Statistical Transparency of Policing Report
Per House Bill 2355 (2017)

November 25, 2019

Oregon Criminal Justice Commission

Michael Schmidt
Executive Director

The mission of the Oregon Criminal Justice Commission is to improve the legitimacy, efficiency, and effectiveness of state and local criminal justice systems.

Authors and Contributors

Kenneth Sanchagrin, JD PhD
STOP Program Director
Oregon Criminal Justice Commission Deputy Director & Research Director

Kelly Officer
Siobhan McAlister
Michael Weinerman
Courtney Rau
Katherine Tallan
Oregon Criminal Justice Commission Researchers

Executive Summary

House Bill 2355 (2017) mandated that by 2021, all Oregon law enforcement agencies must submit data regarding officer initiated traffic and pedestrian stops to the Oregon Criminal Justice Commission, so the Commission could analyze the submitted data for evidence of racial or ethnic disparities on an annual basis. To accomplish these ends, the Commission, along with the Oregon State Police and the Oregon Department of Public Safety Standards and Training (DPSST), created the Oregon Statistical Transparency of Policing (STOP) Program. This is the first annual report to the Oregon Legislature by the STOP Program examining data received pursuant to HB 2355.

Since the passage of HB 2355, the STOP Program developed a standardized method for data collection, developed and offered data collection software to law enforcement agencies, and is receiving data from more than fifty Oregon law enforcement entities. As required by the Bill, this inaugural report includes results for analyses examining the largest twelve law enforcement agencies in the state. In 2020, the STOP Program will report on the 65 largest agencies in the state and by 2021 it will report on all law enforcement agencies in Oregon.

Table E1 reports descriptive statistics for the first year of stop data, which represents stops made from July 1, 2018 through June 30, 2019, for the twelve reporting agencies. Across all agencies, the vast majority of the reported data were for traffic stops, although there was significant variation across agencies regarding the share of traffic versus pedestrian encounters. With regard to race the majority of stops in Oregon involved White individuals, which, in and of itself, is not surprising given the demographic makeup of Oregon as a whole. Overall, a little over one-fifth of stops involved a non-White individual. Finally, males were stopped more often than females and non-binary individuals.

Once the stop had been initiated, stopped individuals either were subject to no further action or merely given a warning in a little over 60 percent of stops. Other outcomes, including receiving a citation or being arrested, varied widely across agencies and are discussed in detail in the main body of the report. Finally, with regard to searches, approximately 3 percent of all stops resulted in a search of some type.

To examine the traffic and pedestrian stop data acquired by the STOP Program for racial/ethnic disparities, STOP Program researchers utilized three methods. The first method, which is used to examine the initial decision to stop an individual, was the Veil of Darkness Analysis (VOD). The VOD Analysis takes advantage of natural variations in daylight and darkness throughout the year and is based on the assumption that it is easier for an officer to discern race/ethnicity during the day when it is light versus the night when it is dark. Accordingly, the VOD Analysis compares stop rates for minority individuals to those for White individuals during the time

Table E1. Descriptive Statistics for Aggregate Tier 1 Stop Data

Variable	Percent
Traffic Stop	97.2%
Race/Ethnicity	
White	77.4%
Black	5.1%
Hispanic	12.1%
Asian/PI	3.4%
Native American	0.6%
Middle Eastern	1.4%
Gender	
Male	66.5%
Female	33.1%
Non-Binary	0.5%
Age	
Under 21	10.1%
21 – 29	24.4%
30 – 39	24.8%
40 - 49	16.8%
50 and Older	23.9%
Stop Disposition	
None	2.9%
Warning	57.3%
Citation	37.0%
Juvenile Summons	0.0%
Arrest	2.9%
Search Conducted	2.9%

windows surrounding sunrise and sunset. If, as demonstrated by the statistics that result from the VOD Analysis, minority individuals are more likely to be stopped in the daylight when race/ethnicity is easier to detect, then there would be evidence of a disparity.

