Assessing Racial Disparities in Traffic Stops

Frank R. Baumgartner
frankb@unc.edu

Leah Christiani
christiani@unc.edu

Derek A. Epp
depp@austin.utexas.edu

Kevin Roach
kevroach@live.unc.edu

Kelsey Shoub
shoub@unc.edu

Abstract

We develop the largest database so far collected of differences in the rates at which drivers of different racial groups are subject to search after traffic stops, across hundreds of US police jurisdictions. We then show the robustness of a simple test of disparity: the search-rate ratio. A search rate ratio is simply the ratio of two search rates. This ratio can be calculated for any agency which publishes statistics on rates of traffic stops and their outcomes; hundreds of police agencies across the nation. We analyze over 60 million individual traffic stop records; this covers the states of Connecticut, Florida, Illinois, Maryland, North Carolina, Ohio, and Texas. For each jurisdiction within these states where a minimum number of traffic stops and searches have occurred, we calculate both the simple ratio described above as well as the most complete logistic regression that the data collection allows. Depending on the state, such a model includes controls for race, gender, ethnicity, age, the reason the driver was stopped, the time of day and day of week, and (for a few states) such things as the race of the officer, the age of the car, and whether the car had out-of-state plates. We then compare these logistic odds ratios to the simple search rate ratios. We show that these two measurements track well with each other. In this paper, we focus on establishing that shared methodological understanding. Substantively, we also document that across hundreds of police agencies in annual reports dating from the 1990s to present, racial disparities are not only large, but that they are virtually ubiquitous across the country. Finally, we show that disparities are systematically greater in municipal police departments than in county sheriff's offices or state highway patrols. They are, however, substantial and disturbing in each type of agency.

Prepared for presentation at the Annual Meeting of the American Political Science Association, San Francisco CA, August 31–September 3, 2017

Introduction

The United States is currently going through a period of renewed attention to questions of racial justice. The concern that police officers direct undue (and sometimes fatal) scrutiny to African Americans and other minorities has prompted state, local, and media organizations to seek out and engage with data on police-citizen encounters, so that the extent of any disparities can be fully documented. Such inquiries are long overdue and, one might think, mostly straightforward. If a jurisdiction keeps some record of police actions – searches, arrests, or citations – and notes the demographic characteristics of people subjected to these actions, then it is a simple matter to determine if African Americans are more or less likely to experience punitive outcomes as compared to whites. Many police agencies across the country do in fact maintain such datasets (or have previously done so) and analysts have used this data to show that racial disparities tend be to be large and pervasive: across the country in cities large and small, African Americans are more likely to be searched, arrested, and ticketed than their white counterparts.

There is, however, persistent concern that what might at face value look like a racial disparity can actually be attributed to some other factor that merely correlates with race. For example, it is possible that the police are more likely to pull over people in run-down cars. If African Americans are more likely to drive run-down cars then we might mistake a disparity that has its origins in another place as being about race. There are a number of potentially confounding possibilities: perhaps African Americans are more likely to drive at night, or to break the law, or to live in neighborhoods with high murders rates that have been targeted by law enforcement for more intensive police patrols.

Of course, social scientists are quite familiar with the importance of controlling for confounding factors. What makes it difficult in this case is that while many jurisdictions across the US have some rudimentary data on policing that allows for a simple assessment of racial disparities, few have the kind of detailed, comprehensive data that would allow an analyst to control for enough rival possibilities confidently to rule out a spurious association. This is unfortunate as it suggests that much of the available data is inadequate to answering urgent questions about the role of race in law enforcement outcomes.

We present findings in this paper suggesting that such a perspective is not justified empirically. We show that, in fact, the simple search rate ratio is an extremely robust indicator of racial difference. We arrive at this conclusion by analyzing datasets on traffic stops that took place in Connecticut, Florida, Illinois, Maryland, North Carolina, Ohio, and Texas over the last 20 years; a total of around 60 million observations. Our approach is to conduct whatever multivariate analysis is possible with the variables collected. If a police agency collected only demographic information then we calculate the "search-rate ratio," which is simply the likelihood that a black driver is searched after being pulled over relative to the likelihood of search for white drivers. (A ratio of 2.0 would indicate that black drivers are twice as likely to be searched.) But when more contextual data is available, such as the age of the vehicle or the time of day the stop was made, then we estimate the likelihood that a black driver will experience a search using logistic regression, including whatever control variables are available. When we compare the parameter estimates with the search-rate ratios for the same jurisdictions we find that they are very highly correlated. One can predict the logistic regression coefficients very well with the simple search-rate ratio. It appears then that the logistic regressions are not telling us very much about racial disparities beyond what we learn from very simple search rate comparisons, which can be made for most jurisdictions where traffic stops data is collected. That these measures should tell the same story is not surprising to anyone who believes that race actually is a deciding factor in how motorists are treated by law enforcement. Still, our findings should reassure analysts that there does not appear to be any systematic, nation-wide factor that is confounding with race.

Comparing Two Racial Disparity Measures

How should inequity in treatment by police during a traffic stop be measured? Techniques to do so range from logistic regressions (Baumgartner, Epp, Shoub, and Love 2017; Gelman, Fagan, and Kiss 2007) to the estimation of latent thresholds to search (Simoiu, Corbett-Davise, and Goel forthcoming). While the specific techniques to do so vary, they have one thing in common: they require a lot of information about every possible traffic stop and its outcome to be known. However, the data to calculate such values is not always available. As a result, we use an alternative measure of inequity in outcomes that requires much less information be available and requires no micro-level information be gathered by the researcher. This is a search rate ratio.

However, because a rate ratio does not and cannot take into account any contextual factors, researchers might be concerned that it does not capture the same information or process that methods using more information about the stop do. If race is actually a systematically small factor in the decision to search, and it is actually the context of the stop that largely determines whether a driver is searched, we would expect a search rate ratio to not provide the same information or measure the same thing as more complicated methods. Here we test this. We compare a simple search rate ratio with the odds ratio coefficient that comes from a logistic regression that controls for everything possible in the dataset provided.

