constant	-3						0.0498		
District variance (se)	0.152	0.109							
Observations	3,446								
Number of groups	22								
LR chi square test	5.88 (p < .01)								
Sample 2									
Black non-Hispanic	0.00198	0.154	0.0129	0.99	-0.299	0.303	1.002	0.741	1.355
Constant	-3.017						0.0489		
District variance (se)	0.256*	0.127							
Observations	3,477								
Number of groups	22								
LR chi square test	26.39 (p < .001)								
* = p < .05									

Note. Impact of Black vs. White non-Hispanic stopped civilians on search outcome from propensity score model using Caliper matching within .06 of a standard deviation on the propensity score. Non-matched cases dropped. B = coefficient; SE = standard error; Z = Z test; OR = odds ratio. LCL and UCL = respectively, lower and upper bounds of 95 percent confidence interval
Source: Jan-June 2016 ISRs from CPD.
* = p < .05

10.3.5 Propensity score models: White non-Hispanic vs. Hispanic only

For sample 1, propensity score matching within .06 of a standard deviation of propensity meant that the analysis focused on 1,520 Hispanics and 1,824 matched White non-Hispanics. For sample 2, the corresponding numbers were 1,831 White non-Hispanics in 1,494 Hispanics. Black civilians were excluded from this analysis.

Preliminary analyses confirmed for both samples that the probabilities of the non-black civilian being Hispanic varied significantly across districts (results not shown). They also confirmed that when analyzing just the matched cases, significant variation on the search outcome across districts persisted in both samples (results not shown).

Table 48 Propensity score model estimates of impact of Hispanic vs. White non-Hispanic civilians on search outcome

Propensity score model ethnicity impact, Caliper = .06	B	SE	Z	p =	LCL	UCL	OR	OR-LCL	OR-UCL

Sample 1									
Hispanic	0.335*	0.151	2.225	0.0261	0.0399	0.630	1.398*	1.041	1.878
Constant	-2.978						0.0509		
District variance (se)	0.0298	0.0443							
LR chi squared test	0.73, ns								
Observations	3,344								
Number of groups	22								
Sample 2									
Hispanic	0.334*	0.147	2.272	0.0231	0.0458	0.621	1.396*	1.047	1.862
Constant	-2.972						0.0512		
District variance (se)	0.153	0.089							
LR chi squared test	14.42; p < .001								
Observations	3,325								
Number of groups	22								

Note. Impact of Hispanic vs. White non-Hispanic stopped civilians on search outcome from propensity score model using Caliper matching within .06 of a standard deviation on the propensity score. Non-matched cases dropped. B = coefficient; SE = standard error; Z = Z test; OR = odds ratio. LCL and UCL = respectively, lower and upper bounds of 95 percent confidence interval
Source: Jan-June 2016 ISRs from CPD.
* = p < .05

10.3.5.1 Results

Results from caliper matched propensity score models with just matching cases appear in Table 48. Results from both samples indicated that Hispanic stopped civilians had about 40 percent higher predicted odds of [being searched versus not searched] compared to the predicted odds of matched non-Hispanic White stopped civilians.

In sample 1, White non-Hispanic civilians' predicted probability of being searched was .048. The corresponding figure for Hispanic civilians in the sample was .066. In sample 2 the predicted probability for White non-Hispanics was .049 and the corresponding predicted probability for Hispanic civilians of being searched was .067. In both samples these results were statistically significant ($p < .05$).

In short, both samples suggested that Hispanic stopped civilians' chances of being searched were significantly higher than the chances of matched White non-Hispanic civilians.

10.3.5.2 Diagnostics

Summary measures of covariate balancing after matching suggested selection on observed covariates was not a concern. In sample 1 Rubin's B = 18.8 and Rubin's R = 1.08. In sample 2 the corresponding numbers were 18.4 and 0.98. Examining individual covariates suggested a bit more concern about this matter. For both samples, after matching, there was at least one covariate where a significant mean difference remained. Further, the ratio comparing variances of the White cases and Hispanic cases were for each covariate outside the suggested boundaries.

Sensitivity analyses suggested extreme sensitivity to selection on unobserved covariates. In both samples, minor changes in the odds of differential "assignment" to ethnicity due to unobserved factors (Γ =1.05) resulted in the observed significant ethnicity impact disappearing.

10.3.6 Summing up on search outcome and race and ethnicity

The points to take away about the link between race and ethnicity of stopped civilians and whether or not they were searched include the following.

Race is not relevant. Neither the regression models nor the propensity score matching models revealed significant net differences between Black and White non-Hispanic stopped civilians on this outcome.

Ethnicity may be relevant. Although the regression models for both samples failed to find a significant ethnicity impact, the matching propensity score models for both samples did find one. That link however is probably best interpreted as correlational and not causal for the following reason. The propensity score matching models appear extremely sensitive to selection on unobserved factors.

10.4 DID A SEARCH RESULT IN A WEAPON BEING DISCOVERED?

After removing custodial searches incident to arrest, and considering only cases where officers also checked the search box, only an extremely low number of searches resulted in weapons being discovered. Given those extremely low numbers, and the large number of covariates involved in the models used here, this outcome was not analyzed.

In sample 1, only 10 searches produced a weapon or a firearm or both after removing searches incident to arrest. In sample 2, the number was 14 after the removal.

10.5 DID THE OFFICER ENGAGE IN ENFORCEMENT?

10.5.1 Regression results

Results from both samples reveal significant net effects of both race and ethnicity on the likelihood that any type of enforcement action would be delivered during the stop. In both samples, Black non-Hispanic stopped civilians had at least 20 percent greater odds of [receiving any enforcement action versus no enforcement action] compared to White non-Hispanic stopped civilians. The discrepancy between Hispanic and non-Hispanic White civilians was about the same magnitude. See Table 49.

