Northeastern University
Community to Community
Policy Equity for All

Research Report 2023-02


AP Photo/Jason DeCrow, File. Dec 17, 2012.

New Jersey State Police Traffic Stops Analysis 2009-21

Northeastern University
School of Public Policy
and Urban Affairs

October 2023

The **School of Public Policy and Urban Affairs** at Northeastern University is committed to interdisciplinary thinking and problem solving on issues of pressing concern at the local, state, national, and international levels – from transportation, housing, energy and community development to social and environmental justice, resilience and sustainability. By educating students in both theory and practice, the School prepares them to become effective contributors to social, economic, and environmental change, in careers spanning the public, private, and nonprofit sectors, as well as academic research.

Community to Community (C2C) is an impact accelerator at Northeastern University designed to deepen the university's commitment to community engaged research at each of its global campus locations. We strive to move the needle on societal problems at the local level, benefitting the areas our university calls home, while also promoting knowledge transfer across communities that are grounded in the local context.

New Jersey State Police Traffic Stops Analysis, 2009-21

Author
Mathew B. Ross, Ph.D.

School of Public Policy & Urban Affairs
Department of Economics
Northeastern University

Table of Contents

Introduction

In November 2021, the New Jersey Attorney General's Office of Public Integrity and Accountability (NJ-OPIA) engaged the author of this study for the purpose of conducting an independent analysis of traffic stops made by the New Jersey State Police (NJ-SP). Based on the author's extensive experience working with state and local policymakers to develop early warning systems for identifying police disparities, the NJ-OPIA requested that the analysis focus on the central question of whether there was disparate treatment on the part of NJ-SP towards marginalized racial and ethnic groups.[1] After cleaning and linking all of the raw data provided by the New Jersey Office of Law Enforcement Professional Standards (NJ-OLEPS), the analytical sample used in this analysis consisted of 6,177,109 traffic stops made by NJ-SP from 2009 to 2021. In the full analytical sample, 60.52 percent of traffic stops were made of White, non-Hispanic motorists while 18.8 percent were Black/African American and 13.44 percent were Hispanic/Latinx. The overall volume of non-White motorists stopped by NJ-SP increased from 35.34% in 2009 to 46.28% in 2021.

The overarching finding from the analysis of the NJ-SP data from 2009-21 is that there was extremely strong evidence of a large and persistent disparity both in the decision to stop as well as the decision to engage in post-stop enforcement like search, vehicular exits, use of force, and arrest. In general, the results were estimated with a very high degree of statistical confidence, survived multiple robustness tests, and were found across most years and troops/stations. In the opinion of this study's author, these disparities represent strong empirical evidence that NJ-SP is engaged in enforcement practices that result in adverse treatment towards motorists of color.

Following best practices, this study applies an ensemble of the most reliable statistical tests available in the scientific literature. The intuition of this approach is that the shortcomings of any individual test are overcome by the totality of the evidence produced by a multitude of tests examining a broad set of enforcement outcomes.

The statistical tests and associated findings for 2009-21 included the following[2]:

1. Analysis of the Decision to Stop: We apply a solar visibility test that leverages variation in the timing of the sunset throughout the year under the premise that police are marginally better able to detect race/ethnicity during periods of daylight relative to darkness.
   i. Black/African American motorists were 2 percentage points (9.3 percent) and Hispanic/Latinx motorists were 2.9 percentage points (16.1 percent) more likely to have been stopped during periods when their race was more easily visible.

2. Analysis of the Decision to Search: We apply a hit-rate test comparing the likelihood of a vehicular search to yield contraband as well as a test that compares the likelihood of a search occurring conditional on the motivating reasons for the traffic stop.
   i. Conditional on the motivating reason for being stopped, Black/African American motorists were 2.9 percentage points (89.8 percent) and Hispanic/Latinx motorists were 1.5

[1] See Ross, Matthew B., Jesse J. Kalinowski, Kenneth Barone. 2020. "Testing for disparities in traffic stops: Best practices from the Connecticut model." Criminology & Public Policy, v19 i4. https://doi.org/10.1111/1745-9133.12528

[2] Note that all relative effects documented below are calculated based on the relevant dependent mean, i.e. the population share of either Black/African American or Hispanic/Latinx motorists. Please see the detailed results for additional details.

percentage points (46.4 percent) more likely to be searched.

ii. Conditional on being searched, Black/African American motorists were 1.4 percentage points (9.7 percent) and Hispanic/Latinx motorists were 3.9 percentage points (26.6 percent) less likely to have evidence found.

3. Analysis of the Decision to Exit Vehicle: We apply a test that compares the likelihood of a search occurring conditional on the motivating reasons for the traffic stop.
   i. Conditional on the motivating reason for being stopped, Black/African American motorists were 2.8 percentage points (14.65 percent) and Hispanic/Latinx motorists were 1.7 percentage points (9.6 percent) more likely to be asked to exit their vehicle.

4. Analysis of Arrests and Use of Force: We apply a test that compares the likelihood of a search occurring conditional on the motivating reasons for the traffic stop.
   i. Conditional on the motivating reason for being stopped, Black/African American motorists were 2.9 percentage points (87.5 percent) and Hispanic/Latinx motorists were 1.5 percentage points (56.8 percent) more likely to be arrested.
   ii. Conditional on the motivating reason for being stopped, Black/African American motorists were 0.013 percentage points (130 percent) and Hispanic/Latinx motorists were 0.003 percentage points (27.5 percent) more likely to experience force.

In the proceeding sections, we provide a detailed methodological discussion of each individual test for the aggregate period from 2009-21 as well as by individual year and troop/station. The empirical appendix contains additional ancillary results as well as a series of robustness tests that impose an additional layer of modeling or sample restrictions.

Analysis of the Decision to Stop

The canonical challenge to assessing whether there are racial and ethnic disparities in the decision by police to stop a motorist is the lack of an appropriate counterfactual. Put simply, reliable data on the demographic composition of motorists is nonexistent and so there is not an available benchmark for evaluating police traffic stops. To overcome this challenge, we implement a solar visibility analysis following Grogger and Ridgeway (2006). A solar visibility analysis compares the likelihood that a traffic stop is made of a person of color during daylight relative to darkness (see also Ridgeway 2009; Horrace and Rohlin 2019; Kalinowski et al. 2018, 2020, 2022). Ross et al. (2020) cites 18 cities and four states that have relied on a solar visibility analysis to evaluate the decision to stop a motorist and describe the test as being widely considered by practitioners and policymakers as "best practice".[3] The authors demonstrate that, under a certain set of conditions, a change from daylight to darkness in the odds of a stopped motorist being a person of color is equivalent to a change in the odds a person of color is stopped. Under the assumption that the only thing changing between daylight and darkness is the ability of police officers to detect race prior to making a traffic stop, an increase in the likelihood that a non-White motorist is stopped during daylight is indicative of disparate treatment and possible discrimination.

Using data on 6,177,109 traffic stops made by the New Jersey State Police from January 2009 to May 2021, we apply a solar visibility analysis to assess the extent of racial and ethnic disparities in the decision to stop a motorist. To account for the fact that enforcement activity and the underlying composition of motorists may be different at different times of the day, we restrict the analytical sample to traffic stops falling within a window of time when sunset varies seasonally. In particular, we focus on 1,085,980 traffic stops occurring within the "inter-twilight window" between the earliest sunset of the year (approximately 4:30 pm) and the latest end to civil twilight (approximately 9:00 pm). Since some infractions (lighting, seatbelt, and cellphone violations) are correlated with visibility and (possibly) race (via socioeconomic factors), we further restrict the analytical sample to 663,916 moving violations (see Ridgeway 2009). In order to control for potential variation across time and geographic space, we will utilize regression analysis and control for things like time of day, day of week, and year by geographic location or individual officer (see Horrace and Rohlin 2019; Kalinowski et al. 2018). To address potential concerns of endogeneity associated with seasonal differences in driving and enforcement, we will also implement a robustness test focusing on a narrow window of time before/after the spring/fall daylight savings time shift using a regression discontinuity design (see Kalinowski et al. 2020).

Formally, we estimate a linear probability model of the form

$$1[minority_i] = \alpha + \beta\, 1[daylight_i] + dow_i + tod_i + (badge_i * year_i) + \mu_i \quad (1)$$

[3] Applications of the test include Grogger and Ridgeway (2006) in Oakland, CA; Ridgeway (2009) Cincinnati, OH; Ritter and Bael (2009) and Ritter (2017) in Minneapolis, MN; Worden, McLean, and Wheeler (2010,2012) as well as Horace and Rohlin (2016) in Syracuse, NY; Renauer, Henning, and Covelli (2009) in Portland, OR; Taniguchi et al. (2016a,2016b,2016c,2016d) in Durham Greens-boro, Raleigh, and Fayetteville, North Carolina; Masher (2016) in New Orleans, LA; Chanin et al. (2016)in San Diego, CA; Criminal Justice Policy Research Institute (2017) in Corvallis PD, OR; Milyo (2017) in Columbia, MO; Smith et al. (2017) in San Jose, CA; and Wallace et al. (2017) in Maricopa, AZ. Statewide studies relying on this test include Ross, Fazzalaro, Barone, and Kalinowski (2015, 2016, 2017a, 2017b, 2018, 2019a, 2019b) in Connecticut and Rhode Island, Racial & Identity Profiling Advisory Board (2020) in California, and Sanchagrin et al. (2019) in Oregon.

where the dependent variable $1[minority_i]$ is an indicator which is equal to one if traffic stop *i* was made of a person of color and zero otherwise. The primary independent variable $1[daylight_i]$ is an indicator which is equal to one if a traffic stop was made in daylight and zero otherwise. Additional control variables include six indicators for the day of the week (dow_i), 18 indicators for 15-minute time of day increments (tod_i), and 15,412 indicators for each unique officer in each year ($badge_i * year_i$). We note that stops occurring during twilight (neither daylight nor darkness) are dropped from the sample and that each group (i.e. African American/Black, Hispanic/Latinx, and all other race/ethnicity) is only compared against a sample of Caucasian (i.e. non-Hispanic/Latinx) motorists. Across all of the estimates, we cluster standard errors at the officer by year level.

Figure 1 contains a graphical presentation of the results from applying equation (1) to the sample of moving violations made by the New Jersey State Police during the inter-twilight window from 2009-21. Panel (a) presents the difference in the likelihood a Black/African American motorist was stopped in daylight relative to darkness while panel (b) presents the same difference for Hispanic/Latinx motorists. The baseline comparison group in both panels consists of traffic stops made of Caucasian (i.e. White, non-Hispanic/Latinx) motorists. The vertical axis denotes the predicted probability that a traffic stop involved a person of color. The navy-colored bar represents the probability of a stop involving a person of color being made in darkness while the orange-colored bar represents daylight. The annotation in the center of each bar documents the magnitude and statistical significance of the change in the predicted probability a person of color is stopped in daylight relative to darkness.

As shown below in panel (a), we find that Black/African American motorists were 2 percentage points (9.31 percent relative to a dependent mean of 0.22) more likely to be represented in the traffic stop data during daylight relative to darkness. With respect to panel (b), we find that Hispanic/Latinx motorists were 2.9 percentage points (16.1 percent relative to a dependent mean of 0.18) more likely to be represented in the traffic stop data during daylight. The estimate for both groups were found to be highly statistically significant at a confidence level exceeding 99 percent.

The disparity found for both Black/African American and Hispanic/Latinx motorists suggests that the New Jersey State Police are more likely to stop a person of color during periods when they can more easily discern their race/ethnicity which is indicative of potential discrimination.

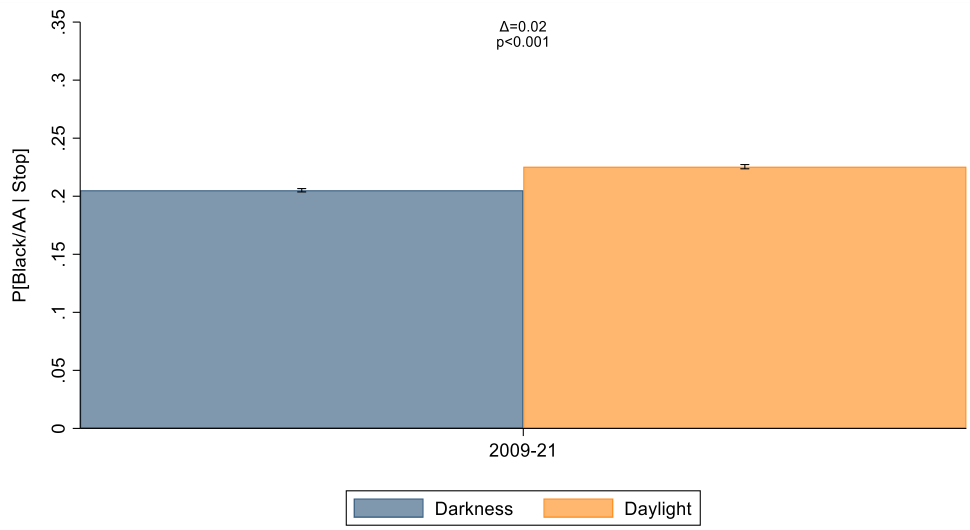
Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.1. Appendix Figure A.3. presents additional regression discontinuity estimates from a more restricted sample of stops within 21 days of the spring/fall daylight savings time shift which show qualitatively similar results.[4]

[4] In this robustness test, we estimate a linear probability model of the form:

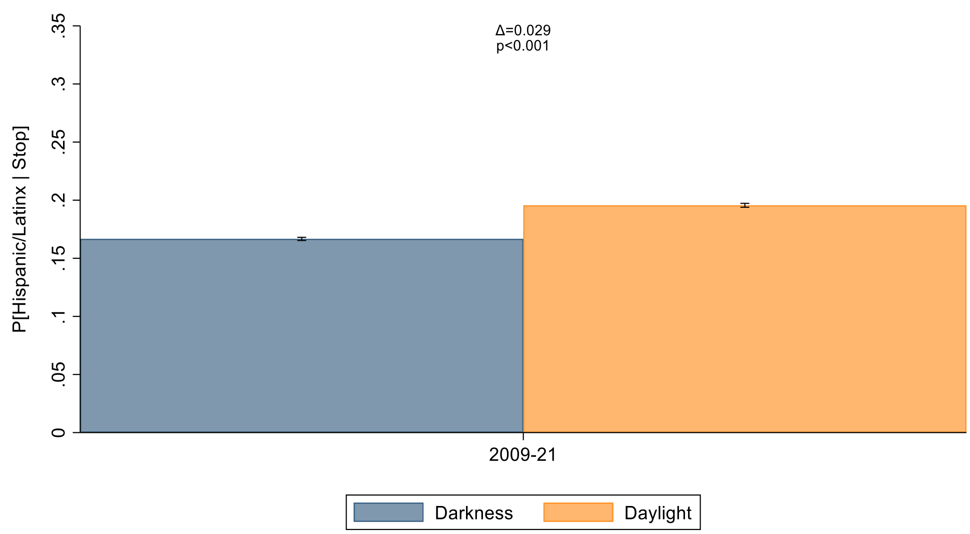
$$1[minority_i] = \alpha + \beta_{dst}1[dst_i] + \beta_{run}(run_i * 1[fall_i]) + \beta_{post}(1[dst_i] * run_i * 1[fall_i]) + dow_i + tod_i + (badge_i * year_i * 1[fall_i]) + \mu_i$$

where the dependent variable $1[minority_i]$ is an indicator which is equal to one if traffic stop *i* was made of a motorist of color and zero otherwise. The primary independent variable $1[dst_i]$ is an indicator which is equal to one if a traffic stop was made in the period after the spring and before the fall daylight savings time shift. The variable run_i is a running variable equal to as little as -21 before the shift to a period of more light and +21 afterwards. We interact this variable with an indicator $1[fall_i]$ for fall vs. spring both alone and in combination with $1[dst_i]$. As before, we include additional control variables include six indicators for the day of the week (dow_i), 18 indicators for 15-minute time of day increments (tod_i), and 30,825 indicators for each unique officer in each year and season ($badge_i * year_i * 1[fall_i]$).