Micro-Level Data on Traffic Stop Outcomes

In order to compare the two measures – a search rate ratio between two groups and the odds ratio coefficient from a logistic regression for the groups – we turn towards the most comprehensive set of micro-level traffic stops data, which was collected by the authors (Baumgartner et al.

2017). This includes information on individual traffic stops from 5 states (Connecticut, Florida, Illinois, Maryland, and North Carolina) and over multiple years. Each state collects information from different sets of agencies within its bounds; some collect information on traffic stops for every agency within state lines while others only collect information on the State Highway Patrol. Additionally, each state has collected its data for different amounts of time, and each state has different requirements about what other characteristics of the stop are reported, such as the time of day, purpose of the stop, age of the vehicle, and so on.

To maximize the number of observations with which to compare the two measures of disparity, we subset data provided by each state by agency and time window. We establish thresholds that a given agency-window must meet to be included in the analysis at all in order to ensure robustness of the measures. First, an agency-window must include at least 10,000 traffic stops. Second an agency-window must include at least 100 stops of white drivers, 100 stops of black drivers, and 100 stops of Hispanic drivers. For each of these subsets, we then calculate the two measures. The first step in creating these subsets was to divide the data into agency-year subsets. Then for those agency-years that meet the thresholds the measures were calculated. For those agency years where the threshold was not met, we combine data across adjacent years for that same agency until the thresholds are met. See the appendix for additional information on this process.

The resulting number of agency-windows that meet these thresholds are presented in Table 1. Additionally, the total number of stops and searches included across these agency-windows by state are shown. The first column indicates the state, the second column indicates the number of agency-windows, the third column indicates the number of stops included, and the fourth column indicates the number of searches conducted.

Table 1. Summary of Data Included in Comparison of Measures by State

State	Agency-Windows (N)	Stops (N)	Searches (N)
Connecticut	5	348,088	6,070
Florida	6	1,011,230	5,833
Illinois	1,046	19,132,923	1,088,130
Maryland	68	2,548,013	75,071
North Carolina	503	9,447,291	517,237
All	1,628	32,487,545	1,692,341

In Table 1, we can see high variation in both the number of agency-windows included by state and the number of stops and searches. The high variation in the number of agency-windows is attributable to what agencies actually collect traffic stops data and make it publicly available. Only North Carolina and Illinois require all agencies in each state meeting some minimum threshold to report, and only these states have near complete compliance with the law. Maryland also requires all agencies to report and make public information on every traffic stop conducted, but multiple agencies (notably the city of Baltimore) do not do so. In Connecticut only the largest agencies must report. In Florida, only the state highway patrol reports and makes public their data. For additional information on the laws mandating the collection of these data, see Baumgartner et al. 2017.

For each identified agency-window, we calculate two measures of racial disparity in searches following a traffic stop. The first is a simple search rate ratio. To calculate this statistic, the only information needed is the number of people stopped by race and the number of people searched by race. First, search rates of two groups, group A and B, are calculated by dividing the number searched belonging to that group divided by the number stopped belonging to that group. If black drivers are group A and white drivers are group B, then this means the percent of black drivers searched is calculated and the percent of white drivers searched is calculated. Second, the percent searched of group A is divided by that of another, group B. Here this means that the black search rate is divided by the white search rate. The result is a number bounded by 0, where

1 indicates equality. Values below 1 indicate group B, white drivers, are searched at higher rates, while values above 1 indicate group A, black drivers, are searched at higher rates. Where SRR stands for search rate ratio, mathematically this is:

$$SRR = \frac{Searches\ of\ Group\ A \big/ Stops\ of\ Group\ A}{Searches\ of\ Group\ B \big/ Stops\ of\ Group\ B}$$

Second, we fit a logistic regression predicting whether a driver is searched using the micro-level data for each agency-window. Regressions vary by state, but consistently include race, gender, age, stop purpose, time of stop, and day of week of the stop. Additional variables depend on what information the state collects and makes public, such as vehicle age. The variables included in each regression by state are shown in Table 2. From each regression, we extract the odds-ratio for the indicators of whether the driver is black and whether the driver is Hispanic.[1] This generated approximately 1,600 coefficients for each racial group.

Table 2. Summary of Variables Included in Regressions by State

Variable	CT	FL	IL	MD	NC
Race	Y	Y	Y	Y	Y
Gender	Y	N	Y	Y	Y
Age	Y	Y	Y	Y	Y
Stop Purpose	Y	Y	Y	Y	Y
Out of State	Y	N	Y	Y	N
Vehicle Age	N	N	N	N	N
High Disparity Officer (Black)	Y	Y	N	Y	N
High Disparity Officer (Hispanic)	Y	Y	N	Y	N
Hour of Day	Y	Y	Y	Y	Y
Day of Week	Y	Y	Y	Y	Y

Note: Y indicates the variable was included. N indicates the variable was not included.

[1] We impose another threshold before we calculate these odds-ratios. We exclude agency-windows where fewer than ten searches of the relevant racial group were conducted. See our appendix for an explanation of why we do so. When the number of relevant searches is extremely low, the standard error of the estimated odds-ratio can be extremely high.

Comparing Search-Rate Ratios and Odds Ratios from Regressions

To test whether the search-rate ratio captures the same information as the odds ratio coming from a logistic regression, we simply compare both the Black - White search-rate ratio and the Hispanic - White search-rate ratio to their respective odds-ratio coefficients. Table 3 presents the results.