Table 49 Predicting any enforcement action: Mixed effects logit models

		Sample 1		Sample 2	
Fixed effects		b	OR	b	OR
Black civilian	dblack	0.256***	1.291***	0.213***	1.237***
Hispanic civilian	dhisp	0.239***	1.270***	0.218***	1.243***
Male	dmale	-0.105**	0.900**	-0.011	0.989
Age 18-25	age1825	-0.441***	0.643***	-0.395***	0.674***
Age 26-35	age2635	-0.201***	0.818***	-0.141**	0.868**
Age 36-45	age3645	-0.126*	0.882*	-0.0588	0.943
Age 46 and up	age46pl	0.119**	1.127**	0.170***	1.185***
February	dfeb	-0.420***	0.657***	-0.384***	0.681***
March	dmar	-0.444***	0.641***	-0.436***	0.646***
April	dapr	-0.545***	0.580***	-0.472***	0.624***
May	dmay	-0.612***	0.542***	-0.555***	0.574***
June	djun	-0.614***	0.541***	-0.635***	0.530***
Weekend	wknddum	-0.0318	0.969	0.018	1.018
3 - 6 AM	dhr0306	0.291**	1.338**	0.216*	1.242*
6 - 9 AM	dhr0609	-0.243**	0.784**	-0.0862	0.917
9 - 12 AM	dhr0912	-0.189**	0.828**	-0.118*	0.889*
12 - PM	dhr1215	-0.413***	0.661***	-0.381***	0.683***
3 - 6 PM	dhr1518	-0.484***	0.616***	-0.404***	0.668***
6 - 9 PM	dhr1821	-0.352***	0.703***	-0.355***	0.701***
9 - 12 PM	dhr2123	-0.239***	0.787***	-0.156**	0.856**
Vehicle stop	dvehstop	0.489***	1.630***	0.583***	1.792***
Missing event no.	eventmis	-2.887***	0.0558***	-1.729***	0.177***
Constant		-0.00504	0.995	-0.203*	0.816*
Random effects					
	District variance	0.0532**		0.0648**	
BIC		33,052		33,001	
Observations		27,058		27,058	
Number of groups		22		22	

Note *** p<0.001, ** p<0.01, * p<0.05

Note. January – June 2016 ISR data, CPD

Note. For sample 1: Null model: LR χ2 test vs. logistic model = 238.63; p < .001; BIC = 33,816

For sample 2: Null model: LR χ2 test vs. logistic model =266.12 ; p < .001; BIC = 33,728

Predicted probabilities appear in Table 50 for both samples, shown separately by race/ethnicity/gender combination. It shows, for example, in sample 1 that among stopped Black non-Hispanic males the predicted probability of receiving any enforcement action was 32.6 percent, compared to a predicted probability of 27.8 percent for White non-Hispanic males.

Table 50. Predicted probabilities, any enforcement action

Sample 1
Gender

	White NH	Black NH	Hispanic	Total
Female	0.296	0.367	0.35	0.355
Male	0.278	0.326	0.302	0.318
Total	0.282	0.331	0.308	0.323

Sample 2
Gender

	White NH	Black NH	Hispanic	Total
Female	0.283	0.348	0.321	0.334
Male	0.284	0.327	0.305	0.319
Total	0.284	0.33	0.307	0.321

Note. NH= non-Hispanic. Source: Jan-June 2016 ISRs, CPD

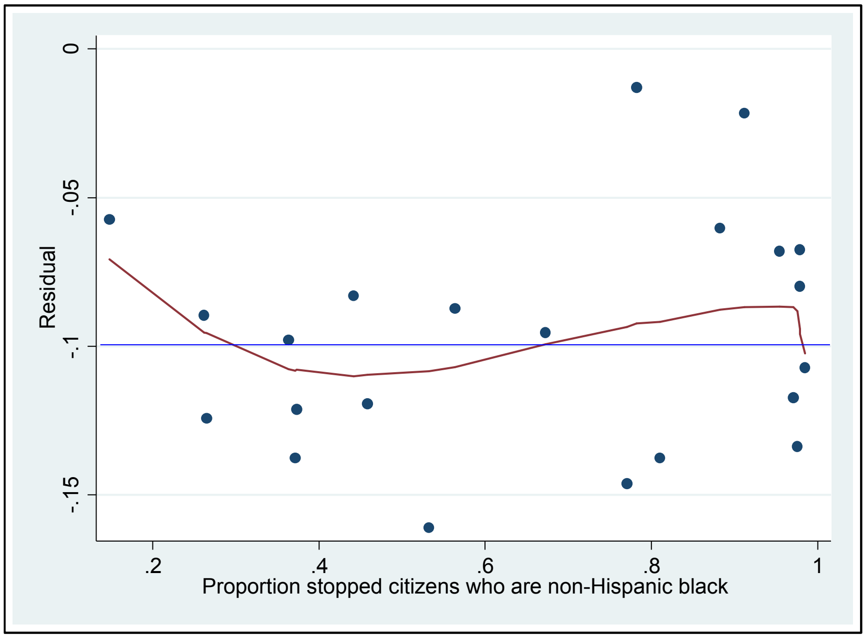
10.5.2 Diagnostics

Diagnostics revealed the following. In both samples predicted probabilities deviated only modestly from observed proportions receiving any enforcement action, at different predicted probabilities (results not shown). The district level residuals' 95 percent confidence intervals in all cases overlapped with the overall residual.

But, that said, two features of residuals suggested a low to moderate level of concern about potential bias due to unobserved selection. In both samples, a modest relationship between residuals and predicted scores surfaced, with residuals increasing slightly as predicted scores increased (results not shown). Further, in both samples a modest district-level relationship surfaced between the proportions of civilians stopped who were Black and non-Hispanic, and the standardized model residuals (Figure 15, Figure 16).

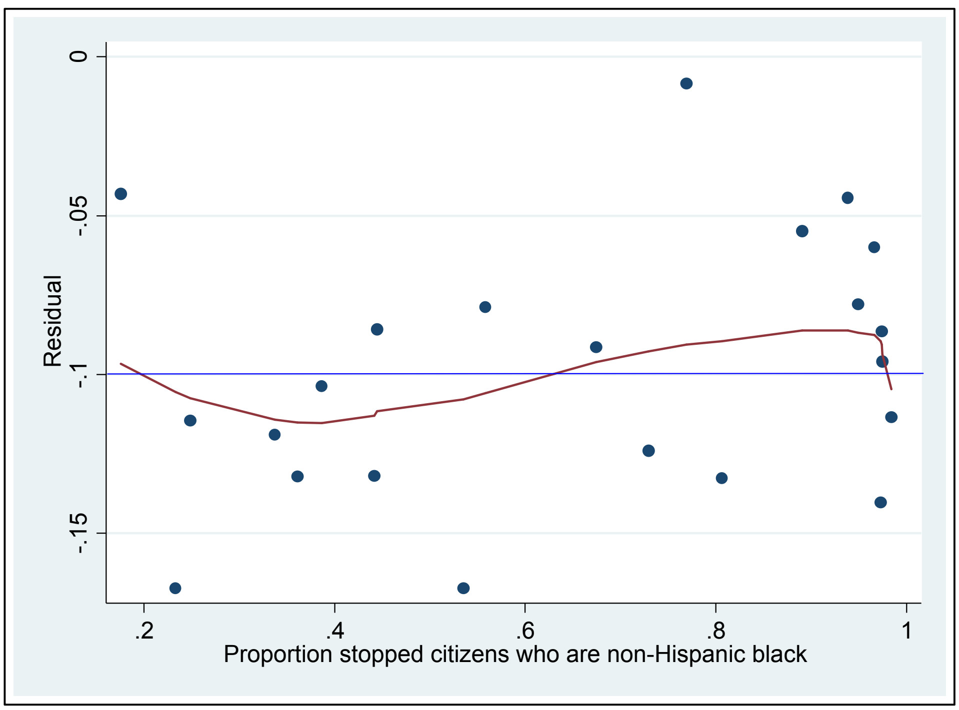
Interpreting this pattern requires a bit of background on residuals in logit models. The standardized Anscombe residuals for sample 1 appear in Figure 17.