Figure 1. Solar Visibility Analysis Estimates, 2009-21



(b) Black/African American



(a) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White, non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weigh by the inverse number of records per traffic stop. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

Examining heterogeneity across different years, we find that the disparities are remarkably persistent and appear to be growing over time. Figure 2 disaggregates the prior set of results by each individual year from 2009 to 2021. As shown below in panel (a), Black/African American motorists were consistently more likely to be stopped by New Jersey State Police in daylight relative to darkness. The magnitude of the disparity ranged from 0.8 percentage points in 2011 to as much as 3.2 and 5.4 percentage points in 2020 and 2021. In all years but 2011, the disparity was estimated with a confidence level exceeding 98 percent. With respect to panel (b), Hispanic/Latinx motorists were also consistently more likely to be stopped by New Jersey State Police in daylight relative to darkness. The magnitude of the disparity ranged from 2.1 percentage points in 2009 to as much as 5.7 and 3 percentage points in 2020 and 2021. In all years, the disparity was estimated with a confidence level exceeding 99 percent.

Across nearly all years in the sample, we find strong evidence suggesting a large and persistent disparity in the decision to stop a person of color.

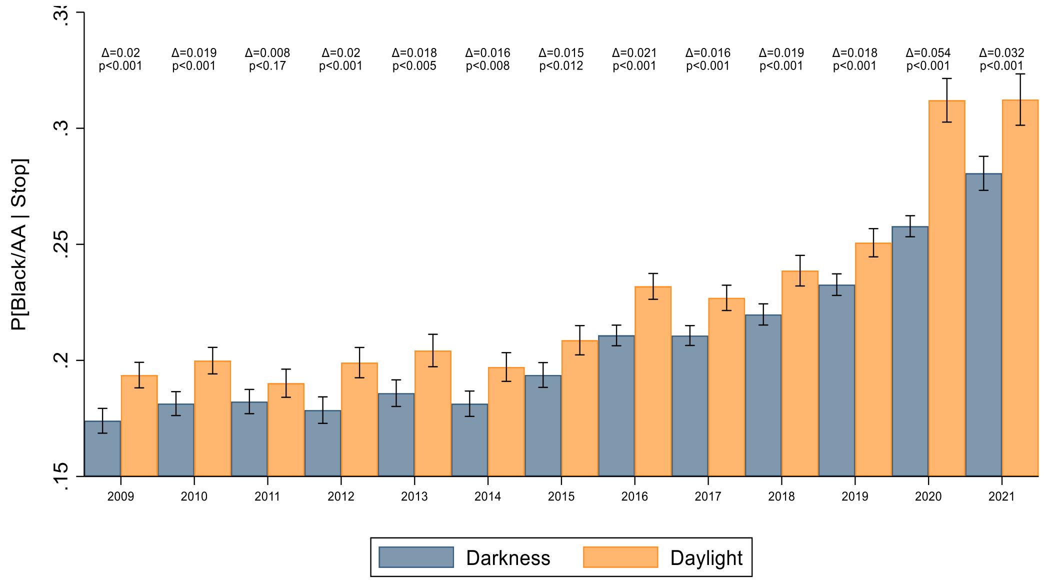
Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.2. Appendix Figure A.4. presents additional year-by-year regression discontinuity estimates from a more restricted sample of stops within 21 days of the spring/fall daylight savings time shift which show qualitatively similar results.

Examining heterogeneity across troops and stations, we also find that the disparities are remarkably consistent. In Troop A, we estimate that Black/African American and Hispanic/Latinx motorists were 2.39 ($p<0.01$) and 0.95 ($p<0.01$) percentage points more likely to be stopped in daylight relative to darkness. In Troop B, we estimate that Black/African American and Hispanic/Latinx motorists were 2.33 ($p<0.01$) and 3.95 ($p<0.01$) percentage points more likely to be stopped in daylight relative to darkness. In Troop C, we estimate that Black/African American and Hispanic/Latinx motorists were 1.4 ($p<0.01$) and 1.9 ($p<0.01$) percentage points more likely to be stopped in daylight relative to darkness. In Troop D, we estimate that Black/African American and Hispanic/Latinx motorists were 3.2 ($p<0.01$) and 1.94 ($p<0.01$) percentage points more likely to be stopped in daylight relative to darkness.

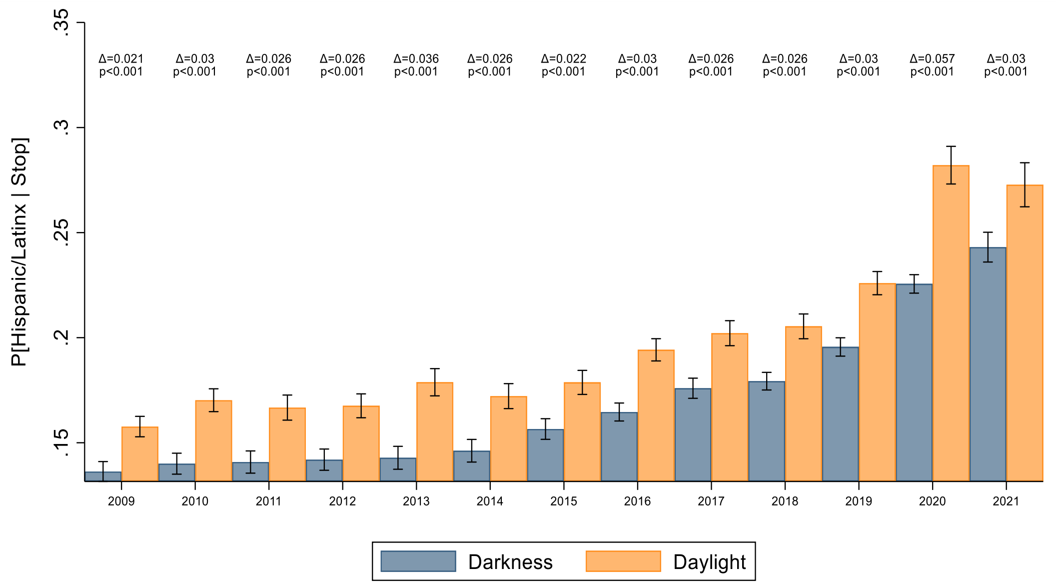
Across nearly all troops in the sample, we find strong evidence suggesting a large and persistent disparity in the decision to stop a person of color.

Appendix Figures B.1-4. contain separate estimates for each individual station within the four overarching troops.

Figure 2. Solar Visibility Analysis Estimates by Year



(b) Black/African American



(a) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White, non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weigh by the inverse number of records per traffic stop. Each regression was estimated using data only for the respective year labeled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

Analysis of the Decision to Search

The challenge of analyzing post-stop enforcement (i.e. search, force, or vehicle exits) for evidence of racial or ethnic disparities is that alternative approaches, which are conditional on observables, may suffer from the well-known "infra-marginality problem." Put simply, disparities in post-stop outcomes might exist due to differences in the distribution of stopped motorists in terms of things observed by police on the scene and not easily observed by analysts from traffic stop data. These unobservable differences are likely to persist even when the researcher controls for a rich set of traffic stop, vehicle, and motorist characteristics. As such, scholars and practitioners have focused on hit-rate style tests following Knowles et al. (2001) as opposed to using an approach that conditions on observables (see also refinements of hit-rate tests by Dharmapala & Ross 2003; Antonovics & Knight 2004; and Anwar & Fang 2006).[5]

Hit-rate tests are motivated by Becker's (1971) model of discrimination where police bias is conceptualized as an officer facing a lower internal cost of engaging in discretionary post-stop enforcement against a person of color relative to a White, non-Hispanic motorist in terms of things like search, force, or vehicle exits. In the absence of disparate treatment and in a world where the police make discretionary post-stop enforcement decisions based on reasonable suspicion or a credible threat, the costs of engaging in enforcement for different groups should be equal. Thus, one should expect the empirical probability of a search yielding contraband (or force or a vehicle exit leading to an arrest) to be equal across racial/ethnic groups even when the guilt rates across these groups differ. Unbiased police officers may engage in discretionary post-stop enforcement against persons of color more often, but only proportional to their likelihood of guilt. If non-White groups face a disproportionate rate of post-stop enforcement relative to their guilt rate, it is indicative that police face a lower cost for engaging in post-stop enforcement and are biased against non-White motorists.[6]

In total, there were 259,652 searches from 2009-21 of which only 6.54 percent resulted in contraband being found. We focus on discretionary searches since there are many that occur incidentally to an arrest, as a result of plain view contraband, or as part of inventorying an impounded vehicle. In particular, we focus on the subset of 18,588 searches where we were able to identify that the officer documented reasonable articulable suspicion or requested consent.[7] Figure 3 presents a hit-rate test of differences in the probability that a discretionary search results in an officer finding contraband. Panel (a) documents the probability of contraband being found as a result of a Black/African American occupant (purple-

[5] Simoiu et al. (2017) also propose a threshold-style test that has the benefit of alleviating potential concerns of infra-marginality in the hit-rate style tests but at the cost expense of adding significant complexity. In an effort to propose a parsimonious solution, we have limited our focus to hit-rate tests.

[6] Note that hit-rate style tests are typically used with searches where the "hit" is contraband being found and is not a discretionary decision on the part of officers. In this analysis, arrest is used as a proxy for contraband being found in searches and for the true guilt rate in vehicle exits and use of force. Imagining that there is also disparate treatment towards people of color in terms of the probability of arrest and that arrests overstate the true guilt rate, we might imagine that a hit-rate style test would be potentially biased against finding discrimination even when it exists. Given the limitations of the NOPD data, using arrest as a proxy for guilt is all that is currently possible in the current analysis.

[7] We identify consent and reasonable articulable suspicion searches using the RMS system and oversight/review data. We caution that the data appear to be most reliable from the second half of 2011 through the first half of 2020. However, results using a sample of all searches are qualitatively very similar to those obtained from this more restrictive subsample of discretionary searches. We feel strongly that measurement error in the identification of consent or reasonable articulable suspicion searches would not bias any of the findings from the hit-rate analysis.

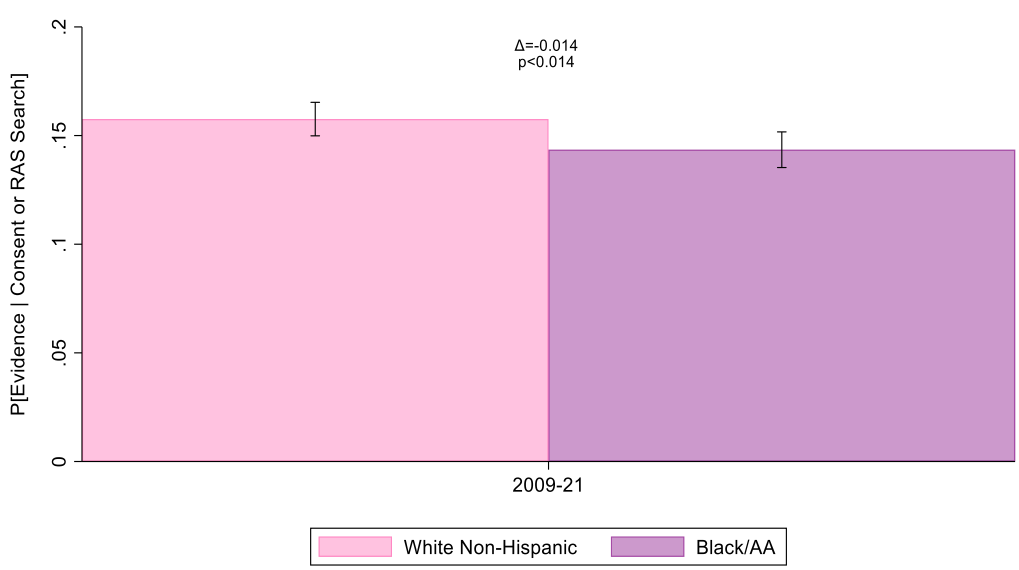
colored bar) being searched while panel (b) documents the probability of contraband being found as a result of a Hispanic/Latinx occupant (orange-colored bar) being searched. The baseline comparison group in both panels consists of White, non-Hispanic occupants (pink-colored bar). The vertical axis denotes the predicted probability that a search resulted in any type of contraband being found. The annotation in the center of each bar documents the magnitude and statistical significance of the difference in the probability a search resulted in evidence being seized for person of color relative to White, non-Hispanic occupants.

As shown below in panel (a) of Figure 3, Black/African American occupants were 1.4 percentage points less likely to have contraband seized as a result of a discretionary search relative to their White, non-Hispanic peers. Similarly, panel (b) reports that Hispanic/Latinx occupants were 3.9 percentage points less likely to have contraband seized as a result of a search relative to their White, non-Hispanic peers. The estimates for both groups were found to be highly statistically significant at a confidence level exceeding 98 percent.

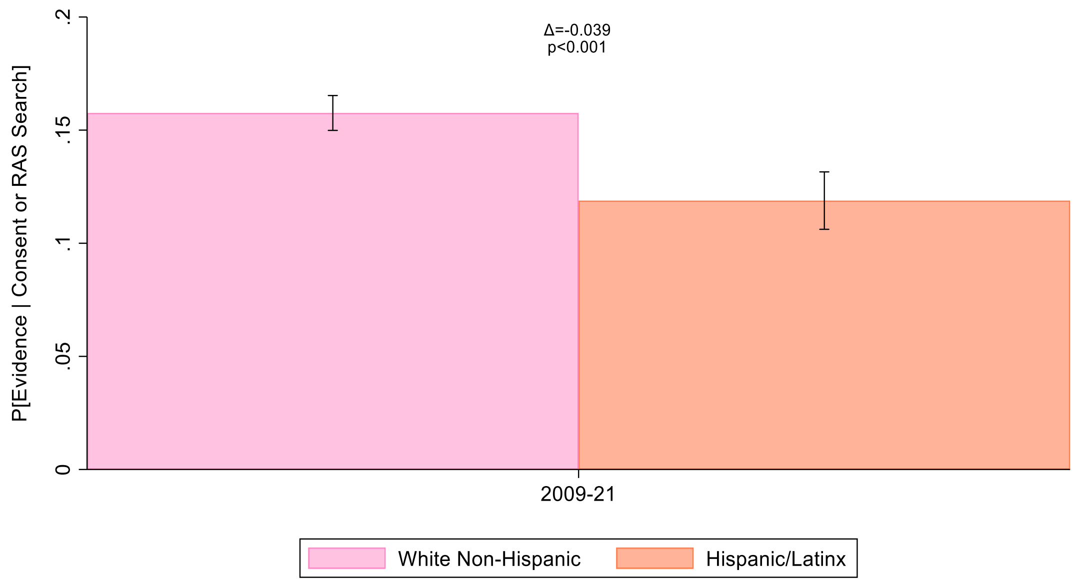
The disparity found for both Black/African American and Hispanic/Latinx groups suggests that the New Jersey State Police apply a lower threshold for searching a person of color which is indicative of potential discrimination.

Additional estimates for all other non-White occupants show a similar pattern and can be found in Appendix Figure A.5. Qualitatively similar results using the larger sample of all searches are contained in Appendix Figure A.7.