Table 3. Comparing Search-Rate Ratios and Odds Ratios from Logistic Regressions

Driver Race	State	Constant	Slope	N	Adjusted R^2
Black	IL	0.42* (0.03)	0.73* (0.01)	981	0.80
	NC	0.31* (0.04)	0.75* (0.02)	501	0.69
	MD	0.26* (0.06)	0.71* (0.03)	68	0.86
	FL	0.14 (0.17)	0.87* (0.07)	6	0.97
	CT	0.74 (0.76)	0.48 (0.27)	5	0.35
	All States	0.34* (0.02)	0.75* (0.01)	1,561	0.82
Hispanic	IL	0.21* (0.04)	0.82* (0.01)	990	0.91
	NC	0.14* (0.14)	0.58* (0.01)	480	0.78
	MD	0.04 (0.05)	0.76* (0.03)	59	0.90
	FL	-0.23 (0.69)	1.17* (0.37)	6	0.64
	CT	-1.10 (1.19)	1.13 (0.48)	5	0.53
	All States	-0.09* (0.02)	0.87* (0.01)	1,540	0.92

Note: * indicates statistically significant at the 0.05 level. Standard errors in parentheses. Entries are the results from a regression predicting the logistic odds-ratio using the search-rate ratio as the predictor variable.

As can be seen in Table 3, we are able to generate 1,5r0 comparisons of Hispanic-white search disparities, and 1,561 black-white comparisons. Each is based on at least 10,000 traffic stops, as discussed above, and many are based on hundreds of thousands. Except for the case of Connecticut, which has only five observations, the correlations between the two measures are extremely high. More importantly, for all the cases again except Connecticut, the search-rate ratio can be used to predict the odds-ratio. Remarkably, the results are highly consistent even though the different models in each state include different predictor variables. For black drivers, the odds-ratio can be predicted to be 0.34 + .75 (search rate ratio), and for Hispanic drivers, it can be predicted as -0.09 + .87 (search rate ratio). These numbers differ only slightly from state to state, making clear that the trends are extremely robust. The fact that the slops tend to be slightly below 1.00 means that other factors do, in fact, tend to dampen the bi-variate linkage between race and the probability of search. However, as we see below, the racial effect is nearly ubiquitous and remains so after statistical controls for other factors such as stop purpose and time of day.

These robust findings are further made apparent in Figure 1, where the search rate ratios and odds-ratio coefficients are plotted against each other for both black (Figure 1a) and Hispanic drivers (Figure 1b). The y-axis in each subfigure presents the estimated odds ratio from the logistic regression for that group. The x-axis in each subfigure presents the estimated search rate ratio. Each dot represents and agency-window. The grey lines in each subfigure indicate the estimated regression line. The data in Figure 1 correspond to the "all states" rows in Table 2, and have the N's of 1,561 and 1,540, respectively.

Figure 1. Comparing Search Rate Ratios to Odds Ratios by Race

a) Black Drivers b) Hispanic Drivers

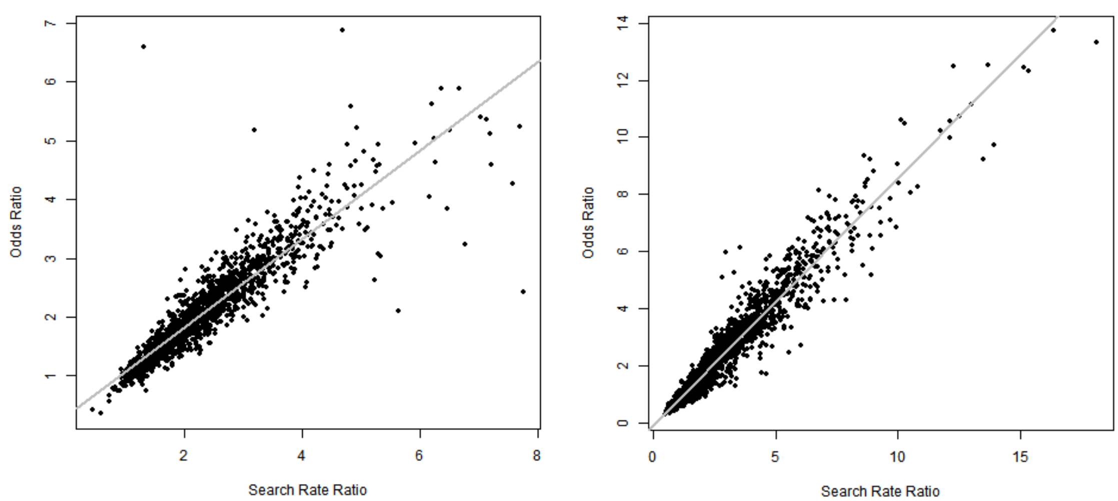


Figure 1, along with Table 2, makes clear that the search-rate ratio is a robust indicator of racial difference in the likelihood of search. Where the search rate ratio is extremely low, so is the multivariate odds-ratio. A small number of cases are poorly predicted as shown in Figure 1a, but in general the results are extremely robust.

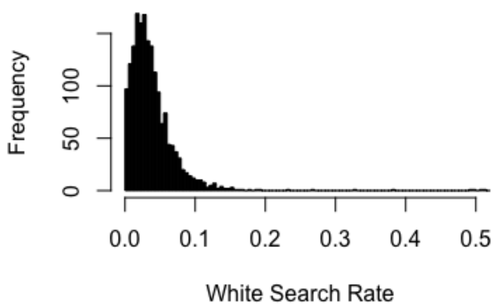
The Ubiquitous Nature of Racial Disparities in Searches

The robust nature of the search-rate ratio allows for an expanded analysis of racial disparities. While we were only able to use micro-level data from five states, many more states collect and publish their own reports about traffic stops that include the total number of stops and searches by race. From these aggregate numbers, we can calculate the search rate by race and subsequently the search rate ratio for blacks and Hispanics, as compared to whites. We use all the agencies included above as well as published reports from 81 additional agencies from various states including Arizona, Colorado, Maryland, Missouri, Nebraska, Oregon, Tennessee, Texas, Vermont, Washington, and West Virginia.