Figure 15 Residual, any enforcement action, and proportions stopped Black civilians: Sample 1



Note. Horizontal reference line reflects average overall residual. Data shown are district level. Curved line = LOWESS smoothed curve. Source: Jan-June 2016 ISRs, CPD.

Figure 16 Residual, any enforcement action, and proportions stopped Black civilians: Sample 2

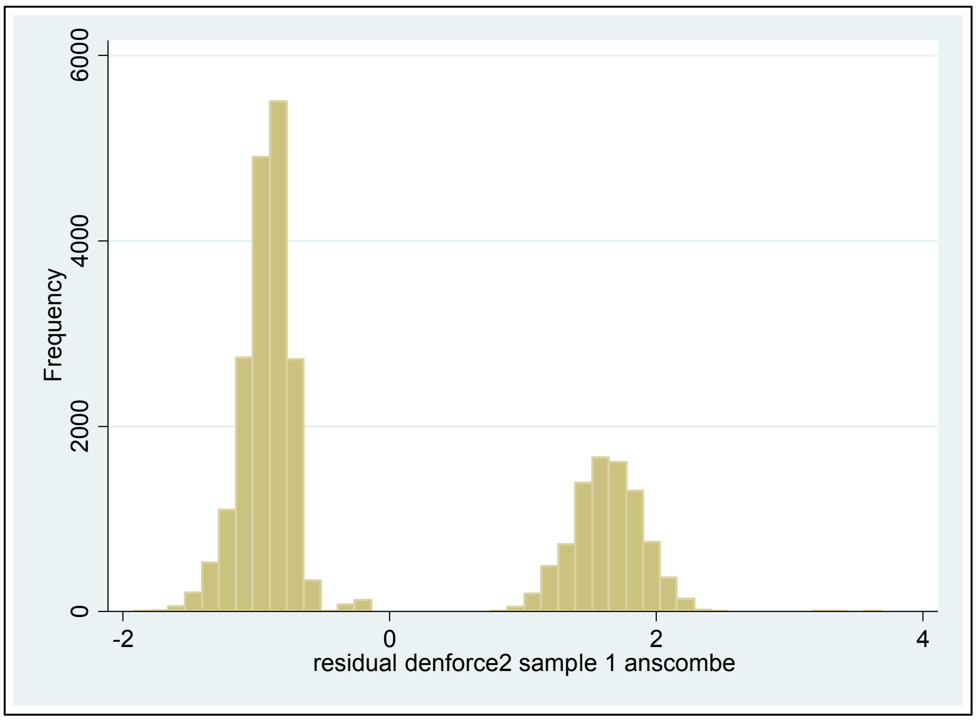


Note. Horizontal reference line reflects average overall residual. Data shown are district level. Curved line = LOWESS smoothed curve. Source: Jan-June 2016 ISRs, CPD.

A positive residual is associated with a stop where any enforcement action occurred. A negative residual is associated with a stop where an enforcement action did not occur. Standardized residual values within +/- 2 or +/- 3 are not considered outliers. There are some positive outliers indicating stops where enforcement happened despite an extremely low predicted probability that that would happen. Therefore, the point suggested by Figure 15, and Figure 16, is that in districts where a large fraction of stopped civilians were Black, there was a larger mix of stops in that district where enforcement took place despite low predicted probabilities.

Again, this patterning of residuals shifts across different predicted probabilities is only a modest pattern. But it is noticeable in both samples. That's why a low-to-moderate degree of concern about unobserved selection is suggested with these regression models.

Figure 17 Any enforcement action residuals: Sample 1



Note. Jan-June 2016 ISRs from CPD

10.5.3 Propensity selection model – Black non-Hispanic vs. White non-Hispanic

10.5.3.1 Results

Propensity score models with matched cases within .060 of a standard deviation on the propensity score were run. Considering only matched cases, the significant race effect failed to resurface (Table 51). Black stopped civilians compared to White stopped civilians had only slightly higher odds of [receiving any enforcement action versus receiving none].

In the first random sample, Black civilians' odds were about 11 percent higher; they were about 7 percent higher in the second random sample. In neither sample, however, were these differences between Blacks and Whites statistically significant.

10.5.3.2 Diagnostics

Diagnostics suggested a low to medium level of concern about potential confounding due to observed selection. Looking at the overall balance diagnostics and sample 1, Rubin's B was above the suggested cutoff value (B=26.6) suggesting a lack of balance on covariates. The overall balance statistics were within an acceptable range for sample 2 (Rubin's B = 22.1; Rubin's R = 1.06).

Looking at the individual covariate diagnostics in sample 1 showed there were a couple of covariates, such as being male and the stop taking place on the weekend, that remained significantly different between the two racial groups even after matching (sample 1). In sample 2 after matching there remained significant differences between the two groups for the stop taking place on the weekend and the stopped civilian being between the ages of 26 and 35.

10.5.4 Propensity selection model – Hispanic vs. White non-Hispanic

10.5.4.1 Results

In contrast to the race results, the ethnicity differences seen in the regression model resurfaced in the propensity score matching models (Table 52), and for both random samples. In both, Hispanics odds of being [receiving any enforcement action versus none] were about 20 percent higher than the odds for matched White non-Hispanics. The size of the ethnicity impact seen in the propensity models, expressed as an odds ratio, was closely comparable to the size of the effect seen for ethnicity in the multiple regression models. Controlling for other factors, and for district context, stopped Hispanics were more significantly likely to be on the receiving end of an enforcement action than were stopped White non-Hispanic civilians.

10.5.4.2 Diagnostics

Overall balance diagnostics suggested that selection on observed covariates was only of low concern (B = 15.8; R = 1.12). Individual variable diagnostics suggested somewhat more concern. There were mean differences on a couple of covariates even after matching, and variance ratios between Hispanic/non-Hispanic White groups continued to be quite dissimilar even after matching.