The second analytical method employed by the STOP Program is the Predicted Disposition Analysis, which examines matched groups using a statistical technique called propensity score analysis to explore whether disparities exist in stop outcomes (i.e., citations, searches, or arrests). If, after matching on all available data points in the stop data (e.g., time of day and day of the week the stop was made, reason for the stop, gender, age), minority individuals are either cited, searched, or arrested more often than similarly situated White individuals, then there would be evidence of a disparity.

Finally, the STOP Program utilized the KPT Hit Rate Analysis, which compares relative rates of successful searches (i.e., those resulting in the seizure of contraband) across racial/ethnic groups. It is based on the assumption that if search decisions by officers are made based on race/ethnicity neutral criteria, then success rates should be similar, if not identical, across different racial/ethnic categories. If, however, search success rates differ and the search success rates for minority individuals are significantly lower than those reported for White individuals, then there would be evidence of a disparity.

To determine if disparities identified in this report warrant additional in-depth analysis and/or technical assistance from the Oregon Department of Public Safety Standards and Training (DPSST), STOP Program researchers reviewed the results of each of the three analyses conducted on the STOP Program data. For each individual analysis, an estimated disparity must meet the 95 percent confidence level for it to be statistically significant. Further, following best practices, for a law enforcement agency to be identified as one requiring further analysis as well as DPSST technical assistance, it must be identified as having a statistically significant disparity in two of the three analytical tests performed on the STOP data.

Using the above mentioned analyses and thresholds, the STOP Program identified one agency that had statistically significant results across two of the tests performed on the data: Portland Police Bureau. Specifically, results indicated that Portland Police Bureau had disparities in the Predicted Disposition Analysis with regard to searches and arrests involving Black individuals and in the KPT Hit Rate with regard to searches of Black individuals. Thus, it is recommended that Portland Police Bureau be examined in greater detail by STOP Program researchers and receive technical assistance from DPSST.

Table of Contents

Table of Contents

Appendices

List of Tables

Table of Contents

List of Figures

1. Background

1.1. HB 2355 (2017)

Efforts by the State of Oregon to collect data regarding stops of individuals made by law enforcement began with the passage of HB 2433 in 1997. While the text of the bill merely mandated that law enforcement agencies develop written policies that included provisions related to data collection, following the passage of HB 2433 the Governor's Public Safety Policy and Planning Council recommended that a full statewide data collection effort be initiated legislatively. It was not until 2001, however, that the Oregon Legislature again considered the collection of stop data. In SB 415 (2001), the Legislature created the Law Enforcement Contacts Policy & Data Review Committee (LECC), provided for the voluntary collection of stop data by agencies, and provided for analysis of collected data by the LECC.

With the exception of a brief hiatus from 2003 to 2005, the LECC engaged with law enforcement agencies throughout the 2000s and 2010s to examine stop data. During this period, however, challenges were encountered related to the creation of a comprehensive database of stops, given that few agencies in Oregon collected stop data and/or elected to partner with the LECC for data analysis. As a remedy, the Legislature passed HB 2355 in 2017, which led to the creation of the Oregon Statistical Transparency of Policing (STOP) Program. The STOP Program represents the culmination of the process started in 1997 and will result in the first statewide analysis of traffic[1] and pedestrian[2] stops in Oregon.

HB 2355, which is codified in ORS 131.930 et seq., created a statewide data collection effort for all officer initiated traffic and pedestrian stops that are not associated with calls for service. The aim of HB 2355 was to collect data regarding discretionary stops, as opposed to stops where discretion was absent. The Oregon Criminal Justice Commission, in partnership with the Oregon State Police and the Department of Justice, worked to develop a standardized method for collecting the data elements required by statute, which include data regarding both the stop itself as well as demographic characteristics of the stopped individual (for a description of the STOP Program data elements utilized in this report, see section 2.3.1.).

To implement the STOP Program, HB 2355 established a three-tiered approach, whereby the largest law enforcement agencies in the state would begin to collect data and report in the first year, followed by medium and small agencies in the next two years, respectively. **Table 1.1** reports the inclusion criteria for each tier as well as the data collection and reporting dates. A full list of agencies broken down by tier can be found in Appendix A.

