Online Appendix for: Comparing Systemic and Individual Sources of Racially Disparate Traffic Stop Outcomes

Kelsey Shoub, University of South Carolina*

July 3, 2021

Abstract

The deaths of George Floyd and Breonna Taylor brought discussions of race and policing to the forefront in the summer of 2020 in the United States, spurring protests and calls for policing reform. However, enacting successful reforms curtailing racially biased policing requires understanding whether bias is widespread, likely tied to systemic sources, or the work of a few racist officers. This study elaborates on these perspectives by drawing on theories of systemic and individual bias that may arise when bureaucrats have ample discretion. Using millions of traffic stops, I construct two measures to evaluate whether disparities are widespread — as indicated by the first perspective — or localized — as suggested by the second. This study finds widespread racial disparities in who is searched following a traffic stop across officers and agencies and that banning consent searches would alter policing patterns more than eliminating outlier officers.

*KSHOUB[at]mailbox.sc.edu

Contents

(A) Building the Data Sets

(B) Threshold Test Model Convergence

(C) Alternative Stop Threshold

(D) Alternative Identification of Highly Disparate Officers

1 Building the Traffic Stops Data Sets

I draw on two data sets made publicly available by each state.[1] Here I discuss each in more detail, such as why these data are collected and made public, what information is contained within them, provide additional information on what traffic stops look like in each state, and how these data are cleaned. Additionally, I further address how to access these data sources from their respective primary sources. For the cleaned and collapsed versions of each data set used in the analysis, please see the associated data verse.

1.1 Connecticut

In 1999, the state of Connecticut passed an anti-racial profiling law (the Alvin W. Penn Racial Profiling Prohibition Act) that prohibits profiling in policing (e.g., stopping, searching, arresting motorists) based on race, gender, age, or ethnicity of that individual. Then, in 2012 and 2013, the state legislature made several changes to the law to create a system to address concerns in Connecticut. Simultaneously, the Racial Profiling Prohibition Project Advisory Board to advise the Office of Policy and Management to help oversee the implementation of policy changes and evaluation of those changes. In addition to these groups, the Institute for Municipal and Regional Policy at Central Connecticut State University was also tasked with policy design, evaluation, and management of the new law. Based on recommendations from the Racial Profiling Prohibition Project, police agencies began collecting and submitting traffic stop data in a new way beginning October 1, 2013. This data is now made publicly available via the Racial Profiling Prohibition Project's collaboration with the CT Data Collaborative[2] and a more limited version is made available via the state's open data portal.[3] New stops are supposed to be uploaded each month to the overarching project and are made public in full, one year increments.

For this project, I downloaded each of the data releases that contain a full year's worth of data. I then append these data sets and clean them. While the names of the

[1] For an overview of state laws and some city ordinances concerning the collection of traffic stop data, see ?. For more information and background concerning Connecticut, see: `http://www.ctrp3.org`. For more information and background concerning North Carolina, see: `https://trafficstops.ncsbi.gov/`.

[2] `http://trafficstops.ctdata.org/`

[3] `https://data.ct.gov/Public-Safety/Traffic-Stops-Racial-Profiling-Prohibition-Project/nahi-zqrt`

agencies reporting are recorded, the agency type (i.e., municipal police department, sheriff's department, state agency, etc.) is not. To identify what type of agency made each stop, I use dictionaries applied to the agency names and retain stops made by municipal police departments and sheriff's departments. In addition to identifying the agency type, I clean the data further to identify driver race and what happens during the stop. With regards to driver race and ethnicity, I narrow the focus of this study to White non-Hispanic drivers and Black non-Hispanic drivers and generate a new variable indicating whether a driver belongs to one of these two group and missing otherwise. Additionally, with regards to searches and contraband, I generate a dichotomous indicator for whether a search took place and whether contraband is found.

These data are then aggregated to create three new data sets: (1) aggregated by officer; (2) aggregated by agency including all stops; and (3) aggregated by agency, with stops made by highly disparate, outlier officers excluded.