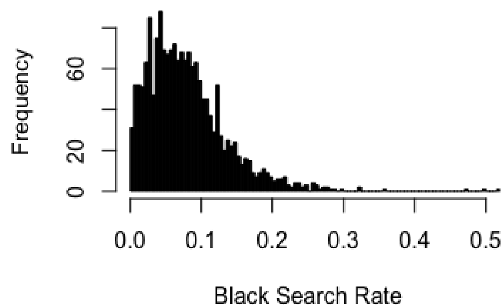
Figure 2 reports the distribution of search rates by race. The average search rate for a white driver is 3.92% while for a black driver it is 8.05% and for a Hispanic driver, 9.55%. The average minority driver is searched at a rate that is twice that of their white counterparts. It is clear that minority drivers are searched at higher rates than whites on average, for every agency-window in our dataset. As is clear from Figure 2, agency-window search rates vary widely. For whites, the search rate ranges from 0.02% to 63.75%. For blacks and Hispanics, the numbers are similar, with rates ranging from 0.05%-67.64% and 0.07%-74.01% respectively. But the means are much higher for minority drivers. And, as we will see, almost every agency has a search-rate ratio above 1.0.

Figure 2: Search Rates by Race

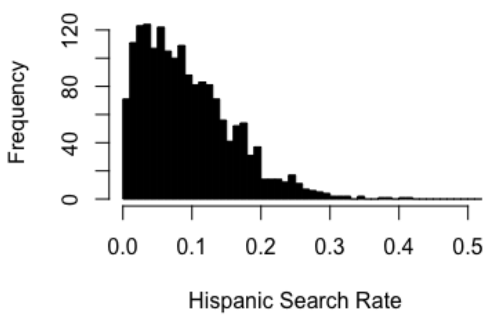
a) White Search Rate



b) Black Search Rate



c) Hispanic Search Rate



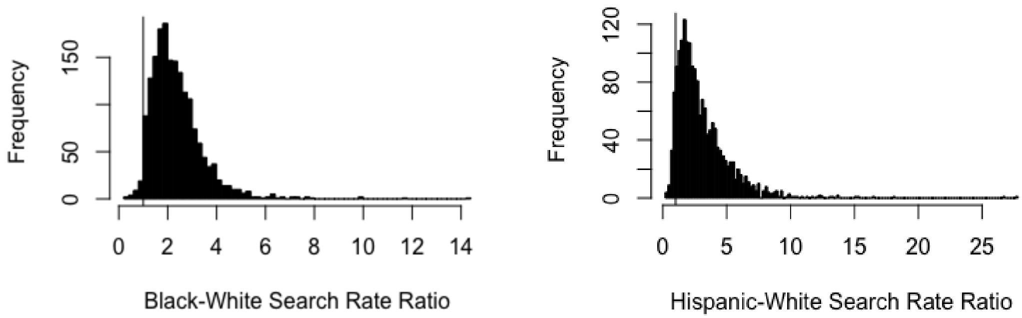
Note: Search rate for every agency-window that met the threshold of 10,000 overall stops including 100 stops and 1 search from each racial group (white, black, and Hispanic). N=1,769.

As demonstrated in the previous section, search rate ratios correlate highly with the coefficients from regressions that control for a number of potential confounding variables. As such, they can be understood as important and informative indicators of racial disparities. The distributions of search rate ratios for blacks and Hispanics, as compared to whites, are plotted in Figure 3. There is a vertical line that marks a search rate ratio of one, the value for which the minority search rate and white search rate are equal. As is clear from both subfigures, the majority of agency-windows have search rate ratios much higher than one. For the black-white search rate ratio, the mean is 2.36, the minimum is 0.30, and the maximum is 14.38. That means that on average, black drivers are searched at a rate that is more than double that of whites. Only 34 observations have a black-white search rate ratio that is less than or equal to one while 1,735 agency-windows have search rate ratios that are greater than one; 34 observations represent fewer than 2 percent of the total.

For the Hispanic-white search rate ratio, the story is similar. The mean is 3.06, the minimum is 0.21, and the maximum is 27.70. A mean of 3.06 indicates that on average, Hispanic drivers are searched at three times the rate of whites. This mean is even greater than that of blacks, who had a black-white search rate ratio mean of 2.36. There is more variation in the Hispanic-white search rate ratio, though. While only 34 agencies had black-white search rate ratios that equaled or dipped below one, 119 agency-windows have Hispanic-white search rate ratios that are less than or equal to one. Still, just like for the black-white search rate ratios, those 119 observations represent a low percentage of the total: just 6.7 percent.

Figure 3: Distribution of Search Rate Ratios

a) Black-white search rate ratios   b) Hispanic-white search rate ratios



Note: Search rate for every agency-window that met the threshold of 10,000 overall stops including 100 stops and 1 search from each racial group (white, black, and Hispanic). N=1,769.

Contrast these findings with the Department of Justice's findings in their report on Ferguson, Missouri. In their sweeping review of the Ferguson Police Department, they identified multiple instances of racial bias on the part of the police force. For traffic stops, the DOJ found that the Ferguson Police Department was 2.07 times more likely to search blacks, following a traffic stop, than whites (see USDOJ 2015). This finding put a national spotlight on Ferguson as a hotbed of racial inequality. However, our findings demonstrate that while searching black drivers twice as frequently as white drivers is substantively unequal and an indicator of great racial bias, it is indeed the common practice for many agencies throughout the country.

In fact, Ferguson would not even rank among the top ten agencies in our dataset with a search rate ratio of 2.07. Instead, it is even below the mean of 2.36. Table 4 reports the ten agency-windows with the highest search rate ratios and the ten agencies with the lowest search rate ratios. Both the highest and lowest outlier agency-windows tend to come from Illinois, which is not surprising as it makes up a large portion of our data. The Lake Forest Police record the highest search rate ratio, at more than 14, indicating that black drivers are searched at a rate that is 14 times that of white drivers. Libertyville Police records a similar high of 11 for the time period 2010-2013. Apart from these, the rest of the outliers have black-white search rate disparities for which black drivers are searched seven to nine times that of white drivers. On the low end, the Cary Police in Illinois have the lowest search rate ratio for 2005-2006, at a value of 0.299. Here, black drivers are searched at a fraction of the rate of whites. Similarly, the rest of the low outliers display values which indicate that black drivers are searched less frequently than white drivers for the specific time period listed. Again, though, it is important to keep in mind that only 34 observations reported black-white search rate ratios at or below one, while 1735 agency-windows reported black-white search rate ratios greater than one.