Table 51 Matched propensity score model results predicting any enforcement action: Black vs. White non-Hispanic

	B	SE	Z	p=	LCL	UCL	OR	OR-LCL	OR-UCL
Sample 1									
Black non-Hispanic	0.102	0.0710	1.436	0.151	-0.0372	0.241	1.107	0.963	1.273
Constant	-0.897						0.408		
District variance (se)	0.0541*	0.0270							
Observations	3,957								
Number of groups	22								
LR chi squared test; 17.51; p < .001									
Sample 2									
Black non-Hispanic	0.0665	0.0703	0.946	0.344	-0.0713	0.204	1.069	0.931	1.227
Constant	-0.871						0.418		
District variance (se)	0.0623	0.0331							
Observations	4,025								
Number of groups	22								
LR chi squared test; 17.09; p < .001									

Note. *** p<0.001, ** p<0.01, * p<0.05

Note. January – June 2016 ISR data, CPD

Table 52 Matched propensity score model results predicting any enforcement outcome: Hispanic vs. White non-Hispanic

	B	SE	Z	p=	LCL	UCL	OR	OR-LCL	OR-UCL
Sample 1									
Hispanic	0.189**	0.0719	2.634	0.0084	0.0485	0.330	1.208**	1.050	1.391
Constant	-0.948						0.387		
District variance (se)	0.0818*	0.0370							
Observations	3,853								
Number of groups	22								
LR chi squared test (df=1)	34.22***								
Sample 2									
Hispanic	0.183*	0.0718	2.549	0.0108	0.0423	0.324	1.201*	1.043	1.382
Constant	-0.933						0.393		
District variance (se)	0.0747*	0.0371							
Observations	3,859								
Number of groups	22								
LR chi squared test (df=1)	29.11***								

Note. *** p<0.001, ** p<0.01, * p<0.05

Note. January – June 2016 ISR data, CPD

Sensitivity tests indicated that potential selection on unobserved factors was potentially more of a problem. In sample 1, even a 10 percent shift in the odds of being Hispanic/non-Hispanic rendered the ethnic difference non-significant ($\Gamma = 1.10$). In sample 2, the significant difference disappeared with a 5 percent shift ($\Gamma = 1.05$).

10.5.5 Overall conclusion on race/ethnicity and enforcement

In both random samples, regression results revealed significant net impacts of both race and ethnicity on the likelihood that the stopped civilian would receive any enforcement action. Black civilians and Hispanic civilians as compared to White civilians were more likely to be on the receiving end of such actions. Diagnostics suggested a low to moderate level of concern about potential confounds due to unobserved selection. The implication is that a correlational rather than causal interpretation is probably more warranted.

Race impacts failed to re-appear in the propensity matching models.

Ethnicity impacts, however, did resurface with the propensity matching models, and their size was closely comparable to that seen in the regression results. That said, low to moderate concerns about observed selection, and strong concerns about unobserved selection given the results of diagnostics, favor a correlational rather than causal interpretation of this ethnicity impact.

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

The last planned analysis considered the potential roles of race and ethnicity in shaping whether the stop ended in one of two ways: a pat down was delivered but no enforcement action was taken, versus no pat down took place and no enforcement action was delivered. As discussed earlier, the procedural justice literature clearly implies that the former type of stop is more intrusive, and has more potential to reduce perceived institutional legitimacy.

10.6.1 Main modeling approach

The models used here were mixed effects multinomial models with stops nested within districts. Because there are four outcome categories for all possible combinations of enforcement in pat down, propensity score matching models were not feasible. Further, model diagnostics were not undertaken. Given the lack of a cross checking analysis, and the lack of detailed diagnostics on predicted scores and residuals, the results presented here should be considered **preliminary**, and certainly should not be interpreted as more than correlational.

Both samples produced statistically significant and practically sizable race and ethnicity impacts (Table 53, Table 54). In both cases, stopped Black as compared to stopped White civilians, and stopped Hispanic as compared to stopped White civilians, had odds of being patted down but no enforcement that were at least 40 percent higher. Because of the covariates were taken into account and district context was considered, these are **net** race and ethnicity impacts.

Even more sizable impacts emerged for gender. In both samples, males' odds of experiencing [a pat down with no enforcement versus no pat down and no enforcement] were at least 250 percent higher.

Finally, the pat down with no enforcement outcome was much more likely in both samples to occur with vehicle stops.

Table 53 Predicting pat down and no enforcement vs. no pat down and no enforcement: Sample 1

Variable	Variable name	B	SE	t	p=	OR
Black civilian	dblack	0.382	0.0694	5.511	<.001	1.465
Hispanic civilian	dhisp	0.350	0.0737	4.744	<.001	1.419
Male	dmale	1.272	0.0606	20.99	<.001	3.568
Age 18-25	age1825	0.0964	0.0476	2.027	0.0427	1.101
Age 26-35	age2635	-0.108	0.0530	-2.039	0.0415	0.898
Age 36-45	age3645	-0.737	0.0660	-11.17	<.001	0.479
Age 46 and up	age46pl	-1.159	0.0656	-17.68	<.001	0.314
February	dfeb	-0.0779	0.0656	-1.189	0.234	0.925
March	dmar	-0.0607	0.0610	-0.994	0.320	0.941
April	dapr	-0.330	0.0611	-5.411	<.001	0.719
May	dmay	-0.433	0.0599	-7.237	<.001	0.649
June	djun	-0.587	0.0613	-9.579	<.001	0.556
Weekend	wknddum	0.144	0.0363	3.962	<.001	1.155
3 - 6 AM	dhr0306	0.597	0.141	4.245	<.001	1.817
6 - 9 AM	dhr0609	-1.131	0.115	-9.800	<.001	0.323
9 - 12 AM	dhr0912	-0.772	0.0743	-10.40	<.001	0.462
12 - PM	dhr1215	-0.575	0.0697	-8.259	<.001	0.563
3 - 6 PM	dhr1518	-0.449	0.0715	-6.290	<.001	0.638
6 - 9 PM	dhr1821	-0.387	0.0656	-5.902	<.001	0.679
9 - 12 PM	dhr2123	-0.302	0.0670	-4.513	<.001	0.739
Vehicle stop	dvehstop	0.561	0.0767	7.311	<.001	1.752
Missing event no.	eventmis	0.108	0.155	0.702	0.483	1.114
	M1[district]	1				
	Constant	-1.245				
	District variance (se)	0.116	0.0367			
	Observations	27,051				

Note. Results from generalized multinomial structural equation model with stops nested within districts.
Results only shown for one contrast: pat down and no enforcement vs. no pat down and no enforcement.
Latter group was reference category.
Other multinomial contrasts were run as part of the same model, but are not shown here.
Source: Jan-June 2016 ISRs from CPD.
Seven cases dropped with discrepant scoring on any enforcement action vs. individual enforcement actions.