1.2 North Carolina

The North Carolina state legislature passed a law in 1999 mandating the universal recording and reporting of traffic stops by highway patrol officers. This was expanded two years later, such that almost all law enforcement agencies in North Carolina that conduct stops must record (via a standardized form) and report them to the North Carolina State Bureau of Investigation. For more information and background, see **?**. To access the raw data from the source, one must e-mail and request both credentials to download the data (i.e., a username and password) and instructions on how to access the SFTP server from the North Carolina State Bureau of Investigation. These datasets are regularly updated to contain up to date information. I also make available the files we used to conduct this analysis on our dataverse.

For standard traffic stops and for drivers, officers are required to record information on characteristics for all drivers and for all stops. However, for check point stops, information is only required to be recorded and provided if a search is conducted, and passenger information is only (sometimes) recorded if a search is conducted. As such, inclusion, of passengers or stops made as a part of a check-point are excluded as they would bias the analysis.

As with the Connecticut data set, there is not an indicator for what type of agency the officer who made the stop belongs to, but the data set does include the agency name. As such, we once again identify sheriff's departments, municipal police departments, state agencies, and other types of agencies using dictionaries applied to the agency name (e.g., "Alamance County Sheriff's Office"). Additionally, to match the time span available in Connecticut, the North Carolina data set is cropped to include only 6 years of data: January 1, 2014 thru December 31, 2019.

In addition to identifying the agency type and the stops occurring within the designated window, I clean the data further to identify driver race and what happens during the stop. As with the Connecticut data sets, I generate a new race-ethnicity variable indicating whether a drivers is white non-Hispanic drivers or black non-Hispanic drivers and missing otherwise. With regards to searches, we exclude stops that involve mandatory searches (e.g., executing a search warrant) as officers do not have discretion in these case. This means that the types of searches included are: probable cause searches and consent searches. Finally, we generate an indicator for whether any contraband is found.

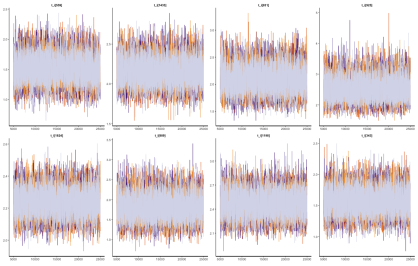
As with the Connecticut data, these data are then aggregated to create three new data sets: (1) aggregated by officer; (2) aggregated by agency including all stops; and (3) aggregated by agency, with stops made by highly disparate, outlier officers excluded.

2 Threshold Test Model Convergence

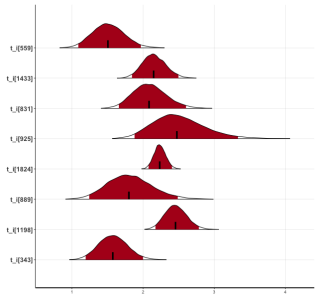
Did the threshold test models converge? To evaluate this, I examine the traceplots, posterior distributions, and the $\hat{R}$. Figure 1 shows a snapshot of the traceplots and posterior distributions for the Connecticut model, while Figure 2 shows the same for the North Carolina. In each figure, a random sample of inferred thresholds is shown for ease of interpretability, as there are hundreds — or thousands – of estimates. The subfigures in the left column of each figure show the traceplots. If the model has stabilized, then these plots should look like random noise – that is what is seen. The right subfigures in the right column of each figure show the distribution of that same sample of inferred search thresholds. As can be seen, they are approximately normal (i.e., look like a bell curve). Finally, the $\hat{R}$ is less than 1.01 for all parameters, which again indicates that our models converged.

In sum, for both Connecticut and North Carolina, the models are stable and the distributions around point estimates are approximately normal.