Figure 3. Discretionary Search Hit-Rate Estimates, 2009-21



(a) Black/African American

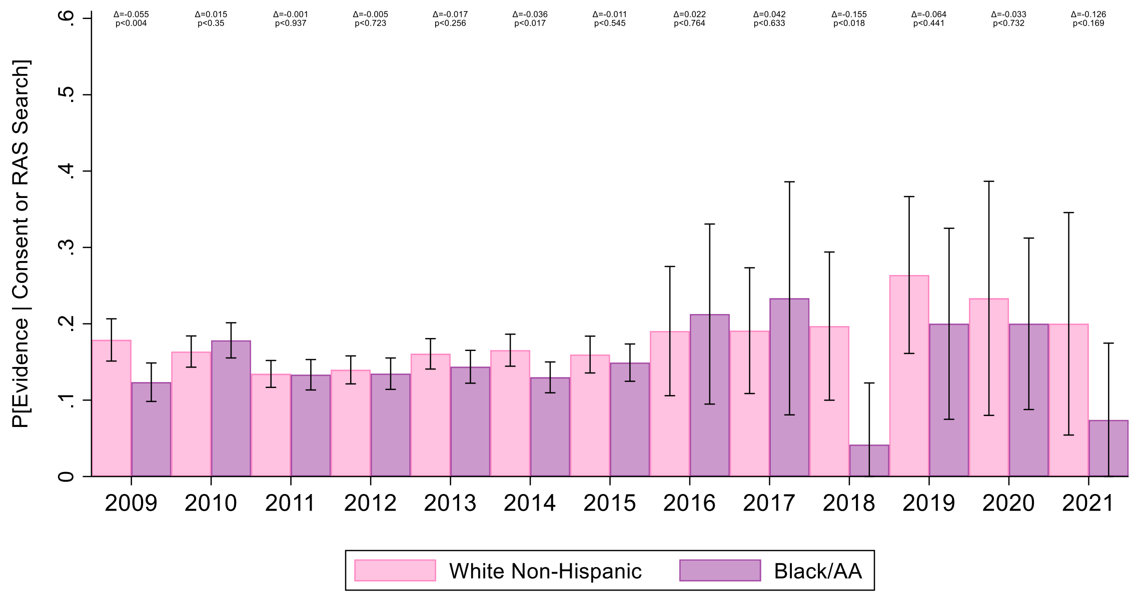


(b) Hispanic/Latinx

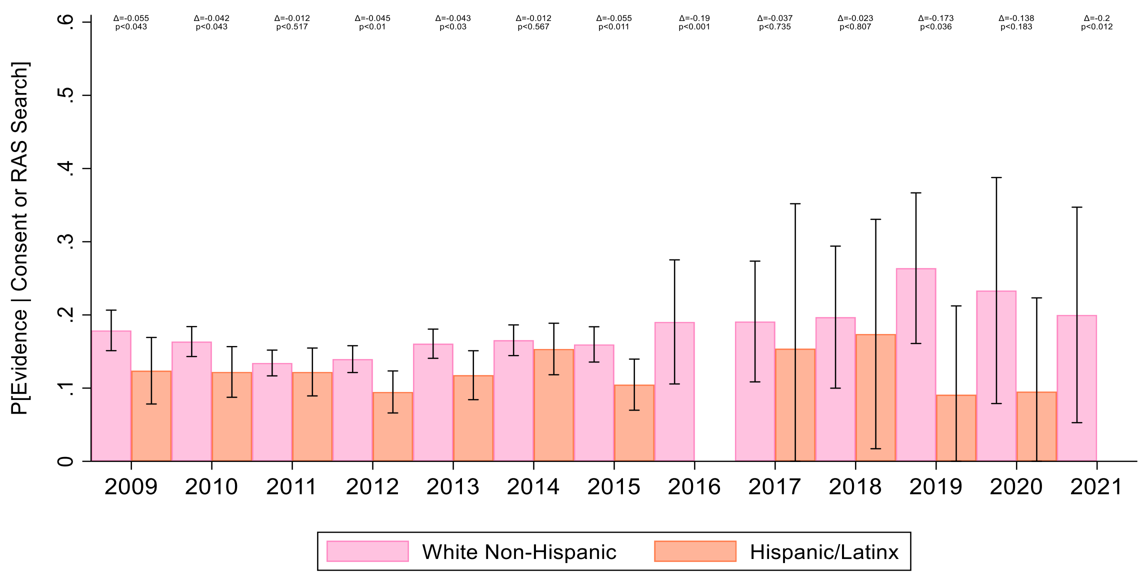
Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weigh by the inverse number of records per traffic stops. Each regression was estimated using data only for the respective year labeled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

Examining heterogeneity across different years is difficult due to the large variations in the size of the sample across different years. Figure 2 disaggregates the prior set of results by each individual year from 2009 to 2021. As shown below in panel (a), discretionary searches of Black/African American occupants were statistically less likely to yield contraband in 2009, 2014, and 2018. The magnitude of the disparity ranged from 2.1 percentage points in 2009 to as much as 5.7 and 3 percentage points in 2020 and 2021. Additional estimates for all other non-White occupants show a similar pattern and can be found in Appendix Figure A.5. Only in the subset of years previously discussed were the disparities statistically different from zero. However, we note that the estimates contained in Appendix Figure A.8 for all searches were more consistently negative and statistically significant across most years. We also note that the sample of discretionary searches is too small for many individual stations/troops to conduct an analysis at that level.

Figure 4. Discretionary Search Hit-Rate Estimates by Year



(b) Black/African American



(a) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on a sample of discretionary searches which include those where the offer has provided reasonable articutable suspicion or have request consent. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether evidence was found on an indicator for race or ethnicity. Each regression was estimated using data only for the respective year labled on the horizontal axis. Eicker–Huber–White standard errors were used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value.

Conditional Outcome Analysis of Post-Stop Decisions

Unfortunately, the version of the New Jersey State Police RMS data provided for this study was not sufficiently detailed to run hit-rate tests for additional post-stop outcomes like vehicle exits and use of force. In particular, it was not possible to discern whether an arrest was the precipitating event or the result of asking an occupant to exit their vehicle (e.g. DUI enforcement or a perceived potential threat) or an application of force (e.g. response to resistance or detainment). Since we believe that a subset of these activities were incidental to an arrest, it was not possible to run hit-rate tests using arrest as an outcome as we did for discretionary searches and contraband. In an attempt to provide some descriptive evidence with respect to these other post-stop enforcement activities (exits, force, and arrest), we proceed by applying an outcome test where we condition on a very rich set of information pertaining to the motivating reason for the traffic stop. Although we acknowledge that this approach has limitations which are discussed in the prior section, we feel comfortable applying such a test because we have a particularly rich set of information about the motivating reason for each traffic stop.

Formally, we estimate a linear probability model of the form

$$\begin{aligned} 1[post_i] = \alpha + \beta\, 1[minority_i] \\ + dow_i + tod_i \\ + (badge_i * month_i) \\ + (statute_i * month_i) \\ + \mu_i \end{aligned} \quad (2)$$

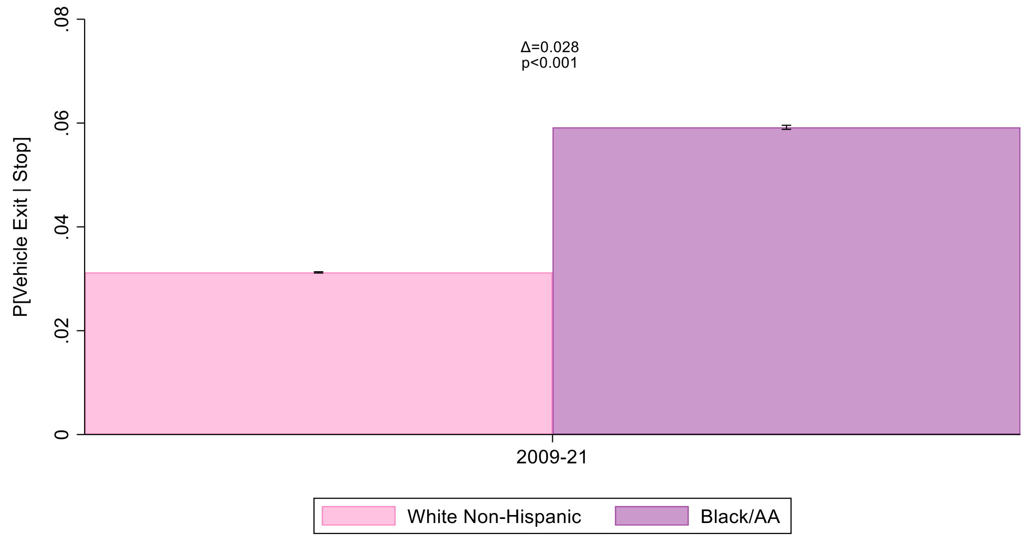
where the dependent variable $1[post_i]$ is an indicator which is equal to one if the traffic stop i resulted in a post-stop outcome like exit, force, or an arrest.[8] The primary independent variable $1[minority_i]$ is an indicator which is equal to one if traffic stop i was made of a person of color and zero otherwise. Additional control variables include six indicators for the day of the week (dow_i), 95 indicators for 15-minute time of day increments (tod_i), 163,939 indicators for each unique officer in each month and year ($badge_i * month_i$), and 23,589 indicators for each law enforcement statute in each month and year ($statute_i * month_i$). Across all of the estimates, we cluster standard errors at the officer by year by month level.

Figure 5 reports the probability that a motorist of color was asked to exit their vehicle while conditioning on a large number of observable features recorded by the officer about the traffic stop. Panel (a) documents the conditional probability of a Black/African American occupant (purple-colored bar) being asked to exit their vehicle relative to a White occupant (pink-colored bar). In particular, we find that Black/African American occupants were 2.8 percentage points (14.65 percent relative to a dependent mean of 0.19) more likely to be asked to exit their vehicle. Similarly, panel (b) documents the conditional probability of a Hispanic/Latinx occupant (orange-colored bar) being asked to exit their vehicle relative to Caucasian occupant (pink-colored bar). We find that Hispanic/Latinx occupants were 1.7 percentage points (9.6 percent relative to a dependent mean of 0.18) more likely to be asked to exit their vehicle. The estimates for both groups were found to be highly statistically significant at a confidence level exceeding 99 percent.

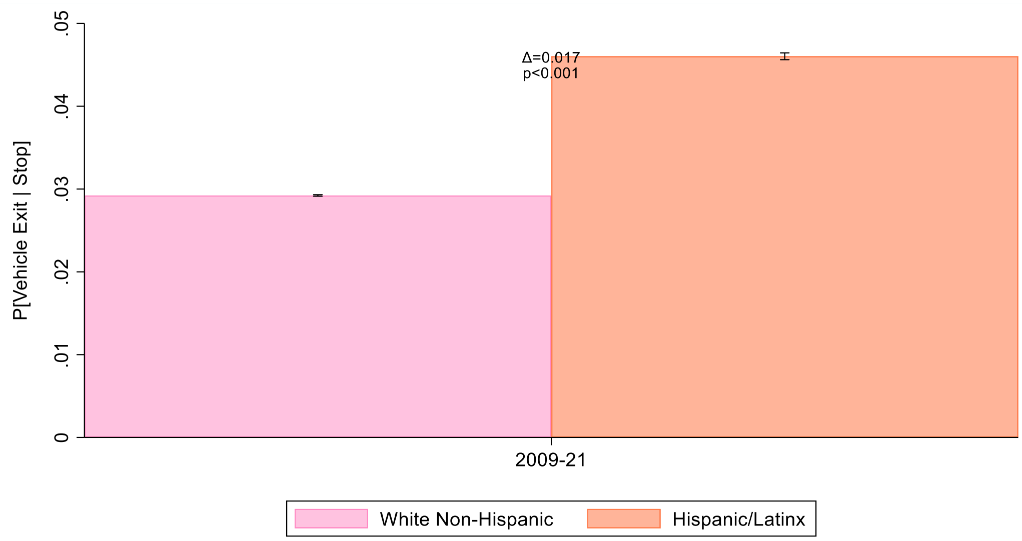
The disparity found for both Black/African American groups suggests that, conditional on the circumstances surrounding a traffic stop, the New Jersey State Police are more willing to ask a person of color to exit their vehicle. Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.11.

[8] For completeness, we also include a conditional outcome test for any search. These results are contained in Appendix Figures A.9 and A.10 and show a relatively large disparity consistent with the findings of the hit-rate test.

Figure 5. Conditional Outcome Estimates for Exit 2009-21



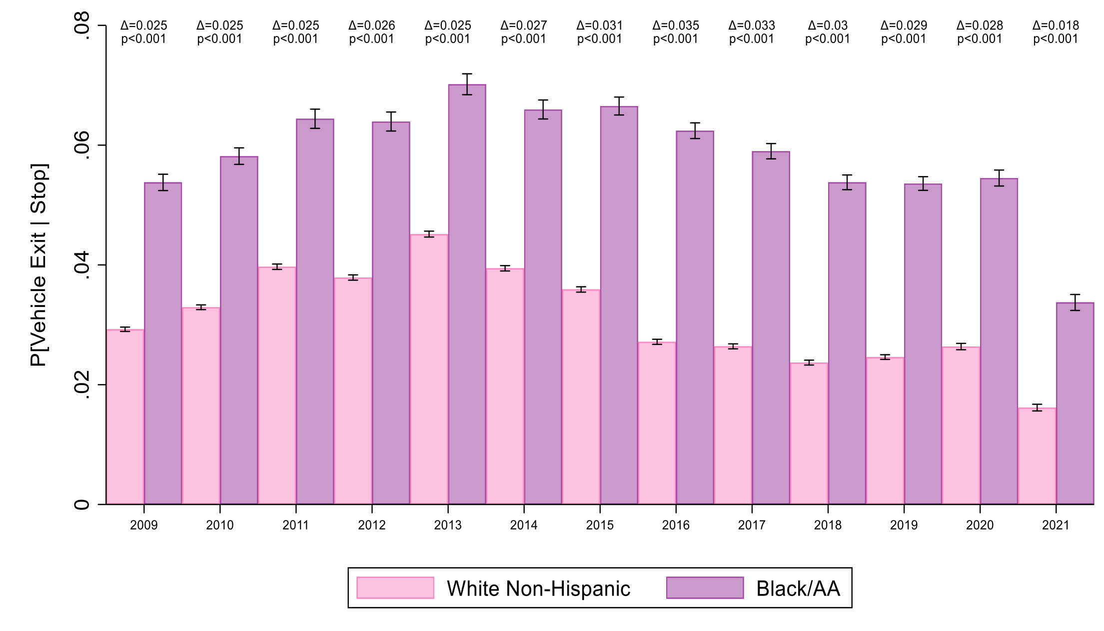
(a) Black/African American



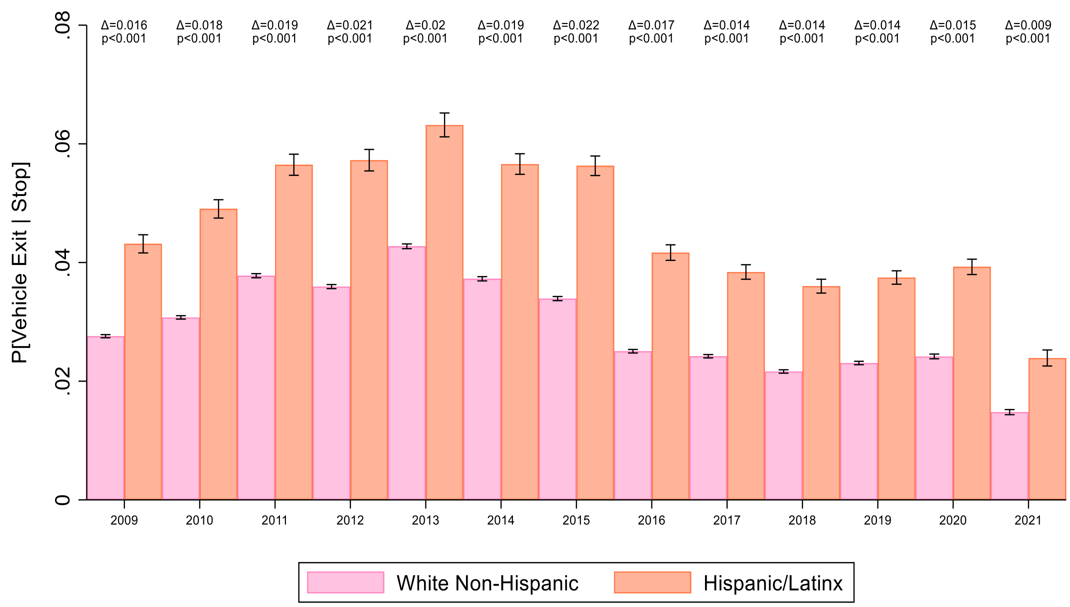
(b) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on a sample of discretionary searches which include those where the offer has provided reasonable articulable suspicion or has requested consent. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether evidence was found on an indicator for race or ethnicity. Each regression was estimated using data only for the respective year labeled on the horizontal axis. Eicker–Huber–White standard errors were used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value.