Table 54 Predicting pat down and no enforcement vs. no pat down and no enforcement: Sample 2

Variable	Variable name	B	SE	t	p=	OR
Black civilian	dblack	0.408	0.0695	5.879	<.001	1.504
Hispanic civilian	dhisp	0.360	0.0736	4.893	<.001	1.433
Male	dmale	1.284	0.0596	21.55	<.001	3.611
Age 18-25	age1825	0.0878	0.0475	1.848	0.0646	1.092
Age 26-35	age2635	-0.0938	0.0531	-1.765	0.0775	0.910
Age 36-45	age3645	-0.494	0.0639	-7.742	<.001	0.610
Age 46 and up	age46pl	-1.158	0.0661	-17.52	<.001	0.314
February	dfeb	-0.0576	0.0645	-0.893	0.372	0.944
March	dmar	-0.0529	0.0605	-0.873	0.383	0.948
April	dapr	-0.260	0.0602	-4.326	<.001	0.771
May	dmay	-0.395	0.0590	-6.701	<.001	0.674
June	djun	-0.474	0.0596	-7.960	<.001	0.623
Weekend	wknddum	0.0814	0.0362	2.252	0.0243	1.085
3 - 6 AM	dhr0306	0.648	0.137	4.737	<.001	1.912
6 - 9 AM	dhr0609	-1.082	0.114	-9.454	<.001	0.339
9 - 12 AM	dhr0912	-0.758	0.0748	-10.14	<.001	0.469
12 - PM	dhr1215	-0.484	0.0699	-6.920	<.001	0.616
3 - 6 PM	dhr1518	-0.330	0.0715	-4.617	<.001	0.719
6 - 9 PM	dhr1821	-0.314	0.0658	-4.772	<.001	0.731
9 - 12 PM	dhr2123	-0.167	0.0672	-2.486	0.0129	0.846
Vehicle stop	dvehstop	0.694	0.0769	9.033	<.001	2.002
Missing event no.	eventmis	-0.0140	0.157	-0.0891	0.929	0.986
	M1[district]	1				
	Constant	-1.402				
	District variance (se)	0.113	0.0360			
	Observations	27,054				

Note. Results from generalized multinomial structural equation model with stops nested within districts.
Results only shown for one contrast: pat down and no enforcement vs. no pat down and no enforcement.
Latter group was reference category.
Other multinomial contrasts were run as part of the same model, but are not shown here.
Source: Jan-June 2016 ISRs from CPD.
Four cases dropped with discrepant scoring on any enforcement action vs. individual enforcement actions.

10.6.2 Alternative models

Alternative models were run using canonical discriminant analysis. [16] Three orthogonal discriminant functions were generated which, collectively, sought to classify stops into one of the four groups used in this analysis. In addition to the predictors listed in the above tables, district context was controlled by adding in dummy predictors for districts 2 through 25.

Roughly, these discriminant functions, in both samples, correctly classified about 82 percent of those in the no-pat-down-no-enforcement group, and about a third of those in the pat-down-but-no-enforcement group. The multivariate F indicated the predictor variables as a set clearly distinguished between these four groups of stops ($p < .001$ by MANOVA, details not shown).

[16] This is candisc in Stata.

More importantly, the Black variable, had a sizable standardized discriminant function coefficient on discriminant function 2 (-.23 in both samples) and, in both samples, this second discriminant function explained a sizable (27 percent in both samples) and significant ($p < .001$ in both samples) portion of the variation based on group membership. The Hispanic variable had a sizable (-.20) and closely comparable standardized coefficient. As in the main analytic model, gender appeared more important for this outcome.

The discriminant analysis provides less precision than the multinomial model because it cannot take account of clustering within districts like a multilevel model can, and because it is trying to discriminate all four groups at once. Nevertheless, the group mean on Black and Hispanic was clearly different between the no-pat-down-no-enforcement group and the pat-down-but-no-enforcement groups. And both the Black and Hispanic variables each had sizable standardized loadings. So the alternative analytics seem to support the main takeaway lesson from the main multinomial model: both race and ethnicity help distinguish between membership in these two groups.

11 DISCUSSION

11.1 LIMITATIONS AND STRENGTHS

11.1.1 Limitations

Numerous limitations must be kept in mind when considering the findings of this report. These include the following.

1. Analyses depended on one source of information for police behavior: investigative stop report (ISR) data compiled by the Chicago Police Department. These are administrative reports of officer behavior that have been processed by the department. Unknown at this time is how the picture painted by these data would align with other sources of information on police behavior. Policing research has a vigorous four decade tradition based on on-site assessments of police-civilian interactions (Reiss, 1971).
2. Project time constraints and other factors resulted in models leaving out additional potentially relevant covariates that could have been used and were available in the ISR data. Models with different sets of predictors have the potential to generate different patterns of statistical significance. On the other hand, we used gamma diagnostics to try and gauge how much of a difference these other factors would need to make before significant impacts disappeared.
3. Project time constraints prevented linking up the variables used here with other potentially important and relevant predictors from sources beyond ISR data. Those might include, for example, indicators either about arrests or about some classes of calls for service when those individual arrests or calls took place close in space and time to each individual stop being analyzed. In other words, a more detailed picture including additional attributes describing the context of specific stops, could have been built up given more time. From a policing perspective an argument can certainly be made that additional features of stop context related to both calls and arrests proximate in space and

time could be relevant. Again, with different predictors different patterns of statistical significance might have been observed.

4. Because there are only (exclusive of District 31) 22 police districts, this small number of geographic units argued against including district level predictors in these models for a range of technical reasons (Bryan & Jenkins, 2016; Schmidt-Catran & Fairbrother, 2016). All that could be done here was to allow each police district to have its own mean score on each outcome. Including contextual predictors at the district level may have altered the impacts seen here of various stop-level predictors.
5. Project time constraints prevented including checking for spatial autocorrelation of various outcomes and, if needed, controlling for same by introducing a spatially lagged predictor as an outcome.
6. Project time constraints precluded testing additional varieties of the propensity score matching models (e.g., using Mahalanobis distance for matching).
7. Project time constraints precluded additional diagnostic assessment of the regression models. Most importantly, leverage and influence have yet to be examined.
8. Two of the outcomes examined here correlate significantly with each other. The current models and the alpha level used may be creating a slightly inflated experiment-wise alpha (Type I error) level.

11.1.2 Potential strengths

The above limitations should be considered in the context of several potential strengths.

1. Models examined each outcome, save the last categorical one, using at least two alternate forms of analysis. Testing links across multiple analytics provided clues to how robust any observed patterns were across modeling approaches.
2. A simple random sampling strategy divided records into two independent random samples. Doing so permitted learning whether a significant link between a predictor and an outcome, if observed, appeared in *both* random samples. If it did, that increased confidence in the durability of that link. In effect, a significant link in both samples amounts to an internal replication of a finding. More specifically, it means that the connection observed did not depend on some features of a small number of specific cases that just happened to wind up in one sample vs. another.
3. A very rough a priori statistical power analysis suggested that statistical power was ample (80 percent or better) for gauging relatively modest age and ethnicity impacts.
4. Where possible, at least some model diagnostics were completed to gauge the extent to which observed and unobserved selection were problematic.
5. For outcomes clearly involving sequential selection, appropriate selection models were used.