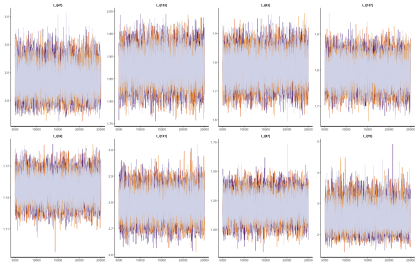
Figure 1: Examining Convergence for the Connecticut Models



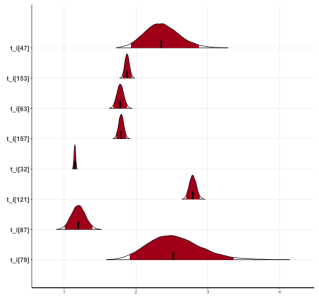
(a) Traceplot: Officers



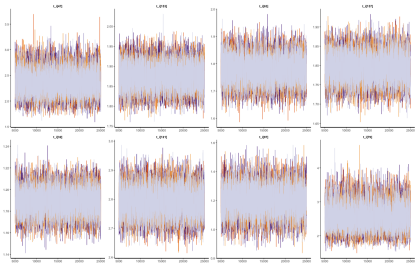
(b) Distribution: Officers



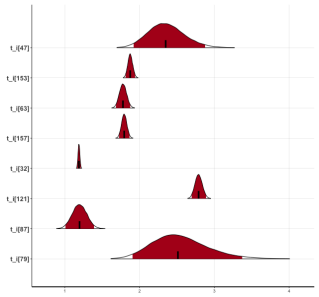
(c) Traceplot: Agencies, Full



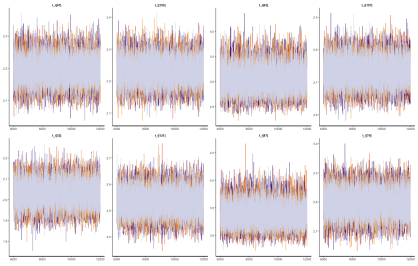
(d) Distribution: Agencies, Full



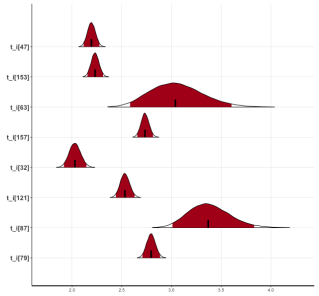
(e) Traceplot: Agencies, No Officer Outlier



(f) Distribution: Agencies, No Officer Outliers

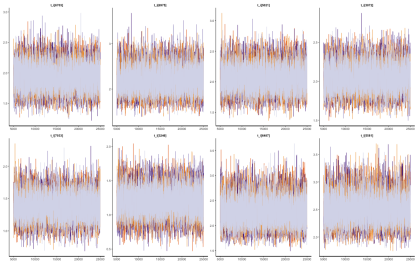


(g) Traceplot: Agencies, No Consent Searches

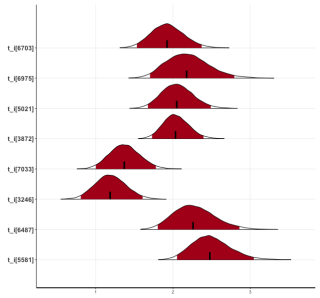


(h) Distribution: Agencies, No Consent Searches

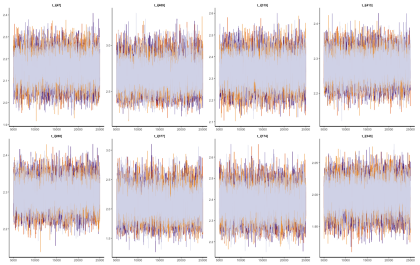
Figure 2: Examining Convergence for the North Carolina Models