Figure 6. Conditional Outcome Estimates for Exit by Year



(b) Black/African American



(a) Hispanic/Latinx

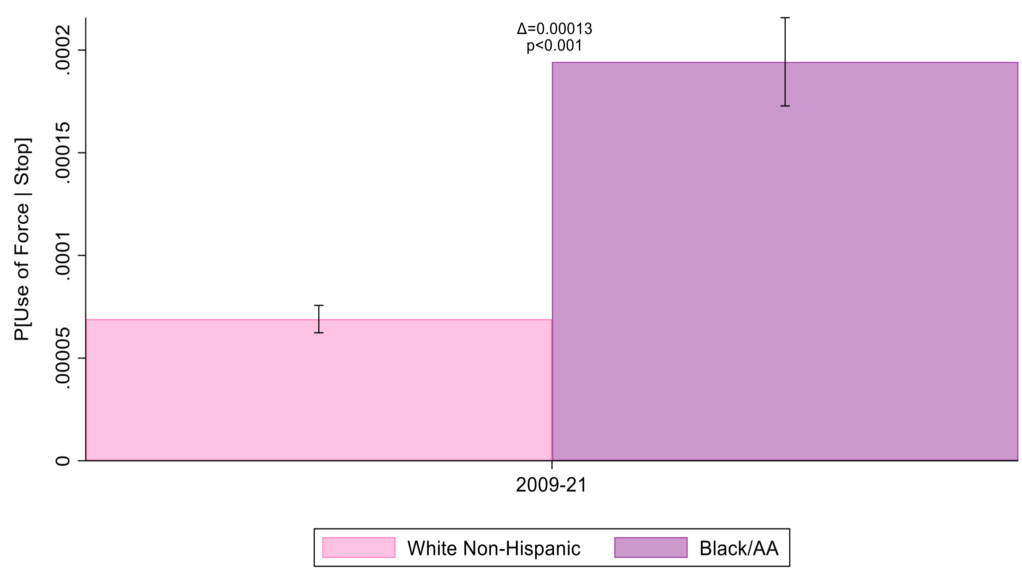
Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective year labeled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value clustered at the badge by month and by year levels.

Figure 6 explores heterogeneity in the conditional probability of a person of color being asked to exit their vehicle in each individual year from 2009 to 2021. In panel (a) the probability that a Black/African American occupant was asked to exit their vehicle was consistently higher than a White occupant. The magnitude of the disparity ranged from a low of 2.5 percentage points from 2009-13 to as much as 3.5 percentage points in 2016. In panel (b) the probability that a Hispanic/Latinx occupant was asked to exit their vehicle was also consistently higher. The magnitude of the disparity ranged from a low of 0.9 percentage points in 2021 to as much as 2.2 percentage points in 2015. The estimates for both groups across all years were found to be highly statistically significant at a confidence level exceeding 99 percent.

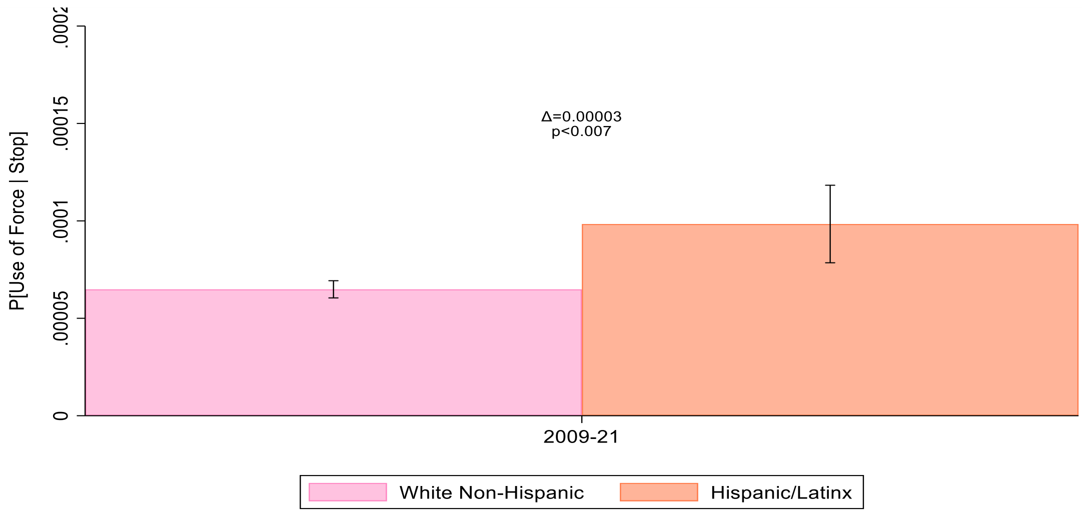
Across nearly all years in the sample, we find strong evidence suggesting a large and persistent disparity in the decision to ask a Black/African American or Hispanic/Latinx occupant to exit their vehicle.

Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.12. Examining heterogeneity across troops and stations, we also find that the disparities are remarkably consistent. In Troop A, we estimate that Black/African American and Hispanic/Latinx occupants were 3.45 (p<0.01) and 02.77 (p<0.01) percentage points more likely to be asked to exit their vehicle. In Troop B, we estimate that Black/African American and Hispanic/Latinx occupants were 2.28 (p<0.01) and 1.36 (p<0.01) percentage points more likely to be asked to exit their vehicle. In Troop C, we estimate that Black/African American and Hispanic/Latinx occupants were 5.24 (p<0.01) and 2.69 (p<0.01) percentage points more likely to be asked to exit their vehicle. In Troop D, we estimate that Black/African American and Hispanic/Latinx occupants were 1.57 (p<0.01) and 1.18 (p<0.01) percentage points more likely to be asked to exit their vehicle.

Figure 7. Conditional Outcome Estimates for Use of Force, 2009-21



(c) Black/African American



Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant experienced force on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month and by year levels.

(d) Hispanic/Latinx

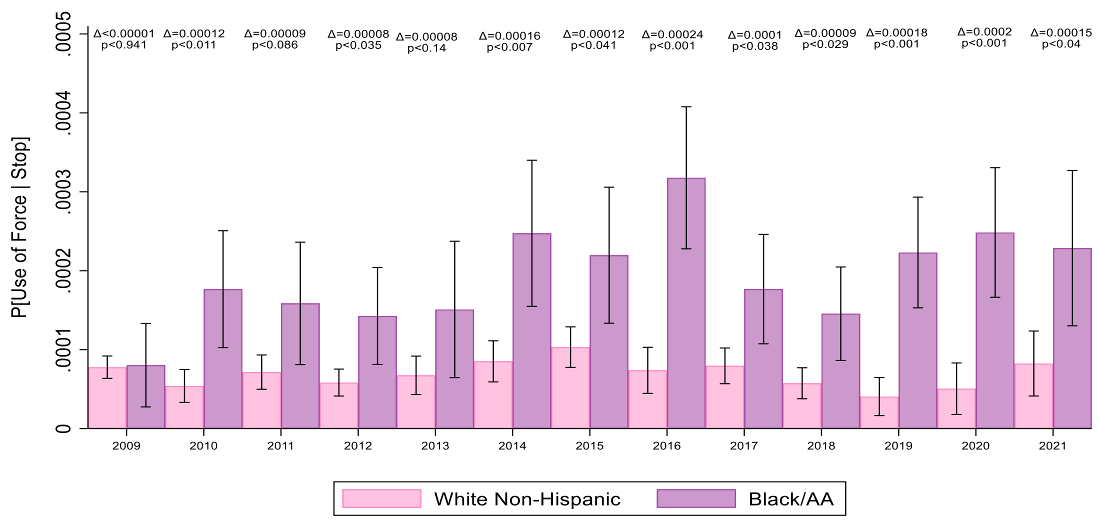
Across nearly all troops in the sample, we find strong evidence suggesting a large and persistent disparity in the decision to ask a Black/African American or Hispanic/Latinx occupant to exit their vehicle.

Appendix Figures B.5-8. contain separate estimates for each individual station within the four overarching troops.

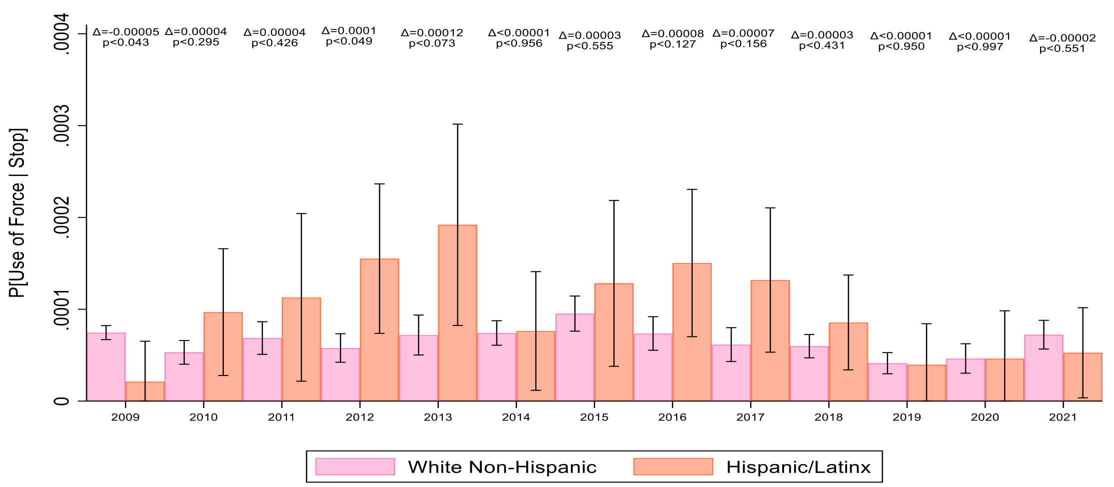
Figure 7 reports the probability of the use of force being applied to a person of color while conditioning on a large number of observable features recorded by the officer about the traffic stop. Panel (a) documents the conditional probability of force for a Black/African American occupant (purple-colored bar) being relative to a Caucasian occupant (pink-colored bar). In particular, we find that Black/African American occupants were 0.013 percentage points (130 percent relative to a dependent mean of 0.010) more likely to experience force. Similarly, panel (b) documents the conditional probability of a Hispanic/Latinx occupant (orange-colored bar) experiencing force relative to a Caucasian occupant (pink-colored bar). We find that Hispanic/Latinx occupants were 0.003 percentage points (27.5 percent relative to a dependent mean of 0.007) more likely to experience force. The estimates for both groups were found to be highly statistically significant at a confidence level exceeding 99 percent. *The disparity found for both groups suggests that, conditional on the circumstances surrounding a traffic stop, the New Jersey State Police are more willing to use force.* Additional estimates for all other all other non-White groups show a similar pattern and can be found in Appendix Figure A.14.

Figure 8 explores heterogeneity in the conditional probability of use of force against a person of color in each individual year from 2009 to 2021. In panel (a) the probability that a Black/African American occupant experienced force was consistently higher than a Caucasian. The magnitude of the disparity ranged from a low of 0.001 percentage points in 2009 to as much as 0.024 percentage points in 2016. In panel (b) the probability that a Hispanic/Latinx occupant experienced force was also higher than a Caucasian in all years except 2009 and 2021. The magnitude of the disparity ranged from a low of 0.001 percentage points in 2014 to as much as 0.012 percentage points in 2013. *The estimates for both groups across all years were found to be highly statistically significant in many of the years analyzed but we note the limitation of analyzing such a rare event.* We also note that the sample of force incidents was too small by individual troops/stations to conduct separate analyses at that level. Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.14.

Figure 8. Conditional Outcome Estimates for Use of Force by Year



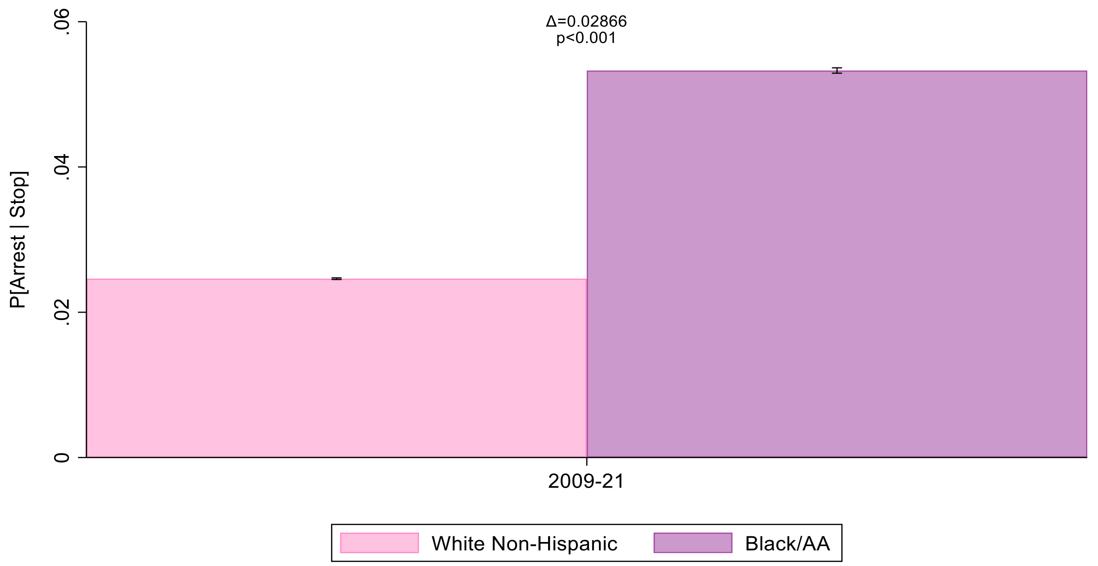
(b) Black/African American



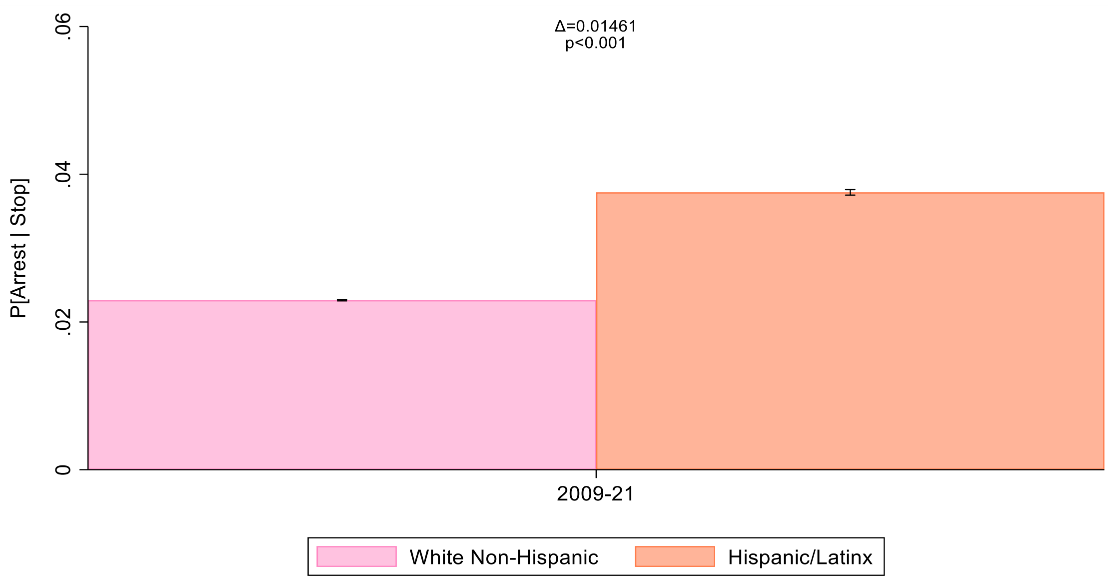
(a) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant experienced force on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month and by year levels.