12 KEY FINDINGS

Patterns of observed race and ethnicity net links with the outcomes, and levels of concern suggested by various diagnostics, along with implications for how to interpret, appear in Table 55.

Pat downs. The strongest pattern revealed by these analyses are net connections between race and whether a pat down occurred, and between ethnicity and this outcome. Both analytic approaches yielded statistically significant net connections in both samples.

Diagnostics of both types of pat down models, however, suggested a moderate level of potential concern about observed and unobserved selection biases. Stated differently, there were other things going on, correlated both with key predictors and the outcome variable, that were not handled sufficiently by the analytics. Given that, the net race and ethnicity impacts are probably best interpreted as correlational. Nonetheless, the links were there, after controlling for other factors, and for district context. As compared to White non-Hispanic civilians, Black and Hispanic civilians were more likely subjected to a pat down.

Pat downs leading to weapons. Previous work on pat down and search hit rates suggested that pat downs of Black and Hispanic civilians would be less likely to lead to recovered weapons. This turned out to be true when examining weapons produced from pat downs, after controlling for other factors and district context. It held for Black as compared to White civilians. Hit rates were significantly lower in both random samples in the regression analyses. The significant net race effect did not resurface using more stringent analytics, although the race effect in one sample was marginally significant. Again, diagnostics suggested some concerns. The conclusion seems to be that there is a net race effect, but it is probably correlational and was just not quite strong enough to be robust across alternate analytics.

Searches. The search outcome results showed no significant net race effects. But significant net ethnicity links appeared, for both samples, using the more stringent alternative analytics. Diagnostics suggested some level of concern, so the conclusion about ethnicity and the search outcome is that the link is probably correlational, but not robust across different approaches.

Any enforcement action delivered. The enforcement outcome yielded robust net ethnicity links across both samples and both analytic approaches. Net race links surfaced only with one analytic approach. The conclusion seems to be, in light of diagnostics, that for both race and ethnicity there is a net connection with this outcome, that for both it is probably best considered correlational, and that for race it is not robust across alternative approaches.

Pat down and no enforcement. The last outcome examined, contrasted two types of stops, no enforcement action and no pat down vs. no enforcement and receiving a pat down. Analyses included both a main and an alternate approach. No diagnostics of either analytic model have yet been completed.

Across both analytic approaches, significant net race and ethnicity effects surfaced. After controlling for other factors and district context, in stops where no enforcement actions were taken by police, Black and Hispanic stopped civilians had much higher odds of being patted down than did stopped White non-Hispanic civilians. Given the potentially corrosive nature of police interactions such as this, this would seem to be an important pattern to address.

These net race and ethnicity links should be considered correlational only at this time, since no diagnostics have been completed, and the patterns seen may or may not be robust across different analytic approaches.

Gross impacts. The above discussion concentrates on statistically significant net impacts of racial or ethnic differences. **Authors recognize that gross ethnic or racial differences represent important findings as well. That is why, as requested by the Parties experts and as agreed, we present all of these ethnoracial differences, and geographic differences, in a number of tables.** How one balances the importance of those gross differences vs. the statistically significant net differences we leave up to the individual readers.

Table 55 Summary of results patterns, and implications, for post stop outcomes

			Outcome		
	pat down	Pat==>Weapon	Search	Enforcement	PD-No E
Result patterns					
Regression models					
Significant net race effect observed?	Y	Y	N	Y	Y
Significant net race effect replicated across both samples?	Y	Y	---	Y	Y
Significant net ethnicity effect observed?	Y	N	N	Y	Y
Significant net ethnicity effect replicated across both samples?	Y	---	---	Y	Y
Diagnostics					
Concern level about observed selection bias	moderate	low-moderate (a)	low	low-moderate	dk
Concern level about unobserved selection bias	moderate		low	low-moderate	dk
Alternate analytics					
Significant net race effect observed?	Y	N (b)	N	N	Y
Significant net race effect replicate across both samples?	Y	---	---	---	Y
Significant net ethnicity effect observed?	Y	N	Y	Y	Y
Significant net ethnicity effect replicated across both samples?	Y	---	Y	Y	Y
Diagnostics					Y
Concern level about observed selection bias	moderate	low-moderate (a)	H: low-moderate	H: low-moderate	dk
Concern level about unobserved selection bias	moderate		H: high	H: high	dk
Conclusion					
Suggested interpretation of significant net race effects	correlational	CBNR	---	CBNR	correlational
Suggested interpretation of significant net ethnicity effects	correlational		CBNR	correlational	correlational

Notes
- (a) For this model, diagnostics did not permit discriminating between concerns about observed vs. unobserved selection bias.
- (b) Marginally significant race impacts in sample 2 would have been statistically significant if a one tailed significance test was used.
- CBNR correlational but not robust across different models
- PD-No E pat down / no enforcement vs. no pat down / no enforcement
- --- not relevant because no significant net effect   dk = unknown
- CBOUR correlational but of unknown robustness across alternative analytics

13 ADDENDUM 1

The table below organizes all stops, for the three ethnoracial groups of key interest in this report, from January 1, 2016-June 30, 2016.

Stops are organized into two rows: those stops where there was no enforcement action of any kind (No Enf) and those stops where there was at least one enforcement action of any kind (Yes Enf), regardless of whether it was an arrest, a citation, an administrative action, PSC, or other.

The columns are organized into two supersets. The right hand set of columns are stops where a search took place (total = 9,595). The left hand set of columns reflect stops where no searches took place (total = 44,521).

Within each search category there are two columns, depending upon whether a pat down occurred or not. There were 14,732 pat downs when no searches took place, and 3,632 pat downs when a search also took place in the same stop. There were a total of 18,364 stops with pat downs.

The numbers in each column are broken out into two separate rows, depending on whether the stop included any enforcement action or not. In 36,691 stops no enforcement action was recorded, and in 17,425 stops some type of enforcement action was recorded.

Any Enforcement Action	No Search			Yes Search			Total
	No Pat	Yes Pat	Total	No Pat	Yes Pat	Total	
No Enf	22,611	12,414	35,025	633	1,033	1,666	36,691
Yes Enf	7,178	2,318	9,496	5,330	2,599	7,929	17,425
Total	29,789	14,732	44,521	5,963	3,632	9,595	54,116
Total pat downs						18,364	
Total searches						9,595	

14 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.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between groups in propensity-score matched samples. *Statistics in Medicine, 28*, 3083-3107.