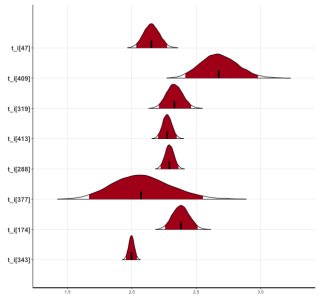
(a) Traceplot: Officers



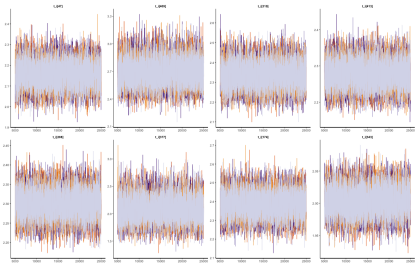
(b) Distribution: Officers



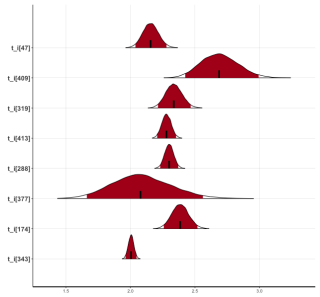
(c) Traceplot: Agencies, Full



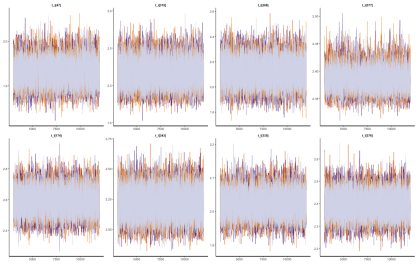
(d) Distribution: Agencies, Full



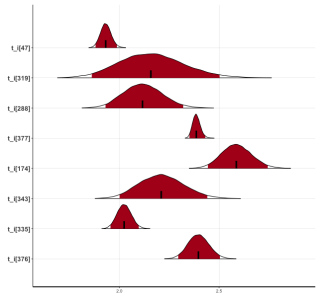
(e) Traceplot: Agencies, No Officer Outlier



(f) Distribution: Agencies, No Officer Outliers



(g) Traceplot: Agencies, No Consent Searches



(h) Distribution: Agencies, No Consent Searches

3 Alternative Stop Thresholds

As discussed in the body of the paper, thresholds for inclusion in the analysis are necessary to ensure the the models converge and estimates are robust. One worry is that the chosen threshold of 50 stops of White drivers, 50 stops of Black drivers, and at least one search of a driver of each race are too low. To address this concern, I replicate the central figures and table from the analysis presented in the paper with a higher stop threshold to 100 stops of White drivers and 100 stops of Black drivers. I do not do this for the final analysis as almost all agencies are included when the threshold is raised to 100 stops.

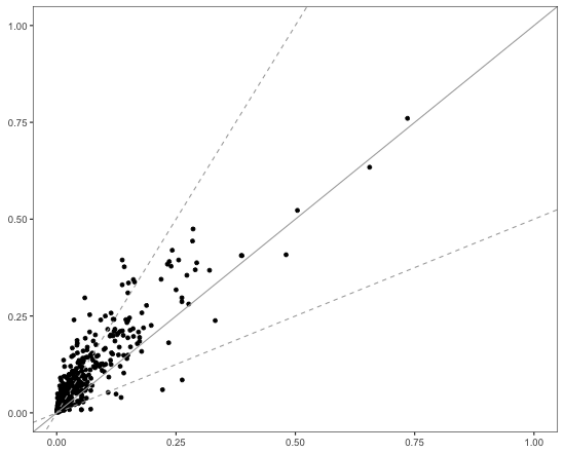
The replicated officer analysis is shown in Figure 3. The Connecticut subfigures show 426 officers from 55 agencies that made 366,231 stops and 21,332 searches. Note that 93 officers would be added if these measures could be calculated without requiring that at least one search occurs. The North Carolina subfigures show 2200 officers from 139 agencies that made 1,777,618 stops and 64,928 searches. Note that 3224 officers would be added if these measures could be calculated without requiring that at least one search occurs. As can be seen, the same patterns are seen: a majority of officers search Black drivers at a higher rate than White drivers and have a lower search threshold for Black drivers than White drivers.

The replicated agency analysis is shown in Figure 4, which replicates the general agency results, and Table 1, which replicates the differnce of means tests. The Connecticut subfigures show 87 agencies of which 15 included at least one outlier officer that made 1,364,601 stops and 57,640 searches. The North Carolina subfigures show 209 agencies of which 28 included at least one outlier officer.Once again, the same patterns are observed here as in the main paper: a majority of agencies search Black drivers at a higher rate and have a lower search threshold for Black drivers compared with White drivers.