Figure 9. Conditional Outcome Estimates for Arrest, 2009-21



(b) Black/African American



(a) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

Figure 9 reports the probability that a person of color was arrested while conditioning on a large number of observable features recorded by the officer about the traffic stop. Panel (a) documents the conditional probability of a Black/African American occupant (purple-colored bar) being arrested relative to a Caucasian occupant (pink-colored bar). In particular, we find that Black/African American occupants were 2.9 percentage points (87.5 percent relative to a dependent mean of 0.03) more likely to be arrested. Similarly, panel (b) documents the conditional probability of a Hispanic/Latinx occupant (orange-colored bar) being arrested relative to a Caucasian occupant (pink-colored bar). We find that Hispanic/Latinx occupants were 1.5 percentage points (56.8 percent relative to a dependent mean of 0.02) more likely to be arrested. The estimates for both groups were found to be highly statistically significant at a confidence level exceeding 99 percent.

The disparity found for both groups suggests that, conditional on the circumstances surrounding a traffic stop, the New Jersey State Police are more willing to arrest a person of color. Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.15.

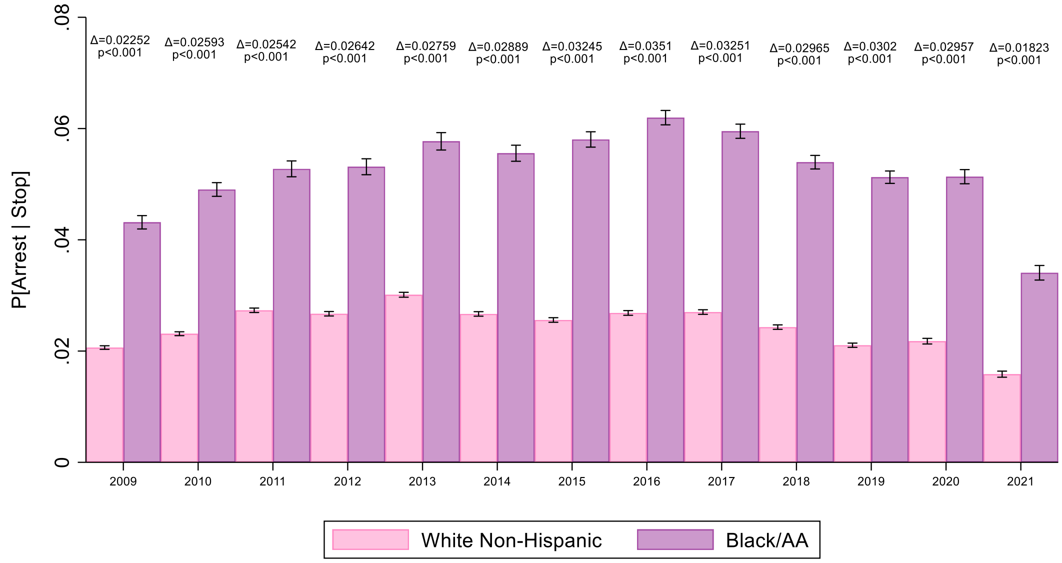
Figure 10 explores heterogeneity in the conditional probability of the arrest of a person of color in each individual year from 2009 to 2021. In panel (a) the probability that a Black/African American occupant being arrested was consistently higher than a Caucasian in all years. The magnitude of the disparity ranged from a low of 1.8 percentage points in 2021 to as much as 3.5 percentage points in 2016. In panel (b) the probability that a Hispanic/Latinx occupant was arrested was also higher than a Caucasian in all years except 2009 and 2021. The magnitude of the disparity ranged from a low of 0.9 percentage points in 2021 to as much as 1.8 percentage points in 2015. The estimate for both groups across all years were found to be highly statistically significant at a level exceeding 99 percent in all of the years analyzed.

Across nearly all years in the sample, we find strong evidence suggesting a large and persistent disparity in the decision to arrest a motorist of color.

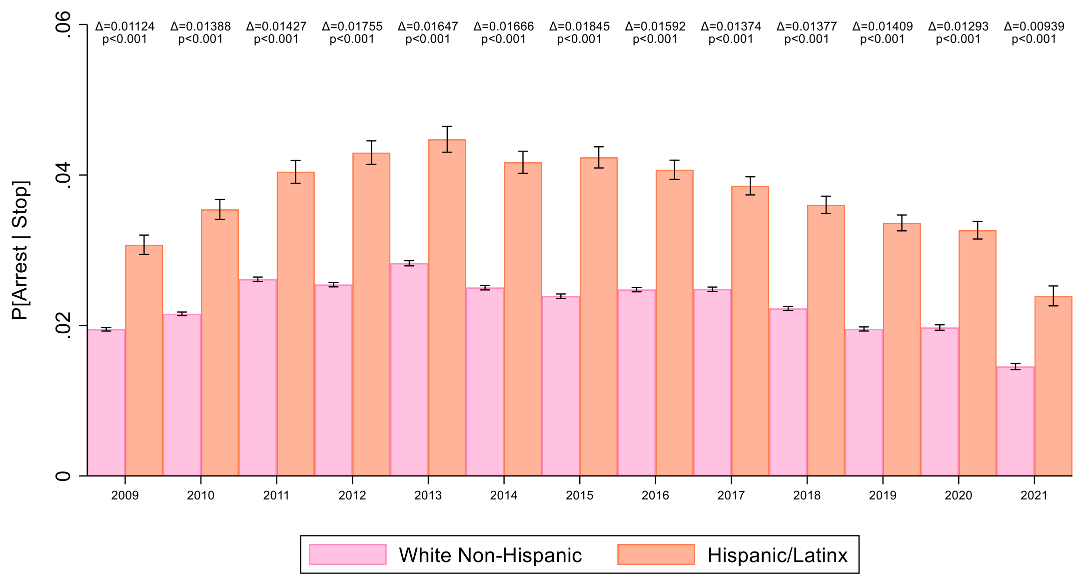
Additional estimates for all other non-White groups show a similar pattern and can be found in Appendix Figure A.16. Examining heterogeneity across troops and stations, we also find that the disparities are remarkably consistent. In Troop A, we estimate that Black/African American and Hispanic/Latinx occupants were 3.4 ($p<0.01$) and 02.35 ($p<0.01$) percentage points more likely to be arrested. In Troop B, we estimate that Black/African American and Hispanic/Latinx occupants were 2.47 ($p<0.01$) and 1.26 ($p<0.01$) percentage points more likely to be arrested. In Troop C, we estimate that Black/African American and Hispanic/Latinx occupants were 5.48 ($p<0.01$) and 2.44 ($p<0.01$) percentage points more likely to be arrested. In Troop D, we estimate that Black/African American and Hispanic/Latinx occupants were 1.55 ($p<0.01$) and 0.95 ($p<0.01$) percentage points more likely to be arrested.

Across nearly all troops in the sample, we find strong evidence suggesting a large and persistent disparity in the decision to arrest a motorist of color. Appendix Figures B.9-12. contain separate estimates for each individual station within the four overarching troops.

Figure 10. Conditional Outcome Estimates for Arrest by Year



(b) Black/African American



(a) Hispanic/Latinx

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective year labeled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level

Bibliography

Adger, C., Ross, M., & Sloan, C. (2022). The Effect of Field Training Officers on Police Use of Force.

Antonovics, K., & Knight, B. G. (2009). A new Look at racial profiling: Evidence from the Boston Police Department. Review of Economics and Statistics, 91(1), 163– 177. https://doi.org/10.1162/rest.91.1.163

Anwar, S., & Fang, H. (2006). An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. American Economic Review, 96(1), 127– 151. https://doi.org/10.1257/000282806776157579

Chanin, J., Welsh, M., Nurge, D., & Henry, S. (2016). Traffic enforcement in San Diego, California An analysis of SDPD vehicle stops in 2014 and 2015.

Criminal Justice Policy Research Institute. (2017). Law enforcement contacts policy and data review committee: 2016 Annual report. Retrieved from https://www.pdx.edu/cjpri/sites/www.pdx.edu.cjpri/files/LECC_2016_Annual_Report_FINAL011917_0.pdf

Dharmapala, D., Ross, S., Dharmapala, D., & Ross, S. (2004). Racial bias in motor vehicle searches: Additional theory and evidence. The B.E. Journal of Economic Analysis & Policy, 3(1), 1– 23.

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. https://doi.org/10.1198/016214506000000168

Horrace, W. C., & Rohlin, S. M. (2016). How dark is dark?: Bright lights, big city, racial profiling. Review of Economics and Statistics, 98(2), 226– 232. https://doi.org/10.1162/REST_a_00543

Kalinowski, J. J., Ross, M. B., & Ross, S. L. (2018). Now you see me, now you don't: The geography of police stops. AEA Papers and Proceedings, 109, 143– 147. https://doi.org/10.1257/pandp.20191026

Kalinowski, J., Ross, M., & Ross, S. L. (2020). Addressing seasonality in veil of darkness tests for discrimination: An instrumental variables approach. Working Papers 2019-028, Human Capital and Economic Opportunity Working Group.

Kalinowski, J., Ross, M., & Ross, S. L. (2022). Endogenous driving behavior in tests of racial profiling in police traffic stops. Journal of Human Resources, forthcoming. https://doi.org/10.3368/jhr.0822-12513R1

Knowles, J., Persico, N. & Todd, P. (2001). Racial bias in motor vehicles: Theory and evidence. Journal of Politcal Economy, 109(1), 203– 229. https://doi.org/10.1086/318603

Masher, J. (2016). What the "veil of darkness" says about new orleans traffic stops. Retrieved from https://nolacrimenews.com/2016/09/08/what-the-veil-of-darkness-says-about-new-orleans-traffic-stops

Milyo, J. (2017). DWB in COMO: Understanding race disparities in vehicle stops and searches. Retrieved from https://www.columbiamissourian.com/news/local/traffic-stop-report-shows-racial-disparities-as-officials-

Racial & Identity Profiling Advisory Board. (2020). Annual Report 2020. State of California. Google Scholar

Renauer, B. C., Henning, K., & Covelli, E. (2009). Benchmarking traffic stop and search data. Portland Police Bureau.

Ridgeway, G. (2009). Cincinnati police department traffic stops: Applying RAND's framework to analyze racial disparities.

Ridgeway, G., & MacDonald, J. (2009). Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. Journal of the American Statistical Association, 104, 486

Ritter, J. A. (2017). How do police use race in traffic stops and searches? Tests based on observability of race. Journal of Economic Behavior and Organization, 132, 82– 98.

Ritter, J. A., & Bael, D. (2009). Detecting racial profiling in minneapolis traffic stops: A new approach. Center for Urban and Regional Affairs: Reporter. University of Minnesota.

Robbins, M., Saunders, J., & Kilmer, B. (2017). A framework for synthetic control methods with high dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. Journal of the American Statistical Association, 112(517), 109– 126.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2015). State of Connecticut traffic stop data analysis and findings, 2013–14. Connecticut: Connecticut Racial Profiling Prohibition Project.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2016). State of Connecticut traffic stop data analysis and findings, 2014–15. Connecticut: Connecticut State Legislature.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2017a). State of Rhode Island traffic stop data analysis and findings, 2016.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2017b). State of Connecticut traffic stop data analysis and findings, 2015–16. Connecticut: Connecticut State Legislature.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2018). State of Connecticut traffic stop data analysis and findings, 2016–17. Connecticut: Connecticut State Legislature.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2019a). State of Rhode Island traffic stop data analysis and findings, 2017.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2019b). State of Connecticut traffic stop data analysis and findings, 2018. Connecticut: Connecticut State Legislature.

Sanchagrin, K., Officer, K., McAlister, S., Weinerman, M., Rau, C., & Tallan, K. (2019). Statistical transparency of policing report per house bill 2355 (2017). Oregon Criminal Justice Commission.

Smith, M. R., Rojek, J., Tillyer, R., & Lloyd, C. (2017). San Jose police department traffic and pedestrian stop study. El Paso, TX: Center for Law and Human Behavior, The University of Texas at El Paso.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016a). A test of racial disproportionality in traffic stops conducted by the Fayetteville Police Department. Research Triangle Park, NC: RTI International.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016b). A test of racial disproportionality in traffic stops conducted by the Greensboro Police. Research Triangle Park, NC: RTI International.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016c). A test of racial disproportionality in traffic stops conducted by the Raleigh Police Department. Research Triangle Park, NC: RTI International.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016d). Exploring racial disproportionality in traffic stops conducted by the Durham Police Department. Research Triangle Park, NC: RTI International.

Wallace, D., Louton, B., Nelson, J., Gricius, M., Tyler, D. H., Coble, S., ... Kramer, K. (2017). Yearly report for the Maricopa County Sheriff's Office: Years 2015–16. Phoenix, AZ: Center for Violence Prevention & Community Safety, Arizona State University.

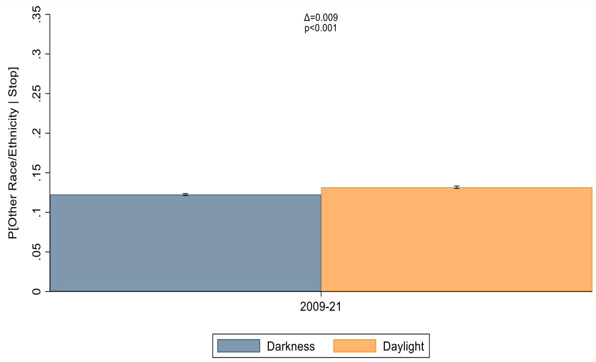
Worden, R. E., McLean, S. J., & Wheeler, A. P. (2010). Stops by Syracuse Police, 2006–2009. Albany, NY: The John F. Finn Institutefor Public Safety, Inc.

Worden, R. E., McLean, S. J., & Wheeler, A. P. (2012). Testing for racial profiling with the veil-of-darkness method. Police Quarterly, 15(1), 92– 111.

Weisburst, E. K. (2022). “Whose help is on the way?” The importance of individual police officers in law enforcement outcomes. *Journal of Human Resources*, 0720-11019R2.