Table 4: High and Low Outliers for Black-White Search Rate Ratios
a) Top ten highest black-white search rate ratios

		Time Window		
State	Agency	Begin	End	Ratio
IL	Lake Forest Police	2011	2013	14.383
IL	Libertyville Police	2010	2013	11.772
IL	Du-Page County Sheriff	2010	2011	9.870
IL	Lake Forest Police	2008	2010	9.834
IL	Maywood Police	2005	2009	7.891
IL	Evanston Police	2010	2010	7.769
IL	Evanston Police	2013	2013	7.648
IL	McHenry Police	2008	2010	7.226
IL	Chicago Police	2010	2010	7.201
IL	Evanston Police	2014	2014	7.173

b) Bottom ten lowest black-white search rate ratios

		Time Window		
State	Agency	Begin	End	Ratio
IL	Cary Police	2005	2006	0.299
IL	Sangamon County Sheriff	2008	2010	0.339
NC	Tarboro Police Department	2009	2014	0.426
IL	Hawthorn Woods Police	2005	2007	0.514
IL	Riverwoods Police	2005	2007	0.590
IL	Sangamon County Sheriff	2005	2007	0.596
IL	Du-Page County Sheriff	2007	2009	0.624
IL	Troy Police	2005	2008	0.650
NC	Appalachian State University	2002	2011	0.669
IL	Cary Police	2010	2013	0.686

The outliers for the Hispanic-white search rate ratio are reported in Table 5. The Lake Forest Police in Illinois again have a high racial disparity for Hispanic drivers, as they did for whites. The Hispanic-white search rate ratio of 27.702 is even higher than their black-white search rate ratio of a little more than 14. This means that from 2011-2013, the Lake Forest Police searched Hispanic drivers at a rate of 27 times the rate they searched white drivers. Because of the strict threshold that we put on the observations in our data, we know that this is not a statistical fluke but instead, a robust value. Further, this seems to be a pattern for the Lake Forest Police because from 2008-2010, they searched Hispanics at a rate that was 26 times that of whites.

The low outliers for the Hispanic-white search rate ratios do have rates below one, indicating that whites are searched at higher rates than Hispanics. Again, though, it is important to remember that only 119 of 1769 observations reported rates that were at or below one. All others reported rates that were higher for Hispanic drivers.

Table 5: High and Low Outliers for Hispanic-White Search Rate Ratios

a): Top ten highest Hispanic-white search rate ratios

		Time Window		
State	Agency	Begin	End	Ratio
Il	Lake Forest Police	2011	2013	27.702
Il	Lake Forest Police	2008	2010	26.739
Il	Cook County Sheriff	2009	2009	18.139
Il	Cook County Sheriff	2008	2008	16.437
Il	Western Springs Police	2009	2012	15.396
Il	Cook County Sheriff	2010	2010	15.126
Il	Glencoe Police	2005	2010	13.754
Il	Cook County Sheriff	2011	2011	13.644
Il	West Chicago Police	2005	2007	13.279
Il	Palatine Police	2007	2008	13.087

b): Bottom ten lowest Hispanic-white search rate ratios

		Time Window		
State	Agency	Begin	End	Ratio
IL	Pulaski County Sheriff	2007	2008	0.2126
IL	Williamson County Sheriff	2005	2010	0.2662
MD	Fruitland	2013	2016	0.3082
IL	Pulaski County Sheriff	2009	2010	0.3843
NC	Gastonia Police Department	2015	2016	0.4243
IL	Macon County Sheriff	2011	2013	0.4485
NC	Tarboro Police Department	2009	2014	0.4734
NC	Kannapolis Police Department	2011	2012	0.4994
NC	Cornelius Police Department	2007	2010	0.5419
IL	Niles Police	2010	2011	0.5491

All in all, when we examine search rates and search rate ratios from agencies across the country, the trend is clear: racial disparities in traffic stops are persistent and ubiquitous. They are not only present in one region of the country or during one time period. Instead, they are common practice throughout the country and throughout every time period in our dataset. While the Ferguson Police Department was highlighted as a particularly problematic department in the US DOJ investigation into its practices, our results contextualize this finding and demonstrate that in fact it was no worse than the norm. Racial disparities in traffic stops are nation-wide and the variance we observe is really a matter of degree.

Comparing Police Departments, Sheriff's Offices, and State Highway Patrols

While high search rate ratios are widespread, we push the analysis further by asking whether the disparities vary in a predictable manner. Specifically, do agencies that afford their officers less discretion have lower disparities? For this, we focus on the comparison of state agencies (ex. State Highway Patrol) to police departments patrolling specific municipalities (ex. Durham Police Department). State agencies on average have a mission to ensure road and transportation safety, and as a result focus on traffic safety. Conversely, police departments must split their time between day-to-day safety and crime investigations; they may use traffic stops as a supplemental investigatory tool at the discretion of the officer. Additionally, we test how each compares to sheriff's departments. County sheriffs are elected, and as a result tend to limit their actions on average in comparison to police departments. Further, sheriff's departments are more likely to be in rural areas and their deputies may be involved in more routine traffic enforcement as compared to police officers in more urban settings.

Table 4 shows the number of agency-window observations included in this analysis, and the black and Hispanic search rate ratios compared to white drivers. As in every other section of this paper, in each agency-window, at least 10,000 stops were made, with at least 100 stops of white drivers, black drivers, and Hispanic drivers.