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

Babu, J., & Jang, W. (2006). *Selection biases: truncation and censoring. Presentation of the Surveys and Population Studies working group, Astrostatitistics Program, Statistics and Applied Mathematical Sciences Institute. [ONLINE: http://sisla06.samsi.info/astro/sps/truncj.pdf ; accessed 7/1/2010]*.

Banks, R. R. (2003). Beyond profiling: Race, policing, and the drug war. *Stanford Law Review, 56*(3), 571-603.

Barnes, K. Y. (2005). Assessing the Counterfactual: The Efficacy of Drug Interdiction Absent Racial Profiling. *Duke Law Journal, 54*(5), 1089-1141.

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

Becker, S. O., & Calaiendo, M. (2007). Sensitivity analysis for average treatment effects. *The Stata Journal, 7*(1), 71-83.

Beckett, K., Nyrop, K., & Pfingst, L. (2006). Race, drugs, and policing: Understanding disparities in drug delivery arrests. *Criminology, 44*(1), 105-137.

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

Browne, W. J., Lahi, M. G., & Parker, R. M. A. (2009). A Guide to sample size calculations for random effects models via simulation ahd the MLPowSim software package. University of Bristol, Centre for Multilevel Modeling. [ONLINE: http://www.cmm.bristol.ac.uk/MLwiN/MLPowSim/index.shtml ; accessed August 14, 2010].

Brunson, R. K. (2005). Young Black Men and Urban Policing in the United States. *British Journal of Criminology, 46*(4), 613-640. doi:10.1093/bjc/azi093

Brunson, R. K. (2006). Gender, Race, and Urban Policing: The Experience of African American Youths. *Gender & Society, 20*(4), 531-552. doi:10.1177/0891243206287727

Brunson, R. K. (2007a). "Police don't like black people:" African-American young men's accumulated police experiences. *Criminology & Public Policy, 6*(1), 71-101.

Brunson, R. K. (2007b). "POLICE DON'T LIKE BLACK PEOPLE": AFRICAN-AMERICAN YOUNG MEN'S ACCUMULATED POLICE EXPERIENCES*. *Criminology & Public Policy, 6*(1), 71-101.

Brunson, R. K., & Gau, J. M. (2011). Officer Race Versus Macro-Level Context: A Test of Competing Hypotheses About Black Citizens' Experiences With and Perceptions of Black Police Officers. *Crime & Delinquency*. doi:10.1177/0011128711398027

Brunson, R. K., & Miller, J. (2006). Gender, race, and urban policing - The experience of African American youths. *Gender & Society, 20*(4), 531-552.

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

Bursik, R. J. J., & Grasmick, H. G. (1993). *Neighborhoods and crime*. Lexington: Lexington.

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).

Carroll, L., & Gonzalez, M. L. (2014). Out of Place: Racial Stereotypes and the Ecology of Frisks and Searches Following Traffic Stops. *Journal of Research in Crime and Delinquency, 51*(5), 559-584. doi:10.1177/0022427814523788

Charpentier, A. (2013). Residuals from a logistic regression. *Freakonometrics. An Open lab-notebook experiment. [ONLINE: http://freakonometrics.hypotheses.org/8210 ; accessed 11/21/2016]*. Retrieved from

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*, 829-836.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Delgado, R., & Stefanic, J. (2012). *Critical Race Theory* (Second ed.). New York: New York University Press.

Engel, R. S. (2008). A Critique of the “Outcome Test” in Racial Profiling Research. *Justice Quarterly, 25*(1), 1-36. doi:10.1080/07418820701717177

Engel, R. S., & Calnon, J. M. (2004). Comparing Benchmark Methodologies for Police-Citizen Contacts: Traffic Stop Data Collection for the Pennsylvania State Police. *Police Quarterly, 7*(1), 97-125. doi:10.1177/1098611103257686

Engel, R. S., Calnon, J. M., & Bernard, T. J. (2002). Theory and racial profiling: Shortcomings and future directions in research. *Justice Quarterly, 19*(2), 249-273.

Engel, R. S., & Tillyer, R. (2008). Searching for Equilibrium: The Tenuous Nature of the Outcome Test. *Justice Quarterly, 25*(1), 54-71. doi:10.1080/07418820701717243

Fagan, J. (2002). Law, social science, and racial profiling. *Justice Research and Policy, 104*, 104-129.

Fagan, J., & Braga, A. A. (2015). *Final Report: An Analysis of Race and Ethnicity Patterns in Boston Police Department Field Interrogation, Observation, Frisk, and/or Search Reports.*

Fagan, J., Geller, A., Davies, G., & West, V. (2009). Street Stops and Broken Windows Revisited: The Demography and Logic of Proactive Policing in a Safe and Changing City. In S. K. Rice & M. D. White (Eds.), *Race, Ethnicity and Policing: New and Essential Readings*. New York: New York University Press.

Fallik, S. W., & Novak, K. J. (2012). The Decision to Search: Is Race or Ethnicity Important? *Journal of Contemporary Criminal Justice, 28*(2), 146-165. doi:10.1177/1043986211425734

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods, 41*(4), 1149-1160. doi:10.3758/BRM.41.4.1149

Freedman, D. A. (2006). On the So-Called "Huber Sandwich Estimator" and "Robust Standard Errors". *The American Statistician, 60*(4), 299-302.

Fridell, L. A. (2005). *Racially Biased Policing: Guidance for Analyzing Race Data from Vehicle Stops. Washington, DC: Police Executive Research Forum &*

Community Oriented Policing Services, US Department of Justice. [ONLINE: http://www.cops.usdoj.gov/pdf/publications/Racially_Biased_Policing_Guidance.pdf ; accessed August 18, 2015]. Retrieved from

Fu, V. K., Winship, C., & Mare, R. D. (2004). Sample selection bias models. In M. Hardy & A. Bryman (Eds.), *Handbook of Data Analysis* (pp. 409-430). Thousand Oaks: Sage.

Gau, J. M., & Brunson, R. K. (2010). Procedural Justice and Order Maintenance Policing: A Study of Inner‐City Young Men's Perceptions of Police Legitimacy. *Justice Quarterly, 27*(2), 255-279. doi:10.1080/07418820902763889

Gelman, A., Fagan, J., & Kiss, A. (2007). An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias. *Journal of the American Statistical Association, 102*(479), 813-823. doi:10.1198/016214506000001040

Gottfredson, M. R., & Gottfredson, D. M. (1988). *Decision Making in Criminal Justice: Toward the Rational Exercise of Discretion* (2nd ed.). New York: Plenum.

Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association, 101*(475), 878-887.

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

Harris, D. A. (1997). "Driving while Black" and all other traffic offenses: The Supreme court and pretexual traffic stops. *The Journal of Criminal Law & Criminology, 87*(2), 544-582.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica, 45*, 153-161.

Hilbe, J. M. (2009). Logistic Regression Models. Boca Raton: CRC Press.

Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied Logistic Regression* (Second ed.). New York: Wiley.

Imbens, G. W., & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5-86.

Jernigan. (2000). Driving while black: racial profiling in America. *Law & psychology review, 24*, 127.

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

Knowles, J., Persico, N., & Todd, P. (2001). Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy, 109*(1), 203-229. doi:doi:10.1086/318603

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

Lundman, R. J., & Kaufman, R. L. (2003). Driving While Black: Effects of race, ethnicity, and gender on citizen self-reports of traffic stops and police actions. *Criminology, 41*(1), 195-220.

MacDonald, J., Stokes, R. J., Ridgeway, G., & Riley, K. J. (2007). Race, neighborhood context and perceptions of injustice by the police in Cincinnati. *Urban Studies, 2007*, 2567-2585.

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

Mastrofski, S. D., Reisig, M. D., & McCluskey, J. D. (2002). Police disrespect toward the public: An encounter-based analysis. *Criminology, 40*(3), 519-551.

McArdle, A., & Erzin, T. (Eds.). (2001). *Zero Tolerance: Quality of Life and the New Police Brutality in New York City*. New York: New York University Press.

Meares, T. L. (2014). The Law and Social Science of Stop and Frisk. *The Annual Review of Law and Social Science, 10*, 335-352.

Persico, N., & Todd, P. E. (2008). The Hit Rates Test for Racial Bias in Motor ‐ Vehicle Searches. *Justice Quarterly, 25*(1), 37-53. doi:10.1080/07418820701717201

Petersilia, J., & Turner, S. (1990). Comparing intensive and regular supervision for high-risk probationers: Early results from an experiment in California. *Crime & Delinquency, 36*(1), 87-111.

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

Pratt, T. C., & Cullen, F. T. (2005). Assessing macro-level predictors and theories of crime: A meta-analysis *Crime and Justice: A Review of Research* (Vol. 32, pp. 373-450).

Reisig, M. D., McCluskey, J. D., Mastrofski, S. D., & Terrill, W. (2004). Suspect disrespect toward the police. *Justice Quarterly, 21*(2), 241-268.

Reiss, A. J., Jr. (1971). *The Police and the public*. New Haven: Yale University Press.

Reskin, B. (2012). The Race Discrimination System. *Annual Review of Sociology, 38*(1), 17-35. doi:doi:10.1146/annurev-soc-071811-145508

Ridgeway, G. (2006). Assessing the effect of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology, 22*(1), 1-29.
Ridgeway, G. (2007a). *Analysis of racial disparities in the New York Police Department's Stop, Question, and Frisk Practices. Technical Report.* Retrieved from Santa Monica, CA:
Ridgeway, G. (2007b). *Disparities in the New York Police Department;s stop, question, and frisk practices. Technical Report*. Retrieved from Santa Monica, CA:
Ridgeway, G. (2009). *Cincinnati Police Department Traffic Stops: Applying RAND's Framework to Analyze Racial Disparities*. Santa Monica, CA: RAND Corporation.
Ridgeway, G., & MacDonald, J. (2010). Methods for assessing racially biased policing [ONLINE: http://www.rand.org/pubs/reprints/RP1427.html]. In S. K. Rice & M. D. White (Eds.), *Race, Ethnicity, and Policing: New and Essential Readings* (pp. 180-204). New York: New York University Press.
Ridgeway, G., & MacDonald, J. M. (2009). Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops. *Journal of the American Statistical Association, 104*(486), 661-668. doi:10.1198/jasa.2009.0034
Ridgeway, G., & Riley, K. J. (2007). Assessing Racial Profiling More Credibly. *Research Brief. Rand Corporation Public Safety and Justice RB-9070-OAK, 2004 [ONLINE: http://www.rand.org/pubs/research_briefs/RB9070/ ; accessed Octrober 1, 2015]*.
Rojek, J., Rosenfeld, R., & Decker, S. (2012). POLICING RACE: THE RACIAL STRATIFICATION OF SEARCHES IN POLICE TRAFFIC STOPS. *Criminology, 50*(4), 993-1024. doi:10.1111/j.1745-9125.2012.00285.x
Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (pp. 1809-2014). New York

John Wiley.
Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*(3-4), 169-188.
Sampson, R. J., Morenoff, J. D., & Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology, 28*, 443-478.

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

Simon, D., & Burns, E. (1997). *The Corner: A Year in the life of an inner-city neighborhood*. New York: Broadway Books.

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

Spybrook, J., Raudenbush, S. W., Congdong, R., & Martinez, A. (2009). Optimal design for longitudinal and multilevel research: Documentation for the 'Optimal Design' software.[ONLINE: http://sitemaker.umich.edu/group-based/optimal_design_software ; accessed 1/18/2010] Retrieved from

Sunshine, J., & Tyler, T. (2003). The Role of Procedural Justice and Legitimacy in Shaping Public Support for Policing. *Law & Society Review, 37*(3), 513-548.

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.

Terrill, W., & Mastrofski, S. D. (2002). Situational and officer-based determinants of police coercion. *Justice Quarterly, 19*, 215-248.

Tillyer, R., Engel, R. S., & Cherkauskas, J. C. (2010). Best practices in vehicle stop data collection and analysis. *Policing-an International Journal of Police Strategies & Management, 33*(1), 69-92. doi:10.1108/13639511011020601

Tillyer, R., Klahm, C. F., & Engel, R. S. (2012). The Discretion to Search: A Multilevel Examination of Driver Demographics and Officer Characteristics. *Journal of Contemporary Criminal Justice, 28*(2), 184-205. doi:10.1177/1043986211425721

Tyler, T. (1988). What is Procedural Justice? Criteria Used By Citizens to Assess the Fairness of Legal Procedures. *Law and Society Review, 22*(1), 103-135.

Tyler, T. (1997). Citizen discontent with legal procedures. *American Journal of Comparative Law, 45*, 869-902.

Tyler, T. (2001). Public Trust and Confidence in Legal Authorities: What Do Majority and Minority Group Members Want from the Law and Legal Institutions? *Behavioral Science and the Law, 19*, 215-235.

Tyler, T. (2003). Procedural justice, legitimacy, and the effective rule of law *Crime and Justice: A Review of Research, Vol 30* (Vol. 30, pp. 283-357).

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

Tyler, T., & Huo, Y. J. (2002). *Trust in the Law: Encouraging Public Cooperation With the Police and Courts*: Russell Sage Foundation.

Tyler, T., & Lind, E. A. (2001). Procedural Justice. *Handbook of Justice Research in Law*, 65-92.

Walker, S. (2001). Searching for the denominator: Problems with police traffic stop data and an early warning system solution. *Justice Research and Policy, 3*, 63-95.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica, 50*(1-25).