Appendix A: Additional Statewide Results

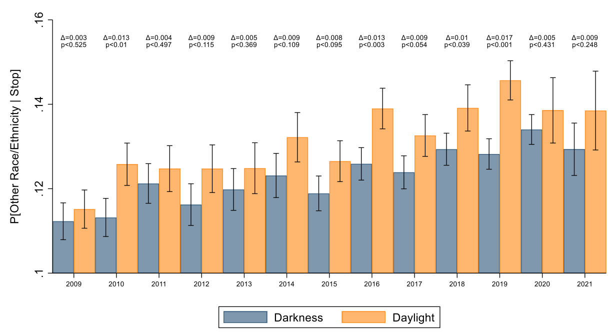
Appendix Figure A.1. Solar Visibility Analysis Estimates, 2009-21



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

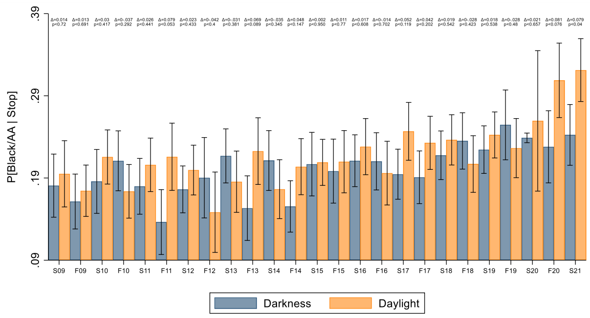
Appendix Figure A.2. Solar Visibility Analysis Estimates by Year



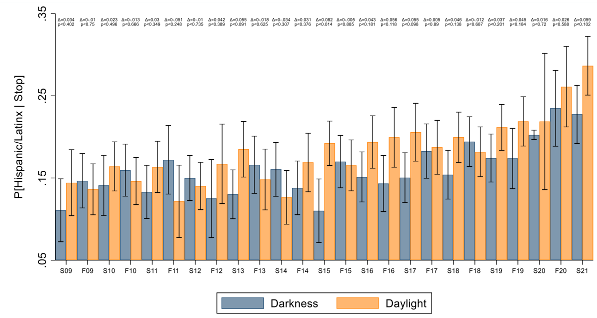
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

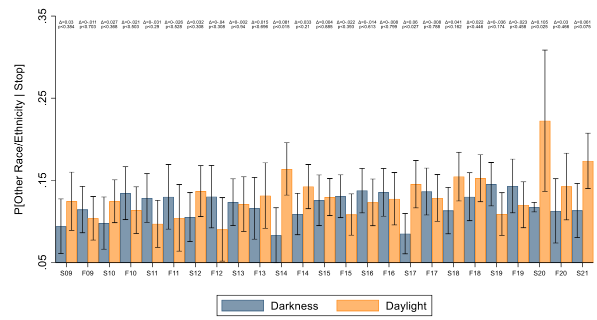
Appendix Figure A.3. DST Solar Visibility Analysis Estimates



(a) Black/African American



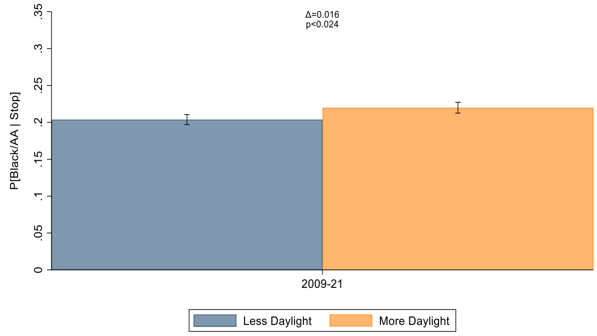
(b) Hispanic/Latinx



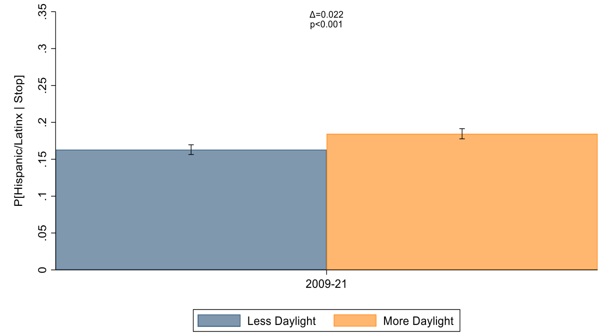
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight and within 21 days of the spring/fall daylight savings time shift. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for the period with more light before/after the spring/fall daylight savings time shift as well as controls for time of day, day of week, and badge by year. We implement a regression discontinuity design and also include a running variable, post-discontinuity runnin variable, and interactions between these variables with an indicator for season The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. Each regression was estimated using data only for the respective year labled on the horizontal axis.The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

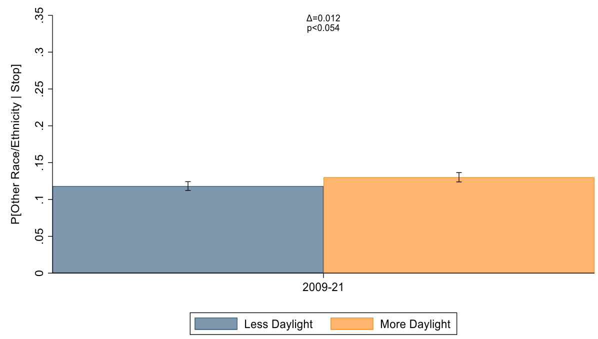
Appendix Figure A.4. DST Solar Visibility Analysis Estimates



(a) Black/African American



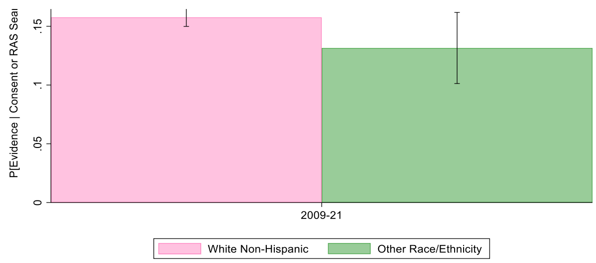
(b) Hispanic/Latinx



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight and within 21 days of the spring/fall daylight savings time shift. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for the period with more light before/after the spring/fall daylight savings time shift as well as controls for time of day, day of week, and badge by year. We implement a regression discontinuity design and also include a running variable, post-discontinuity runnin variable, and interactions between these variables with an indicator for season The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

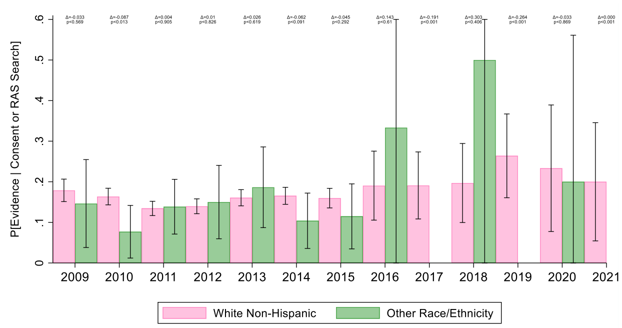
Appendix Figure A.5. Search Hit-Rate Estimates, 2009-21



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample of discretionary searches which include those where the offer has provided reasonable articutable suspicion or have request consent. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether evidence was found on an indicator for race or ethnicity. Eicker–Huber–White standard errors were used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value.

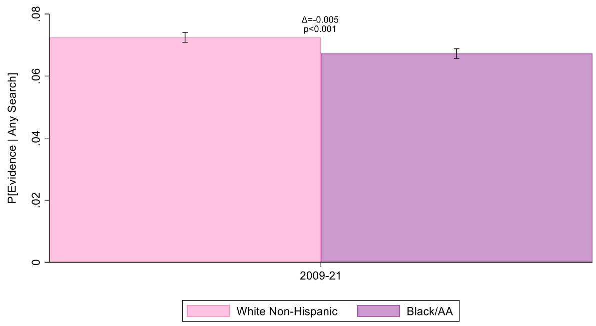
Appendix Figure A.6. Search Hit-Rate Estimates by Year



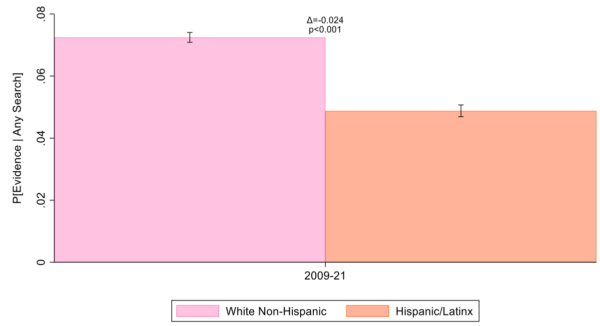
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample of discretionary searches which include those where the offer has provided reasonable articutable suspicion or have request consent. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether evidence was found on an indicator for race or ethnicity. Each regression was estimated using data only for the respective year labled on the horizontal axis. Eicker–Huber–White standard errors were used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value.

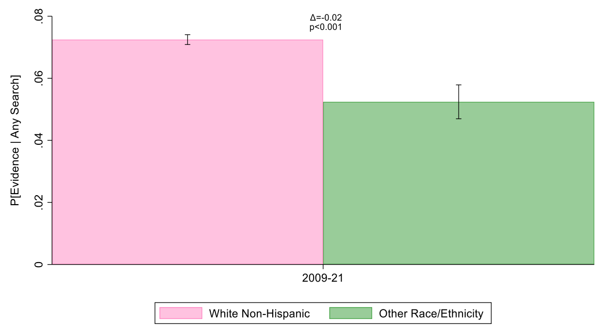
Appendix Figure A.7. Search Hit-Rate Estimates for All Searches, 2009-21



(a) Black/African American



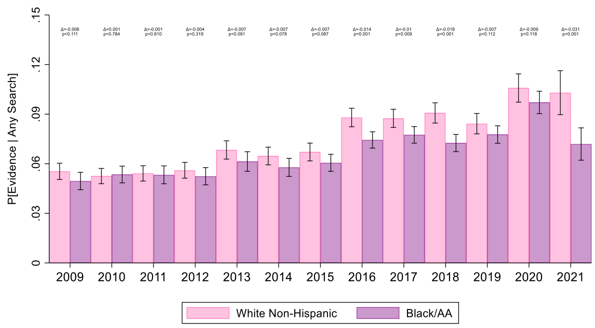
(b) Hispanic/Latinx



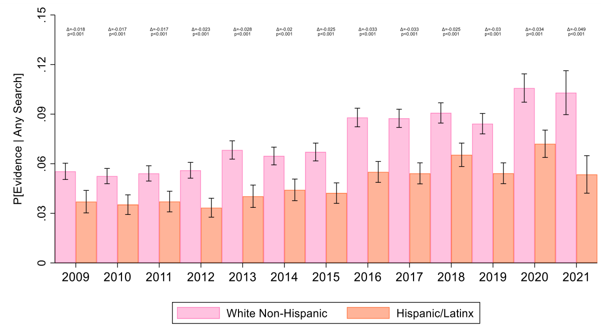
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample of discretionary searches which include those where the offer has provided reasonable articutable suspicion or have request consent. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether evidence was found on an indicator for race or ethnicity. Eicker–Huber–White standard errors were used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value.

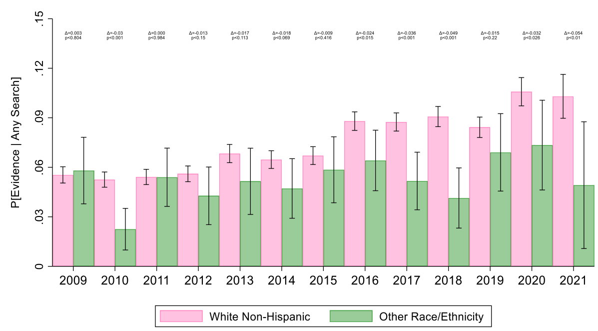
Appendix Figure A.8. Search Hit-Rate Estimates for All Searches by Year



(a) Black/African American



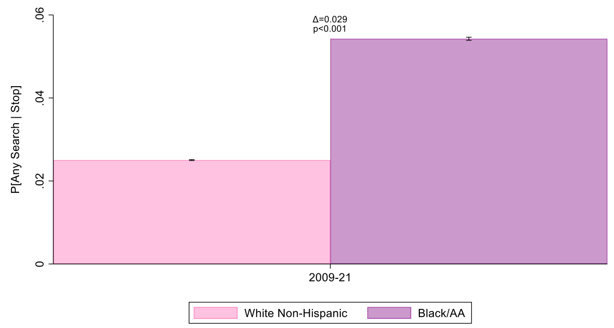
(b) Hispanic/Latinx



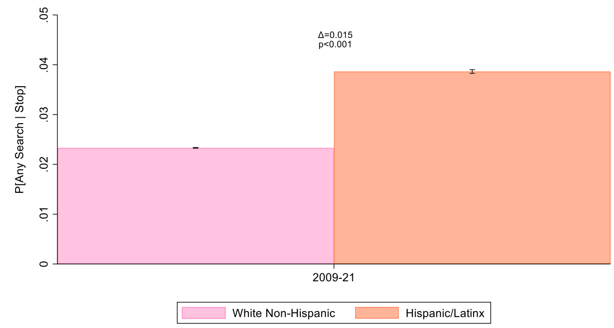
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample of discretionary searches which include those where the offer has provided reasonable articutable suspicion or have request consent. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether evidence was found on an indicator for race or ethnicity. Each regression was estimated using data only for the respective year labled on the horizontal axis. Eicker–Huber–White standard errors were used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value.

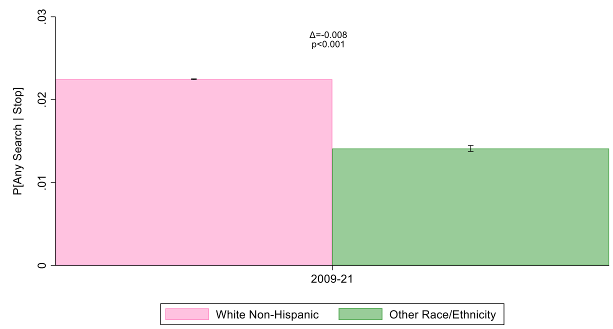
Appendix Figure A.9. Conditional Outcome Estimates for Search 2009-21



(a) Black/African American



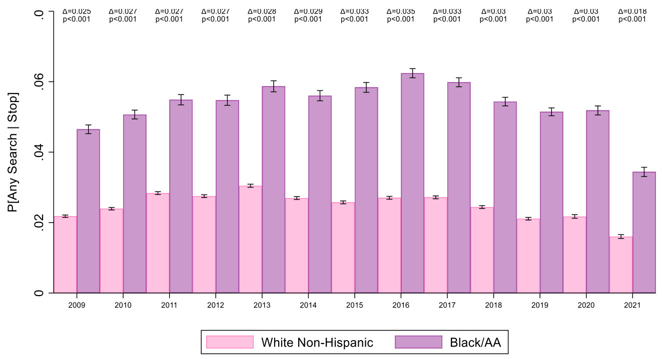
(b) Hispanic/Latinx



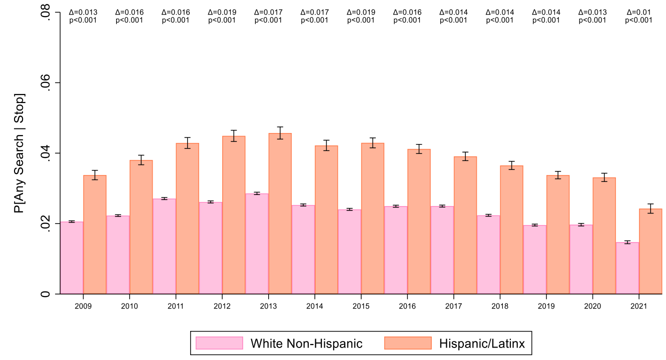
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was searched on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

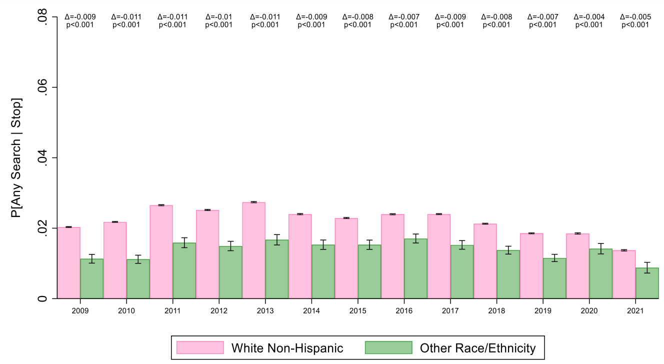
Appendix Figure A.10. Conditional Outcome Estimates for Search by Year