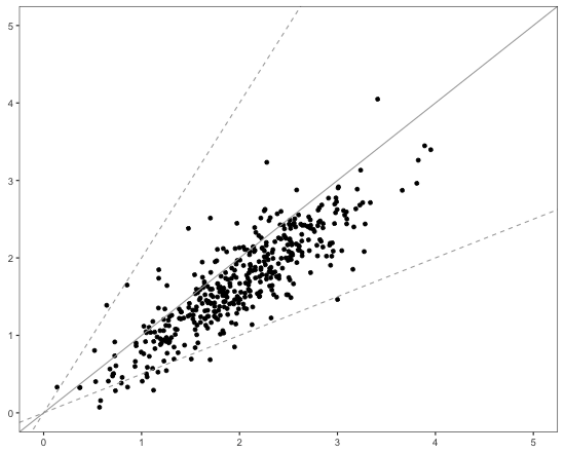
Similarly, the results of the test of differences between the calculated inferred thresholds of agencies that include outlier officers and those that do not remain the same (Table 1). Again, the overall search thresholds for both Black and White drivers for both states increase if outlier officers are excluded, and the threshold ratios and differences of the differences decrease. However, again, the movement in the threshold ratios is relatively small.

In sum, altering the threshold does not change the statistical or substantive results: Black drivers are searched at higher rates than White drivers, and agencies on average have lower search thresholds for Black drivers compared with White drivers. These results are seen across officers from different agencies and states and across agencies from different states.

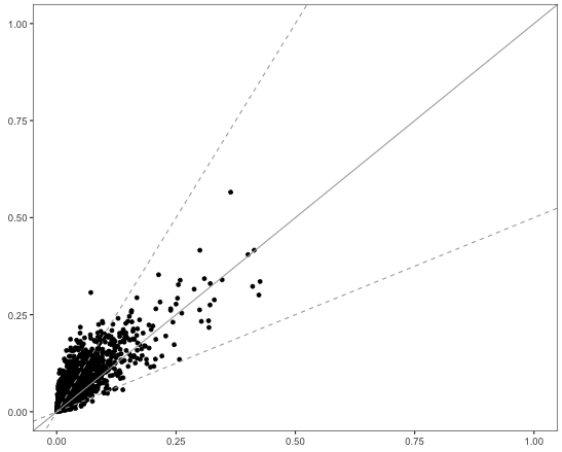
Figure 3: Distribution of (Municipal Police Department) Officer Search Rates and Thresholds by Driver Race and State (100 Stop Threshold)