Table 4. Summary of Agencies Included in Analysis

Type	N	Black - White SRR	Hispanic - White SRR
State Agency	48	1.98	2.39
County Sheriff	211	2.28	2.85
Police Department	1,447	2.41	3.15

Note: Only agency-windows meeting all thresholds included in analysis.

Table 4 shows that police departments on average have the highest search rate ratios, while state agencies have the lowest. However, this does not tell us whether the differences are statistically meaningful. To do this, we conduct three t-tests: the first compares police departments with state agencies; the second compares police departments with county sheriffs; and the last compares county sheriffs with state agencies. The summary of this analysis is presented in Table 5.

Table 5. Comparison of Black - White Search Rate Ratios by Type of Agency

Type 1	Type 2	SRR of Type 1	SRR of Type 2	Difference
Police Department	State Agency	2.41	1.98	0.43 **
Police Department	County Sherriff	2.41	2.28	0.13
County Sherriff	State Agency	2.28	1.98	0.30 **

Note: ** indicates statistical significance at the 0.05 level. * indicates statistical significance at the 0.10 level

Police departments and county sheriff's offices have higher Black - White search rate ratios than state agencies. Both differences are statistically significant at the 0.05 level. The first is a difference of 0.43, and the second is a difference of 0.30. While, these comparisons adhere to expectations, the comparison of police departments and county sheriffs do not. The difference is in the expected direction, but not statistically significance. Additionally, while we see that state agencies have on average lower search rate ratios, on average it is still a value of almost 2. Black drivers are on average searched at a rate nearly double of that of white drivers.

Table 6 replicates the test for Hispanic - White search rate ratios by agency type.

Table 6. Comparison of Hispanic - White Search Rate Ratios by Type of Agency

Type 1	Type 2	SRR of Type 1	SRR of Type 2	Difference
Police Department	State Wide	3.15	2.39	0.76 **
Police Department	County Sherriff	3.15	2.85	0.30 *
County Sherriff	State Wide	2.85	2.39	0.46 *

Note: ** indicates statistical significance at the 0.05 level. * indicates statistical significance at the 0.10 level

Police departments and county sheriff's offices have higher Hispanic - White search rate ratios than state agencies. The first is statistically significant at the 0.05 level, while the second is statistically significant at the 0.10 level. Note that here the "lower" value is still above 2: even state agencies have extremely large disparities, but these grow even larger when moving to sheriffs and then on to police departments, where the black –white ratio is above 2.4 and the Hispanic – white ratio exceeds 3.0.

Conclusion

Racial disparities in the likelihood of search are large, ubiquitous throughout the United States, and easily measured. They can be robustly estimated with a simple search-rate ratio with little fear that a more complex multivariate regression will show different patterns. State-wide agencies focused mostly on traffic control, county sheriff's offices, and municipal police departments differ in statistically significant ways with regards to the degree of racial disparities in the outcomes of their traffic stops, but they do not differ in the direction of these disparities. White drivers consistently face much lower odds of search.

Appendix

Aggregating the Data

This appendix explains how this paper built its macro level observations from micro (stop) level datasets. Each observation in the data must meet the thresholds laid out in the body of the paper. As a reminder, these are a given observation must have 10,000 stops, where 100 of each white drivers, black drivers, and Hispanic drivers must also be stopped. However, when using solely agency-year dyads, these thresholds resulted in a high level of data loss.

Because most variation in outcome rate ratios and associated odds-ratio coefficients comes from variation between agencies not over time, we combine data across years within agencies until the thresholds are met for that agency. The result is an agency-window rather than agency-year subset of the data. For example, if a given agency did not have more than 10,000 total stops and over 100 stops for each race category in its first year of data (say, 2005), we would combine its 2005 observations to its 2006 observations. If this combination met the thresholds, it constituted its own observation in the macro level dataset, and the process would begin again with 2007. If the 2005 and 2006 combination did not break the threshold then we would combine 2005 and 2006 observation with the 2007 observations, repeating this process until the threshold was met (if the threshold was not met for the combination in the last year data was available for that agency, then the data would be dropped). Table A1 reports the number of agency-year observations that initially met the thresholds, as well as the number of observations derived from the method described above. As the table makes clear, we increased the number of usable observations from 599 to 1,628.

Table A1: Summary of observations for different aggregation methods

	Agency-Year Observations		Agency-Window
State	All Obs.	Obs. Above Thresholds	Obs. Above Thresholds
IL	9,588	338	1,046
MD	574	41	68
CT	314	5	5
NC	3,535	209	503
FL	6	6	6
Total	14,017	599	1,628

This process did create observations with different time boundaries. Table A2 reports the summary statistics for the time frame (calculated as the end year for the observation minus the start year for the observation plus 1) for the macro level observations. The time range for the observations ranging from one year up to 15 years, with an average of 2.86 years.

Table A2: Summary Statistics for the Number of Years Each Observation Spans

	Observations	Mean	Std. Dev.	Minimum	Maximum
Time Frame	1,628	2.86	2.26	1	15

Establishing a Search Threshold

This appendix explains our reasoning on the application of a search threshold for our observations, as well as presenting a short simulation to demonstrate the lack of reliability in observations with low numbers of searches. Our comparison of measures required running logistic regressions on the micro level data associated with a macro level observation, and while each observation had a minimum number of stops, there was no requirement threshold placed on the number of searches an observation needed. Because searches are rare events and rare events have the potential for seriously violating the assumptions underlying GLM, we were concerned that observation with a low number of searches in a particular racial category could create unreliable estimates for that race's coefficient. To address this concern, we imposed a further threshold for the number of searches (not stops) in the logistic regressions. In order to decide on the appropriate threshold to use, we followed this procedure.