(a) Black/African American



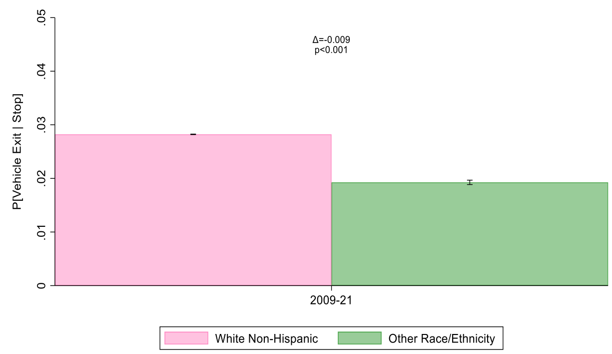
(b) Hispanic/Latinx



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was searched on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

Appendix Figure A.11. Conditional Outcome Estimates for Exit, 2009-21



(c) All Other Race/Ethnicity

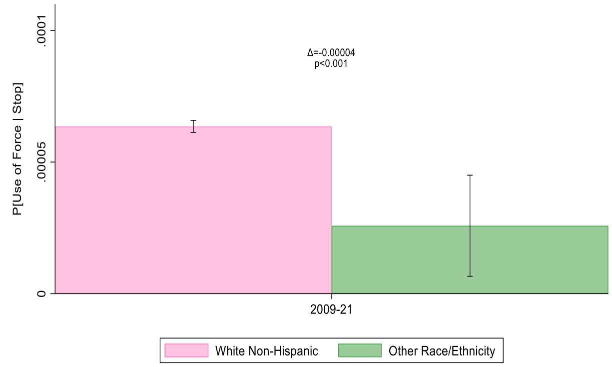
Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

Appendix Figure A.12. Conditional Outcome Estimates for Exit by Year

(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective year labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

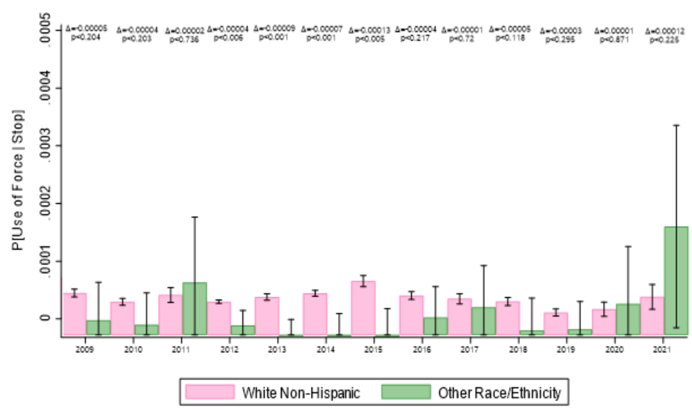
Appendix Figure A.13. Conditional Outcome Estimates for Use of Force, 2009-21



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant experienced force on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

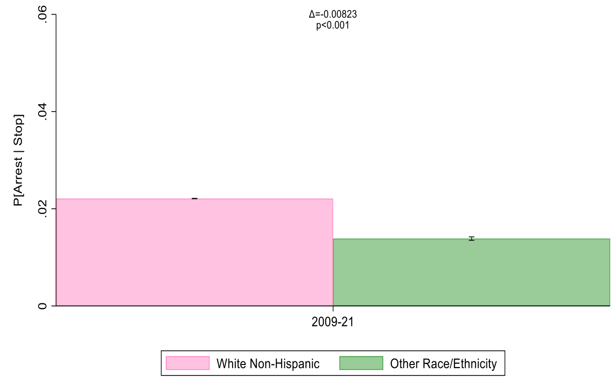
Appendix Figure A.14. Conditional Outcome Estimates for Use of Force by Year



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant experienced force on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective year labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

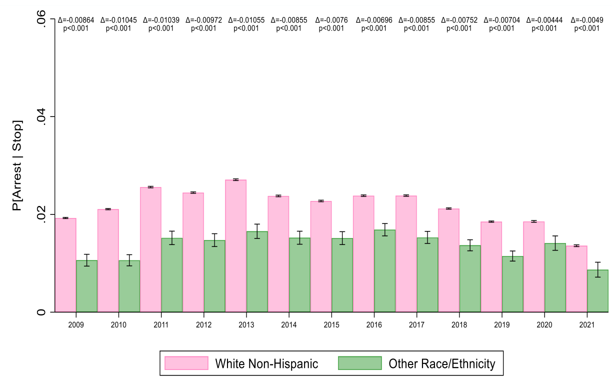
Appendix Figure A.15. Conditional Outcome Estimates for Arrest, 2009-21



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

Appendix Figure A.16. Conditional Outcome Estimates for Arrest by Year

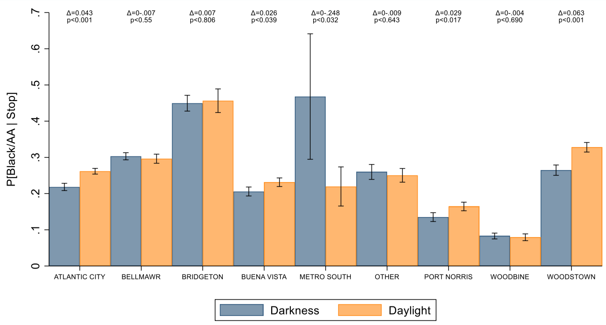


(c) All Other Race/Ethnicity

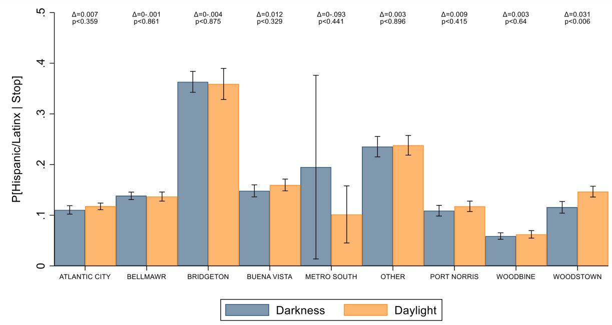
Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective year labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

Appendix B: Additional Station Results

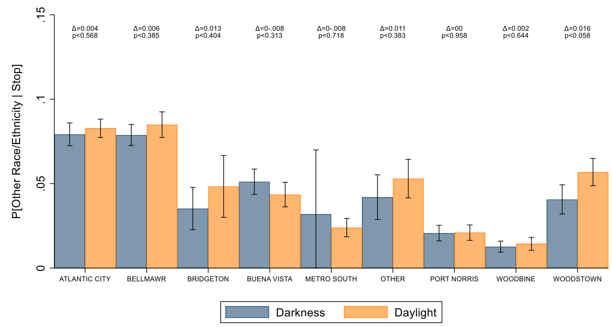
Appendix Figure B.1. Solar Visibility Analysis Estimates by Station for Troop A, 2009-21



(a) Black/African American



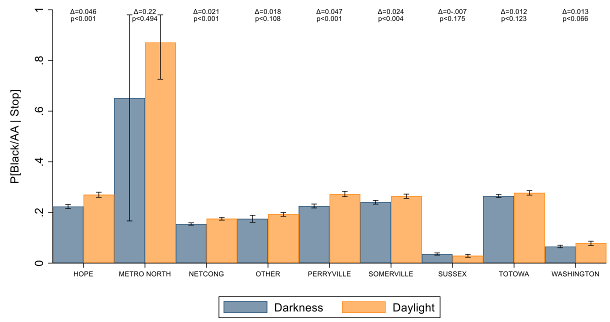
(b) Hispanic/Latinx



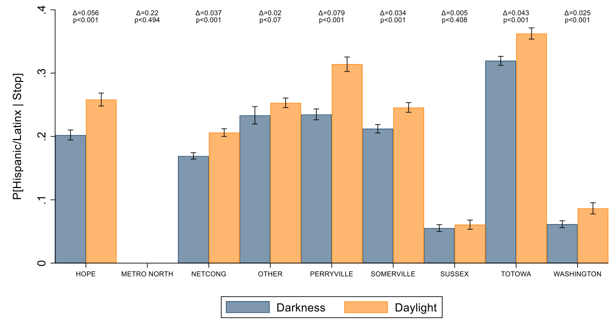
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. Each regression was estimated using data only for the respective station from troop A labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

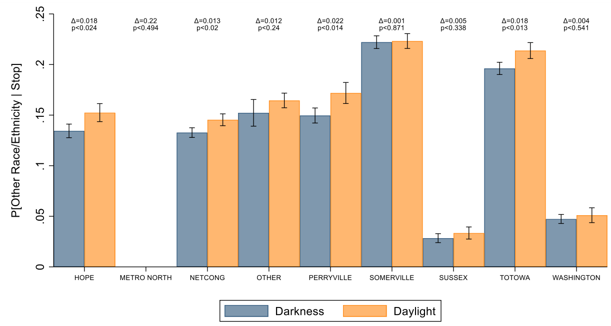
Appendix Figure B.2. Solar Visibility Analysis Estimates by Station for Troop B, 2009-21



(a) Black/African American



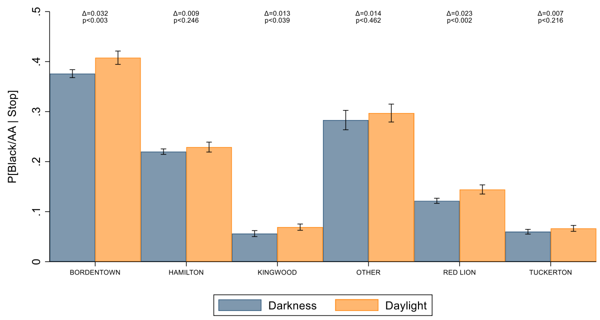
(b) Hispanic/Latinx



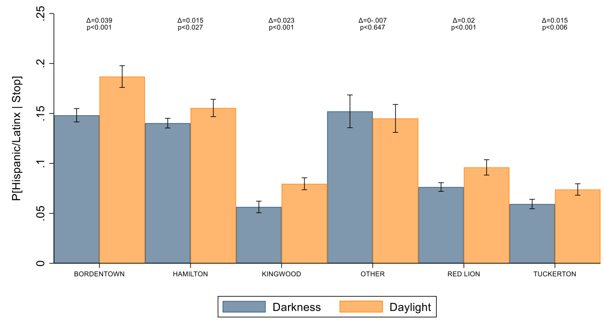
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. Each regression was estimated using data only for the respective station from troop B labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

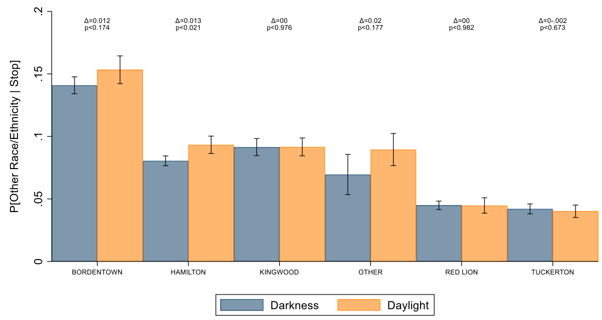
Appendix Figure B.3. Solar Visibility Analysis Estimates by Station for Troop C, 2009-21



(a) Black/African American



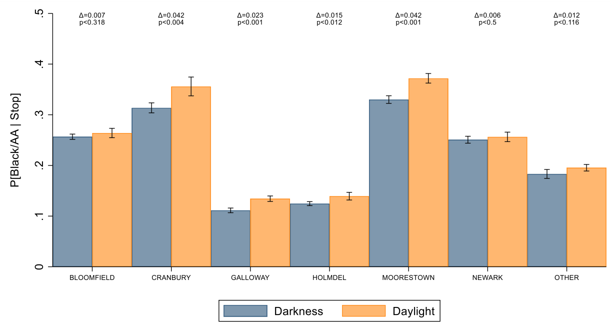
(b) Hispanic/Latinx



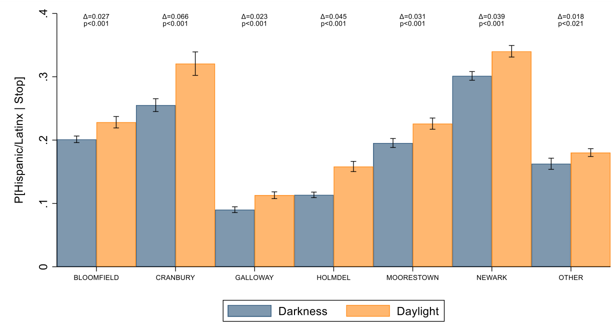
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. Each regression was estimated using data only for the respective station from troop C labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

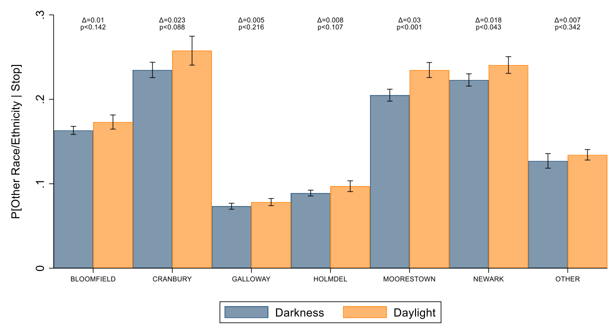
Appendix Figure B.4. Solar Visibility Analysis Estimates by Station for Troop D, 2009-21



(a) Black/African American



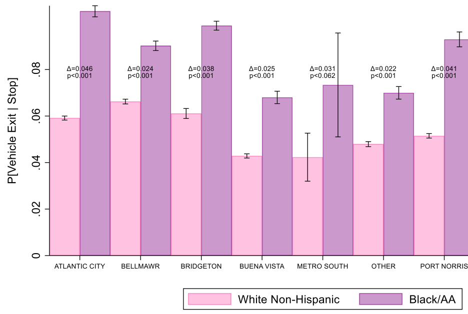
(b) Hispanic/Latinx



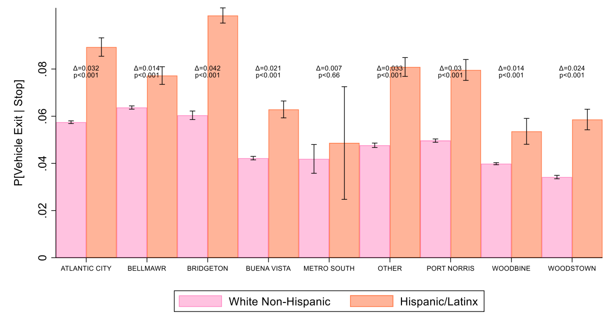
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on a sample traffic stops made of White non-Hispanic and Black/African American or Hispanic/Latinx motorists during the inter-twilight. The bars and estimated change were obtained by plotting the estimated probabilities at the mean of the control variables using a linear probability model regressing an indicator for race on an indicator for daylight as well as controls for time of day, day of week, and badge by year. The unit of observation is a traffic stop but there are a small number of observations with multiple records for different violations that we weight by the inverse number of records per traffic stops. Each regression was estimated using data only for the respective station from troop C labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by year level.