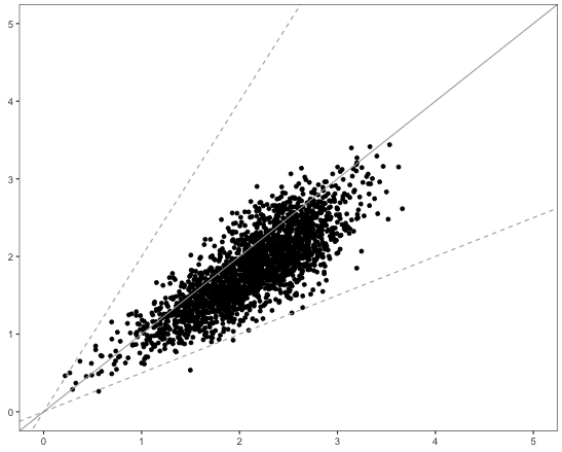
(a) Connecticut, Search Rates



(b) Connecticut, Thresholds



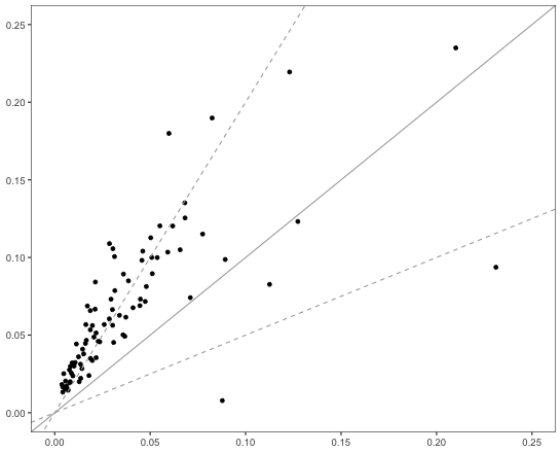
(c) North Carolina, Search Rates



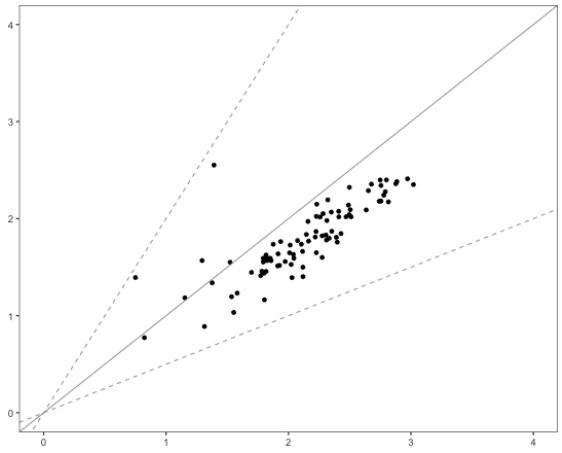
(d) North Carolina, Thresholds

Note: The same thresholds are used for both states and both measures. In all figures, the solid, gray, diagonal line indicates equality, while the dashed grey lines indicate a 2:1 search rate or threshold. In the subfigures showing search rates, dots above the line indicate that Black drivers are searched more frequently than White drivers indicating Black drivers experience disparately negative outcomes, while dots below the solid black line indicate the reverse. In the subfigures showing search thresholds, dots below the line indicate that officers employ lower search thresholds for Black drivers indicating Black drivers experience disparately negative outcomes, while dots above the solid, black line indicate the reverse. In the Connecticut subfigures, 426 officers are included, of which 88% search Black drivers at a higher rate and 88% have a lower search threshold for Black drivers than for White drivers. In the North Carolina subfigures, 2200 officers are included, of which 83% search Black drivers at a higher rate and 82% have a lower search threshold for Black drivers than for White drivers.

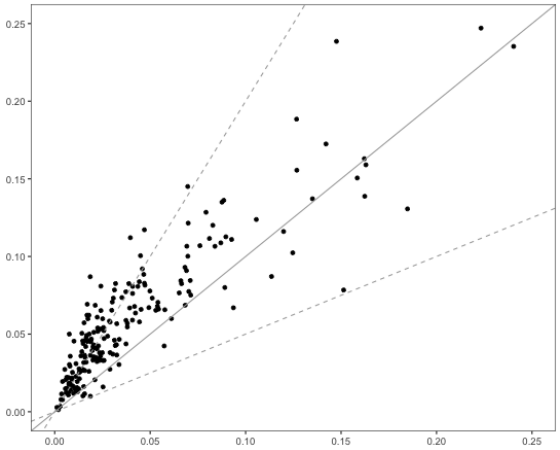
Figure 4: Agency Search Rates and Thresholds with Highly Disparate Officers Excluded