Table 1.1. Three Tier Reporting Approach in HB 2355 (2017)

Tier	Number of Officers	Data Collection Begins	Reporting Begins
Tier 1	100+	July 1, 2018	July 1, 2019
Tier 2	25-99	July 1, 2019	July 1, 2020
Tier 3	1-24	July 1, 2020	July 1, 2021

1 Officer initiated traffic stops are defined as any "detention of a driver of a motor vehicle by a law enforcement officer, not associated with a call for service, for the purpose of investigating a suspected violation of the Oregon Vehicle Code" (ORS 131.930 § 4). Included with traffic stops are stops made of individuals operating bicycles. Stops involving operators of watercraft, however, are not included in the stop database, as watercraft violations fall outside the Oregon Vehicle Code (see ORS Chapter 830).

2 Officer initiated pedestrian stops are defined as "a detention of a pedestrian by a law enforcement officer that is not associated with a call for service. The term does not apply to detentions for routine searches performed at the point of entry to or exit from a controlled area" (ORS 131.930 § 3).

1. Background

In the development of the standardized data collection method, the primary goals of the STOP Program were to ensure that (1) all data collected are as accurate and complete as possible, (2) data collection methods are minimally impactful to each agency's workload and free or affordable for each agency, and (3) data collection methods are minimally impactful on law enforcement personnel to ensure that officer safety is not negatively impacted during the data collection process. As such, the STOP Program contracted with a technology vendor to develop software that could both collect and receive stop data via multiple submission methods.

The STOP Program software solution includes three methods of data collection/input. First, the software solution can receive data from local agencies' records management systems. Under this approach, an agency with the ability to collect stop data through its own preexisting systems can integrate stop data collection requirements into their in-car or e-ticketing system, recording the data internally before submitting the required data fields to the STOP Program in electronic format via a secure data connection. Second, for agencies that either cannot or choose not to integrate the required stop data fields into their preexisting systems, the STOP Program provides a free web application that can be loaded on officers' in-car computers (or other similar devices, like iPads) and used when a stop is made that requires data collection under the requirements in HB 2355. Third, and similar to the previous method, the STOP Program also provides mobile applications free of charge for both iPhones and Android phones, through which officers can submit stop data for qualifying police-citizen interactions.

2. Methodological Approach

2.1. Background

The formal examination of police traffic and pedestrian stop data began in the mid-1990s. For decades, advocacy groups cited anecdotal evidence supporting the notion that law enforcement applies different standards to minority drivers and pedestrians. Specific and systematic measurement of police practices during citizen stops, however, did not occur until court cases alleging racial bias in policing began to be litigated (see *Wilkins v. Maryland State Police* (1993) and *State of New Jersey v. Soto et al.* (1996)). Building on this foundation, the US Department of Justice, along with several other organizations, began hosting conferences related to the improvement of police-community relationships with a specific focus on the collection, analysis, and public reporting of traffic and pedestrian stop data. In response, many states began to mandate the collection of traffic stop data. In states that had yet to legislatively call for data collection, many local jurisdictions and departments began to collect and analyze stop data on their own.

During the approximately three decades that stop data have been studied, the majority of analyses have utilized and continue to utilize population-based benchmarks. This approach compares the demographic breakdown of stopped drivers and/or pedestrians to residential census data. Benchmarks are both intuitive and relatively simple to calculate, but the statistics that result from population-based benchmark comparisons are overly simplistic and often biased or wholly invalid (Neil and Winship 2018). The concerns regarding population-based benchmarks are many and discussed at length in academic research as well as in a companion research brief released by the STOP Program in 2018.[3] The central thrust of these critiques is that comparisons between stop data and residential benchmarks are invalid because the driving population in a given area (which forms the pool of individuals at risk for being stopped) is often unrelated to the residential population of that area. There are myriad reasons for this, including work commuting patterns, tourism, and others, all of which lead to a disjuncture between the residential demographics and the makeup of the driving population.[4]