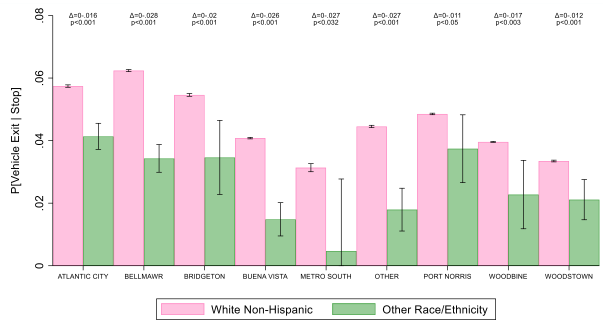
Appendix Figure B.5. Conditional Outcome Estimates for Exit and Troop A, 2009-21



(a) Black/African American



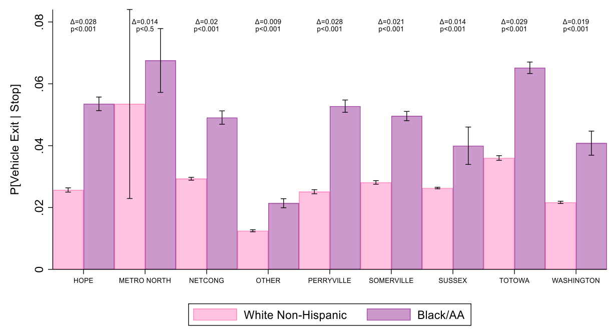
(b) Hispanic/Latinx



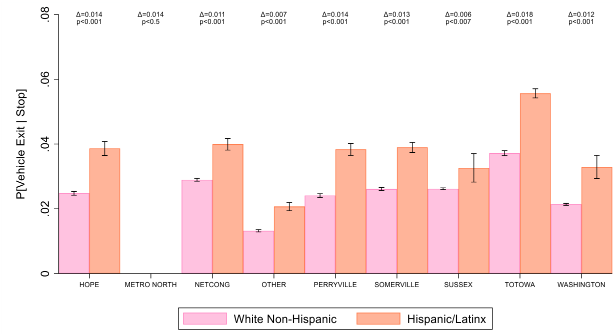
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop A labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

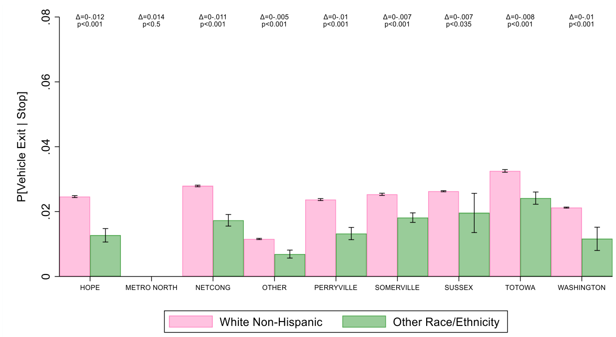
Appendix Figure B.6. Conditional Outcome Estimates for Exit and Troop B, 2009-21



(a) Black/African American



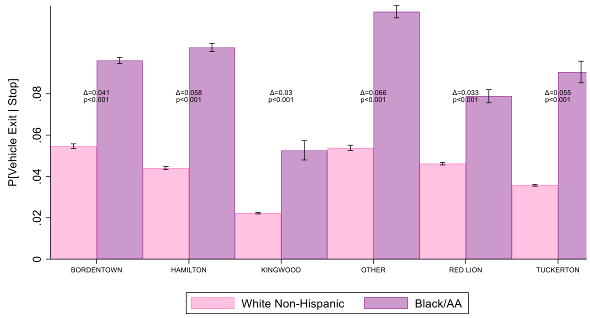
(b) Hispanic/Latinx



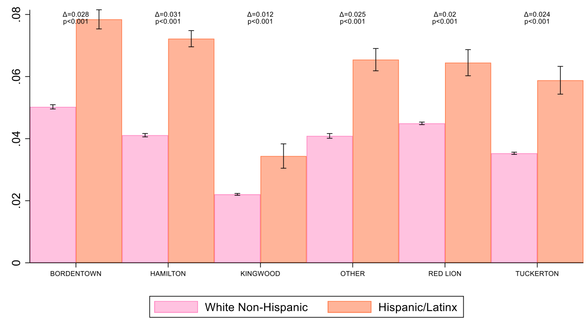
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop B labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

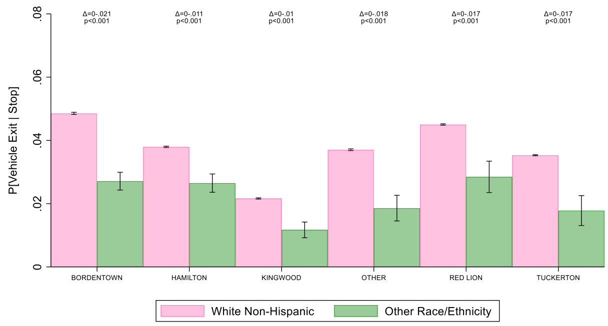
Appendix Figure B.7. Conditional Outcome Estimates for Exit and Troop C, 2009-21



(a) Black/African American



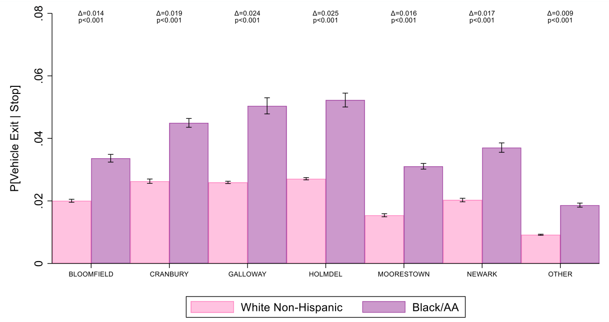
(b) Hispanic/Latinx



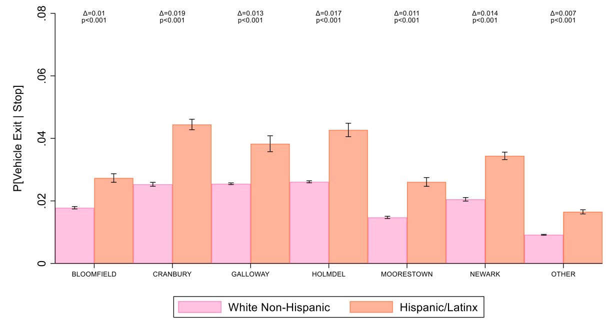
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop C labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

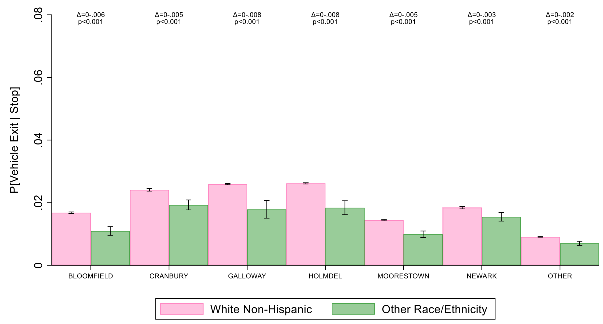
Appendix Figure B.8. Conditional Outcome Estimates for Exit and Troop D, 2009-21



(a) Black/African American



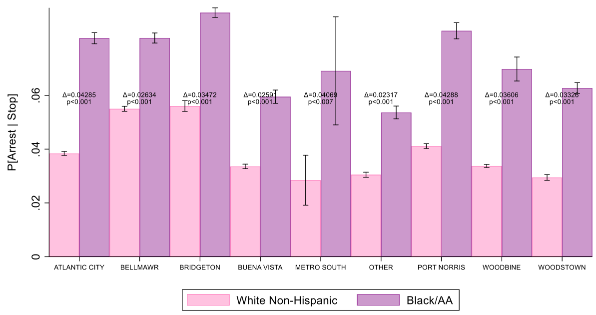
(b) Hispanic/Latinx



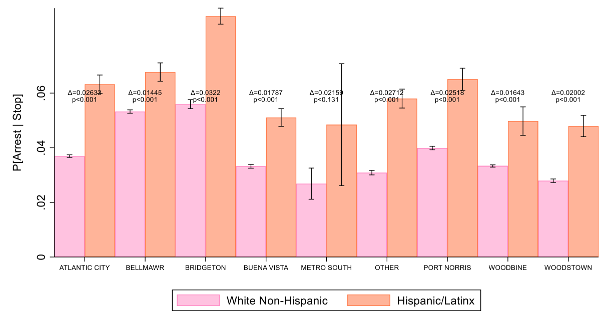
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was asked to exit their vehicle on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop D labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

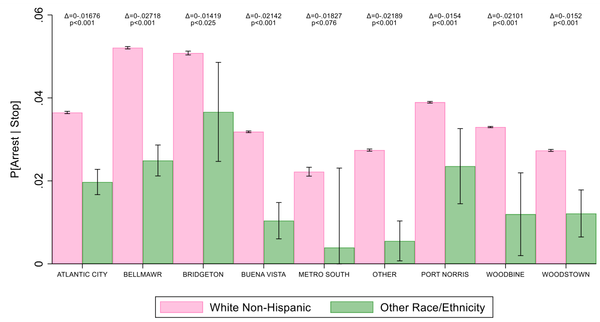
Appendix Figure B.9. Conditional Outcome Estimates for Arrest and Troop A, 2009-21



(a) Black/African American



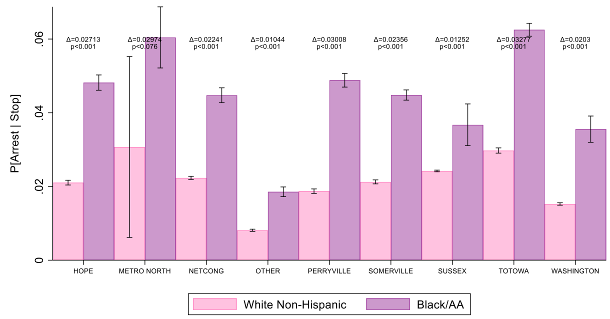
(b) Hispanic/Latinx



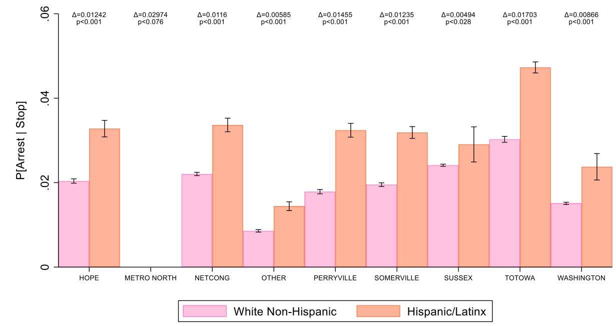
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop A labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

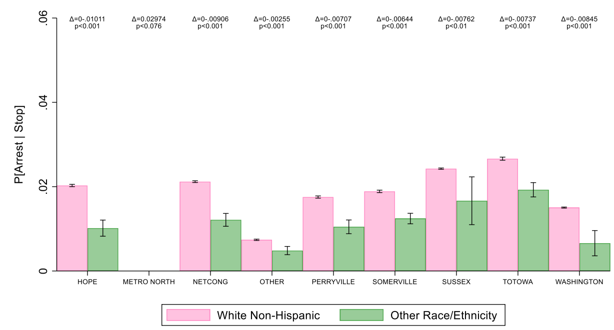
Appendix Figure B.10. Conditional Outcome Estimates for Arrest and Troop B, 2009-21



(a) Black/African American



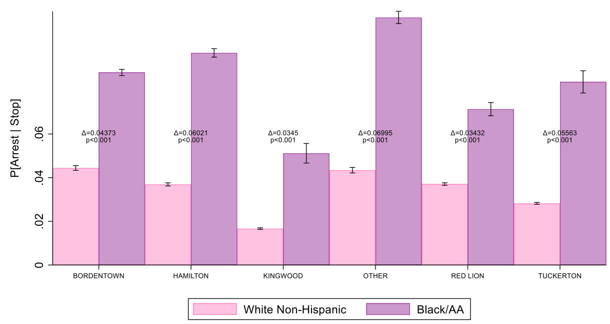
(b) Hispanic/Latinx



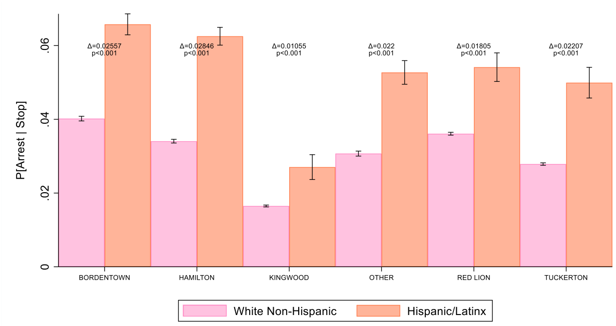
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop A labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

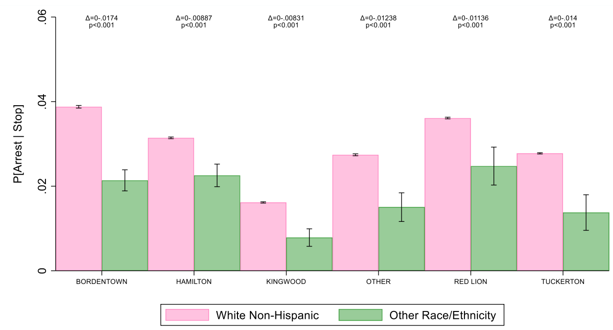
Appendix Figure B.11. Conditional Outcome Estimates for Arrest and Troop C, 2009-21



(a) Black/African American



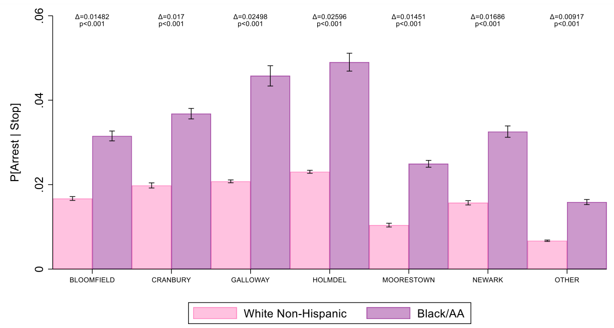
(b) Hispanic/Latinx



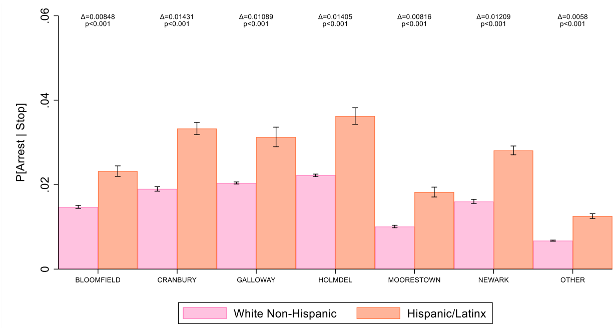
(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop A labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.

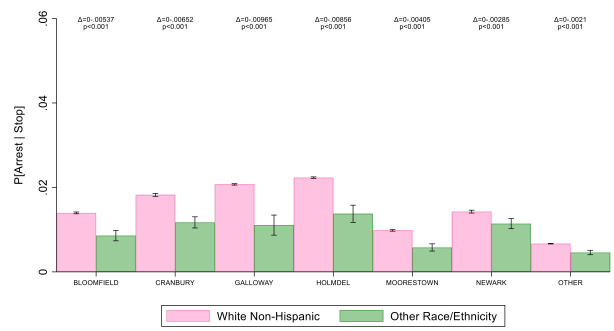
Appendix Figure B.12. Conditional Outcome Estimates for Arrest and Troop D, 2009-21



(a) Black/African American



(b) Hispanic/Latinx



(c) All Other Race/Ethnicity

Notes: The results for panels (a) and (b) are estimated on the universe of traffic stops. The numbers underlying the figure were obtained by estimating a linear probability model regressing an indicator of whether an occupant was arrested on an indicator for race or ethnicity as well as a rich set of controls. The controls in each regression include indicators for time of day, day of week, year by month by badge, and year by month by statute. Each regression was estimated using data only for the respective station from troop D labled on the horizontal axis. The standard errors used to construct the confidence intervals and to conduct the hypothesis test denoted by the p-value were clustered at the badge by month by year level.