We started with a database from five agency with relatively few searches, around 25. We used Lake Forest Police from 2008-2010, Du-Page County Sheriff from 2007-2009, Great Lakes Naval Station from 2005-2009, Illinois Commerce Commission Police from 2005-2010, and Pulaski County Sheriff from 2011-2012. In each, we altered the number of searches in that data set and recorded the effect a diminishing number of searches had on the coefficients. That is, we first ran the regression on the full unaltered dataset and then recorded the coefficients for blacks and Hispanics. Then we changed an observation from a search to no-search, and re-ran the regression, again recording the race coefficients. We did this progressively from 25 down to no searches, in increments of one.

Figure A1 shows the odds-ratios for Blacks and Hispanics as the number of total searches decreases for the Illinois Commerce Commission Police Department. The y-axis shows the associated odds-ratio, and the x-axis shows the total number of searches. The grey line indicates the odds-ratio for Black drivers, and the black line indicates the odds-ratio for Hispanic drivers.

Figure A1: Simulation 1 for the Illinois Commerce Commission Police Department

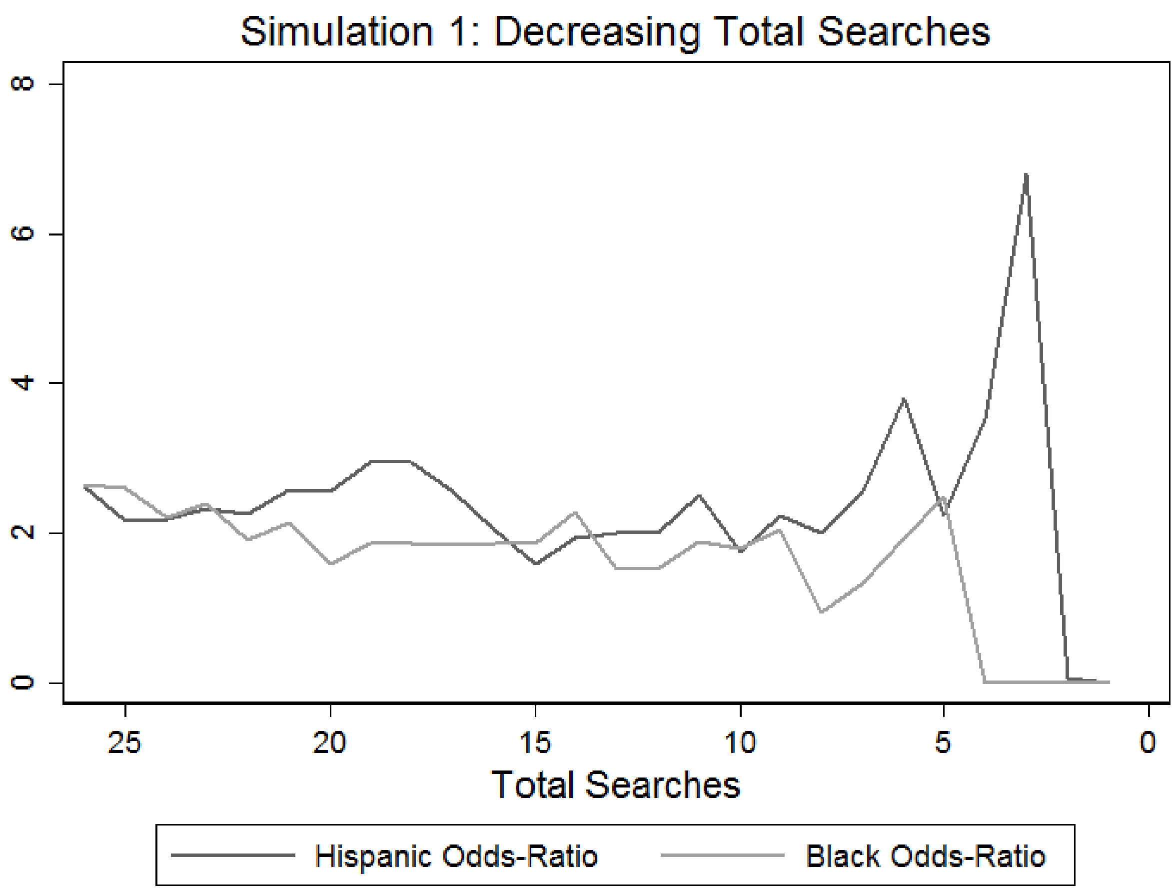


Figure A1 demonstrates the high volatility in logistic regression results when the total number of searches falls below 10. Similar results were found in the other four agencies we simulated. For this reason, we impose a ten-search threshold for all observations, Table A3 reports the data loss associated with this threshold.

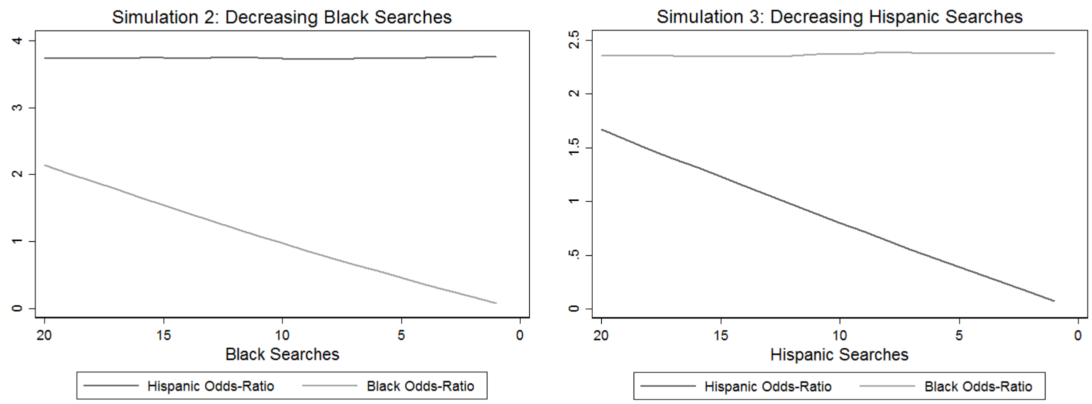
Table A3: Summary of data loss from total search threshold