2.2. Oregon STOP Program Analyses

To address the shortcomings of population-based benchmark analyses, researchers and statisticians have developed several sophisticated statistical approaches that allow for more precise, unbiased estimates of disparities in stop data. The STOP Program decided to utilize three of these analyses. The decision to utilize multiple tests was based on two factors. First, the nature of traffic and pedestrian stops necessitates the use of multiple tests. Initially, it is tempting to view a stop as a single instance of police-citizen contact that can be assessed for the presence or absence of discriminatory behavior by a law enforcement agent. Within the time it takes to execute and conclude a single stop, however, there are numerous opportunities where racially disparate treatment may be present. Beyond the initial contact between a citizen and a law enforcement officer, for instance, race/ethnicity could be a factor in the decision to search a driver, give a citation, or make an arrest. This distinction is critical, because both the data and analytical techniques required to analyze the various decision points found in a single officer initiated stop differ, and STOP Program researchers' data and analyses must (and do) address each of these decision points specifically. While not every stop recorded in the data will contain each decision point (e.g. most stops do not result in arrest and are thus not usable for analysis of arrest

3 See STOP Program Research Brief: Analytical Approaches to Studying Stops Data (October 2018), which can be found at www.oregon.gov/cjc

4 Using 2017 Census data via https://onthemap.ces.census.gov, it is possible to view the impact that work commuting has on Oregon cities and thus to understand the possible scope of the disjuncture between the driving population and residential census population of a given area. In Portland, for instance, the Census estimates that over 240,000 individuals commute into the city for work each day (about 60 percent of the city's workforce).In Beaverton, this pattern is even more pronounced, as over 85 percent of individuals working in Beaverton commute in from outside the city. Notably, commuting patterns do not just affect the Portland metro area, as Eugene, for example, displays a similar pattern. Specifically, it is estimated that 65 percent of individuals working in Eugene, approximately 91,000 people, commute into the city for work each day.

patterns), in aggregate, the stops recorded give information about every possible decision point in an interaction, and the analyses performed utilize that fact.

Second, while the statistical tests utilized by the STOP Program represent the gold standard[5] in law enforcement stop data analyses, the application of multiple tests is also necessary to address the possibility that any single analysis could produce false positives or false negatives. Statistics, by their very nature, are estimates, and it is always possible that some degree of error could influence results, whether stemming from data collection strategies or decisions, errors in reporting, or the like.

The three analyses utilized by the STOP Program are:[6]

Veil of Darkness Analysis. The Veil of Darkness test takes advantage of natural variations in daylight and darkness throughout the year to examine the initial decision to stop an individual. Based on the assumption that it is easier for an officer to discern race/ethnicity during the day when it is light versus the night when it is dark, this analysis compares stop rates for minority individuals to those for White individuals during the time windows surrounding sunrise and sunset. If, as demonstrated by the statistics that result from the Veil of Darkness test, minority individuals are more likely to be stopped in the daylight when race/ethnicity is easier to detect, then there would be evidence of a disparity.

Predicted Disposition Analysis. The Predicted Disposition test examines matched groups using a statistical technique called propensity score analysis to explore whether disparities exist in stop outcomes (i.e., citations, searches, or arrests). In essence, this test matches stop data on all available characteristics, only allowing race/ethnicity to vary between the two groups being compared. This means that the analysis compares White and Black groups, for example, who have identical proportions of gender, age, stop time of the day and day of the week, reason for the stop, season, whether the stop was made in the daylight, and agency and county stop volumes to determine whether one group is cited more often, searched more often, or arrested more often. If, after matching on all of the factors listed above, minority individuals are either cited, searched, or arrested more often than similarly situated White individuals, then there would be evidence of a disparity.

Hit-Rate Analysis. The Hit-Rate test compares relative rates of successful searches (i.e., those resulting in the seizure of contraband) across racial/ethnic groups. It is based on the assumption that if search decisions by officers are made based on race/ethnicity neutral criteria, then success rates should be similar, if not identical, across different racial/ethnic categories. If, however, search success rates differ and the search success rates for minority individuals are significantly lower than those reported for White individuals, then there would be evidence of a disparity.