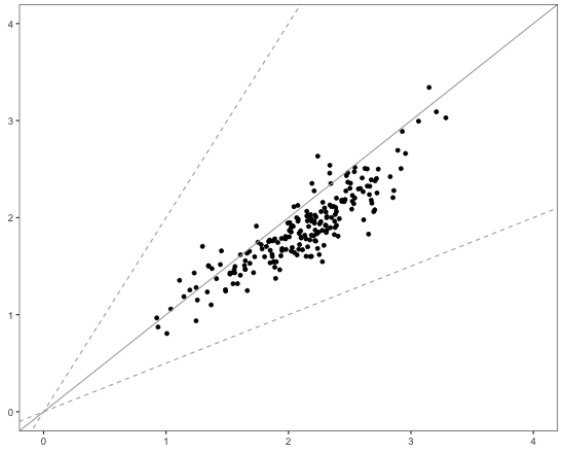
(a) Connecticut, Search Rates



(b) Connecticut, Thresholds



(c) North Carolina, Search Rates



(d) North Carolina, Thresholds

Note: The same thresholds are used for both states and both measures, and outlier officers are detected and excluded in the same manner. In all figures, the solid, gray, diagonal line indicates equality, while the dashed grey lines indicate a 2:1 search rate or threshold. In the subfigures showing search rates, dots above the line indicate that Black drivers are searched more frequently than White drivers indicating Black drivers experience disparately negative outcomes, while dots below the solid black line indicate the reverse. In the subfigures showing search thresholds, dots below the line indicate that officers employ lower search thresholds for Black drivers indicating Black drivers experience disparately negative outcomes, while dots above the solid, black line indicate the reverse. In the Connecticut subfigures, 87 agencies are included with 15 agencies seeing at least one officer being excluded. Of these agencies 95% search Black drivers at a higher rate, and 94% have a lower search threshold for Black drivers than for White drivers. In the North Carolina subfigures, 209 agencies are included with 28 agencies seeing at least one officer being excluded. Of these agencies 90% search Black drivers at a higher rate, and 88% have a lower search threshold for Black drivers than for White drivers.

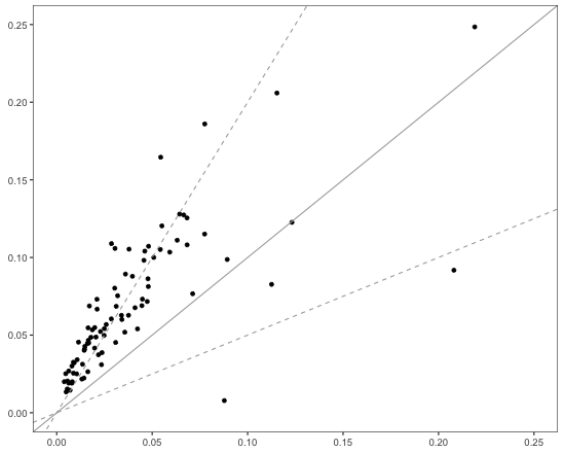
Table 1: Testing the Change in Inferred Search Thresholds When Outlier Officers are and are not Included in Agency Measures

	Includes All Stops	Excludes Disparate Officer Stops	Difference
Connecticut			
Ratio	1.25	1.21	0.04**
Black Drivers	1.35	1.41	-0.06**
White Drivers	1.67	1.70	-0.03**
Difference-in-Difference			0.03*
North Carolina			
Ratio	1.21	1.19	0.02*
Black Drivers	1.78	1.83	-0.05**
White Drivers	2.15	2.18	-0.03**
Difference-in-Difference			0.02*

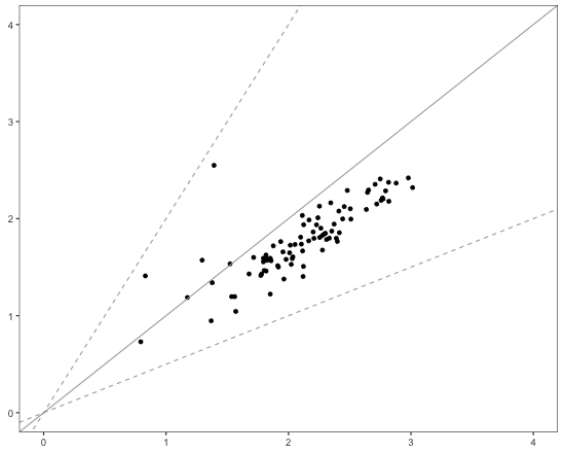
*Note: All tests are paired t-tests. ** indicates $p < 0.05$; * indicates $p < 0.10$. There are 15agencies included in the Connecticut analysis, and 28 agencies included in the North Carolina analysis. Ratios are the White search threshold divided by the Black search threshold. The difference-in-difference for each state compares the change in threshold for Black drivers to that of White drivers.*