		N Below		
State	Total Observation	Total Threshold	Black Threshold	Hispanic Threshold
IL	1,074	28	65	56
MD	68	0	0	9
CT	5	0	0	0
NC	535	32	23	2
FL	6	0	0	0
Total	1,688	60	88	67

The first simulation dropped searches randomly regardless of driver race. Given the instability in the estimates we found when an agency had a low number of total searches, we wanted to establish whether a low number racial category could cause instability in the coefficients. Furthermore, we wanted to check if the coefficient for other racial groups could be influenced by this instability. To test these concerns we conducted an additional two simulations, where real data was altered so that a particular racial group's searches would be decreased, and we recorded the odds-ratios from the logistic regression for each altered dataset. For Black drivers the agency-windows used are: Palatine Police from 2012-2013, Antioch Police from 2008-2010, Boone County Sheriff from 2008-2012, Lake Zurich Police from 2009-2010, and Wood Dale Police from 2005-2007. For Hispanic drivers the agency-windows used are: Tinley Park Police from 2013-2014, Whiteside County Sheriff from 2010-2013, Normal Police in 2013, Columbia Police from 2005-2008, Rock Island County Sheriff from 2008-2012. Figure A2 shows the results for the second and third simulation, where the first shows the results of altering the number of Black searches (Figure A2a) and the second shows the results of altering the number of Hispanic searches (Figure A2b). In each of the subfigures, the y-axis shows the odds-ratio coefficient, and the x-axis shows the number of searches of the relevant group. Additionally, the grey line indicates the odds-ratio for Black drivers, and the black line indicates the odds-ratio for Hispanic drivers.

Figure A2: Decreasing the Number of Searches by Race, Odds Ratios

a) Black Searches (Antioch Police Dept.) b) Hispanic Searches (Whiteside Police Dept.)

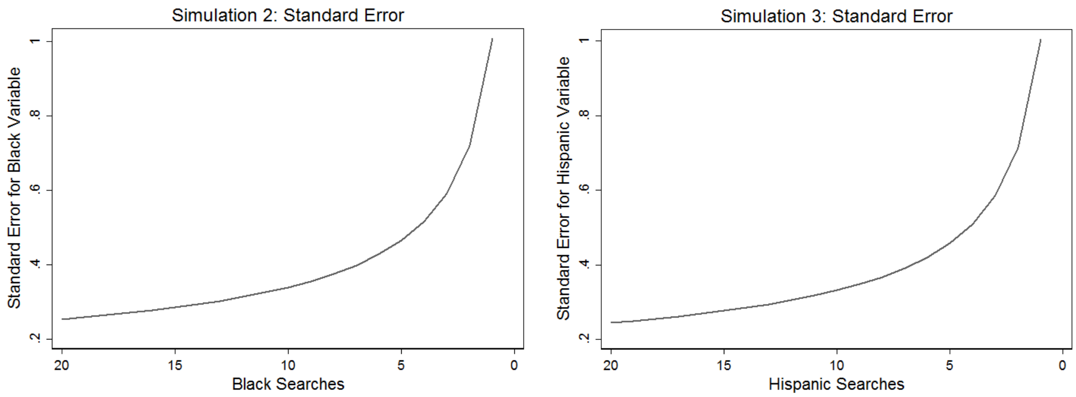


The patterns observed Figure A2 are representative of the findings for all five agencies in their respective simulation. Both figures show the same patterns, a low number of searches for a racial group does not affect the results for other racial groups. Furthermore, decreasing the number of searches for a racial group does not bias the estimate for that group. The downward slope of the line is due to decreasing the number of minority searches, which are being compared against the baseline group in the logistic model, Whites. Therefore, each minority search creates a dataset where minorities are less likely to be searched than their white counterparts are, and this is reflected in the figures. The odds-ratios accurately reflect the declining search rates.

While the odds-ratios were robust to a low number of searches, we discovered in these simulations that observations with low numbers of minority searches were not very efficient due to the lack of variability in the dependent variable for that group. Figure A3 shows the effect a diminishing number of searches have on the standard error for the associated minority group. In Figure A3, the y-axis is the standard error for the coefficient in the regression model for the city. The x-axis once again shows the number of searches of the relevant group.

Figure A3: Decreasing the Number of Searches by Race, Standard Errors

a) Black Searches (Antioch Police Dept.) b) Hispanic Searches (Whiteside Police Dept.)



In Figure A3, we can see that the standard errors appear to be relatively stable until the number of searches gets down to below 10. By the time the number of searches is as low as 5, the standard errors are quite large. Given these results, we imposed a 10-search threshold for each racial group, when considering the odds-ratio for that racial group. Note that if the number of searches of Hispanics was below the threshold, but there were enough blacks and whites to make an estimate, our results indicated that the black odds-ratio was not affected by the lack of Hispanic search observations, and we retained the estimate for blacks.

References

Baumgartner, Frank R., Leah Christiani, Derek A. Epp, Kevin Roach, and Kelsey Shoub. 2017a. Racial Disparities in Traffic Stop Outcomes. *Duke Forum for Law and Social Change* 9: 21-53.

Baumgartner, Frank R., Derek A. Epp, Kelsey Shoub, and Bayard Love. 2017. Targeting Young Men of Color for Search and Arrest during Traffic Stops: Evidence from North Carolina, 2002-2013. *Politics, Groups, and Identities* 5, 1: 107-131.

Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 2007. An analysis of the New York City police department's “stop-and-frisk” policy in the context of claims of racial bias. *Journal of the American Statistical Association* 102.479: 813-823.

Simoiu, Camelia, Sam Corbett-Davies, and Sharad Goel. Forthcoming. The problem of infra-marginality in outcome tests for discrimination. *Annals of Applied Statistics*.

United States Department of Justice (USDOJ), Civil Rights Division. 2015. *Investigation of the Ferguson Police Department.* Washington: US DOJ. March 4.