2.3. Analytical Sample

2.3.1. Data Elements

A total of 396,612 records were submitted by the twelve Tier 1 agencies during the first year of data collection. As required by HB 2355 (2017), agencies submit numerous data points, including information regarding the stop itself as well as information regarding the stopped individual. While HB 2355 is clear regarding the data elements the STOP Program is required to collect, it did not define these elements. To fill this gap, the Oregon State Police assembled a group of stakeholders to formally define the data elements contained in the statute,

5 The analytical approach utilized by the STOP Program is based on the work conducted by the Connecticut Racial Profiling Prohibition Project, which employs research and analytical techniques that have been peer reviewed by academics who specialize in the study of racial/ethnic disparities in law enforcement contacts.

6 More detailed, technical descriptions of these analyses can be found in Appendices C, D, and E.

which included representatives from law enforcement, community groups, state agencies, and the Oregon Legislature.

Date and Time the Stop Occurred. Law enforcement personnel are required to record the date (month/day/year) and time that the stop occurred. Stop times are recorded on a 24-hour clock ("military time") and converted to 12-hour clock time for this report.

Type of Stop. As required by HB 2355, both traffic and pedestrian stops are reported by law enforcement. Included in the database is a binary variable denoting whether the record is for a traffic or pedestrian stop. During the analysis of this data element, it was discovered that in a small number of cases, some stops were coded as "pedestrian" that were clearly for moving or other traffic violations. Similarly, some stops were coded as "traffic" that were clearly violations by pedestrians. These stops were recoded by STOP Program researchers to the appropriate categories.[7]

Perceived Race/Ethnicity of Subject. Law enforcement officers are required by HB 2355 to record their perception of a subject's race/ethnicity (for traffic stops, only the perceived race/ethnicity of the driver is reported). The categories included in the data collection are: White, Black, Hispanic, Asian or Pacific Islander (hereinafter, Asian/PI), Native American, and Middle Eastern.

Perceived Gender of Subject. Law enforcement officers are required by HB 2355 to record their perception of a subject's gender (for traffic stops, only the perceived gender of the driver is reported). The categories included in the data collection are: male, female, and non-binary.

Perceived Age of Subject. Law enforcement officers are required by HB 2355 to record their perception of a subject's age, which is entered as a whole number (for traffic stops, only the perceived age of the driver is reported).

Legal Basis for the Stop. The legal basis for each stop is reported to the STOP Program. This includes violations of: an Oregon statute, a municipal traffic code, a municipal criminal code, a county code, tri-met rules/regulations, or a Federal statute.

Oregon Statutory Violations Detail. For violations of Oregon statute, which represents over 90 percent of all stops, law enforcement provide the specific ORS code corresponding to the violation. In this data element, over 700 different ORS codes were reported during the first year of data collection. To simplify the use of this information in the models conducted in the remainder of this report, the STOP Program research team aggregated these violations into the following categories: moving violations; equipment, cell phone, and seat belt violations; registration and license violations; and "other" violations (e.g., criminal offenses, camping violations).[8]

Disposition of the Stop. The most serious disposition for each stop is reported by law enforcement officers. The categories included in the data collection are: nothing, warning, citation, juvenile summons, and arrest. It is important to note that stops can have multiple dispositions (e.g., an individual could be both cited for a traffic violation and arrested for a crime), however, only the most serious disposition is reported into the STOP Program database. This means that the categories for warnings, citations, and juvenile summons could be undercounted. For the analyses examining stop disposition in this report, the juvenile summons category was removed from the data set because the data included only 93 juvenile summons (0.02 percent of all dispositions).

7 For instance, approximately 300 stops labeled as "pedestrian" were for speeding, with a small subset stopped for speeding more than 21 miles per hour over the limit. Alternatively, almost 60 "traffic" stops were for ORS 814.020, which is specifically when a pedestrian fails to obey a traffic control device. For a full list of these inconsistent stops, please contact the STOP Program.

8 Details on the offenses falling into each category are available upon request.

Whether a Search was Conducted. Law enforcement officers utilize a binary variable to report whether a search was conducted to the STOP Program database.