4 Alternative Identification of Highly Disparate Officers

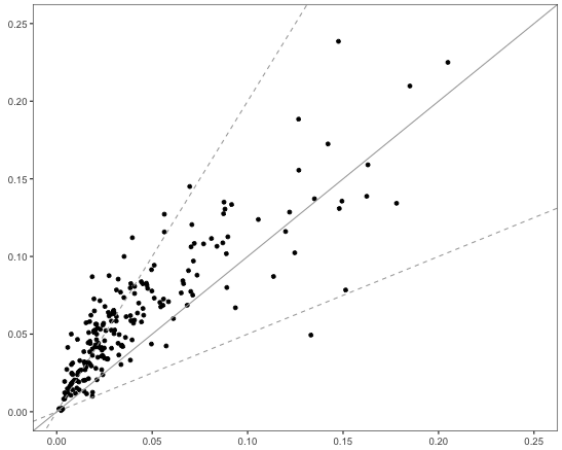
Figure 5: Agency Search Rates and Thresholds with Highly Disparate Officers Excluded



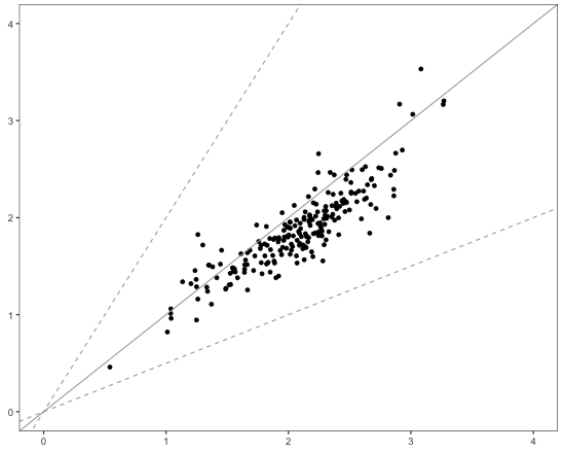
(a) Connecticut, Search Rates



(b) Connecticut, Thresholds



(c) North Carolina, Search Rates



(d) North Carolina, Thresholds

Note: The same thresholds are used for both states and both measures. Both outlier officers and officers only searching drivers of one race are excluded. In all figures, the solid, gray, diagonal line indicates equality, while the dashed grey lines indicate a 2:1 search rate or threshold. In the subfigures showing search rates, dots above the line indicate that Black drivers are searched more frequently than White drivers indicating Black drivers experience disparately negative outcomes, while dots below the solid black line indicate the reverse. In the subfigures showing search thresholds, dots below the line indicate that officers employ lower search thresholds for Black drivers indicating Black drivers experience disparately negative outcomes, while dots above the solid, black line indicate the reverse. In the Connecticut subfigures, 88 agencies are included with 51 agencies seeing at least one officer being excluded. Of these agencies 95% search Black drivers at a higher rate, and 94% have a lower search threshold for Black drivers than for White drivers. In the North Carolina subfigures, 217 agencies are included with 127 agencies seeing at least one officer being excluded. Of these agencies 88% search Black drivers at a higher rate, and 87% have a lower search threshold for Black drivers than for White drivers.

Table 2: Testing the Change in Inferred Search Thresholds When Outlier Officers are and are not Included in Agency Measures

	Includes All Stops	Excludes Disparate Officer Stops	Difference
Connecticut			
Ratio	1.21	1.20	0.01*
Black Drivers	1.768	1.768	<0.01
White Drivers	2.11	2.10	0.008
Difference-in-Difference			0.008
North Carolina			
Ratio	1.133	1.126	0.007
Black Drivers	1.94	1.93	0.01
White Drivers	2.19	2.16	0.03**
Difference-in-Difference			0.02**

*Note: All tests are paired t-tests. ** indicates $p < 0.05$; * indicates $p < 0.10$. There are 51 agencies included in the Connecticut analysis, and 127 agencies included in the North Carolina analysis. Ratios are the White search threshold divided by the Black search threshold. The difference-in-difference for each state compares the change in threshold for Black drivers to that of White drivers.*