Justification for the Search. Law enforcement officers can provide several bases for a search using the following categories: consent search, consent search denied, or "other" search. The "other" search category includes frisks, probable cause searches, and other administrative searches. Multiple data points are allowed so that the data can include several search justifications. For example, if an officer initially requests to search an individual but consent is not given, an officer may then perform a search based on probable cause. In this example, the officer could record both "consent search denied" as well as "other search" into the database.

Search Findings. Seven categories were predefined by the STOP Program stakeholder engagement group with regard to search findings. These categories are: nothing, alcohol, drugs, stolen property, weapon(s), other evidence, and other non-evidence. Officers are permitted to report up to six search findings to the STOP database so that searches resulting in the seizure of multiple types of contraband are properly documented.

Stop Location. Law enforcement officers are required by HB 2355 to record the location of the stop. The form in which these data are submitted varies by agency. Some agencies report X,Y coordinates, while others submit textual descriptions of the location (e.g., 123 Main Street, intersection of Main and Maple Streets).

Finally, the STOP Program created two of its own variables for use in its analyses. Following best practices, variables representing both the *daily agency stop volume* and *daily county stop volume* were created. For agency stop volume, the aggregate number of stops for a single date are divided by the maximum number of daily stops for the agency unit in question. Thus, if an agency stopped 1,000 drivers on its busiest day, this would be the denominator against which all other days would be compared. A measure of the county stop volume would be calculated the same way, although all stops made by agencies within a single county would be included together.

2.3.2. Sample

While the overall number of records was substantial, the STOP Program team faced challenges with regard to sample size when the data were broken down into subsamples based on race/ethnicity. In cases where the sample size is too small, statistical analyses cannot be conducted due to the increased risk of a false negative or false positive result. For instance, while one agency stopped a total of 339 Black individuals during the past year only 5 of those 339 stops resulted in a search, which is an insufficient number of observations to conduct the Hit-Rate analysis. The small sample size issue is further complicated by the fact that two of the three analyses are multivariate, so that each subset of variables must also meet the appropriate thresholds to ensure that results are as accurate as possible. For example, if an analysis includes measures for race/ethnicity, gender, and stop reason (e.g., moving violation, equipment violation), then an adequate number of observations must be present for each combination of the three included variables in order to perform the analysis.

To determine appropriate thresholds for sample size, the STOP Program relied on established criteria set in the academic and professional literature. Drawing on standards described by Wilson, Voorhis, and Morgan (2007), the STOP Program used the following sample size thresholds reported in **Table 2.3.1**.

The sample size issue identified above had a significant impact on the STOP Program research team's ability to conduct analyses on all of the racial/ethnic groups found in the stop database. **Table 2.3.2** reports the breakdown by race/ethnicity and agency for Tier 1 agencies. As shown in **Table 2.3.2**, in several cases the total number of stopped individuals for certain racial/ethnic groups falls under the thresholds defined in **Table 2.3.1**. Further, once the STOP Program research team began to analyze subsets of the data (e.g., only those individuals who were searched, or arrested; those observations that met the standards to be included in the Veil of Darkness), the counts for these groups quickly fell under the thresholds.

Table 2.3.1.
Sample Size Thresholds for Conducting Statistical Analyses

Statistical Test	Sample Size Threshold
Veil of Darkness	Minimum of 100 observations for an individual racial/ethnic group[9]
Predicted Disposition	Minimum of 100 observations for an individual racial/ethnic group[10]
Hit-Rate	Minimum 30 observations per racial/ethnic group analyzed; no cell with less than 5 observations

Table 2.3.2.
Race/Ethnicity Reporting by Tier 1 Agency for All Reported Stops

Agency Name	Asian/PI	Black	Hispanic	Middle Eastern	Native American	White
Beaverton PD	1,215	1,577	3,411	525	166	13,676
Clackamas Co SO	933	1,028	2,486	311	188	20,725
Eugene PD	269	634	597	0	40	9,581
Gresham PD	416	1,375	1,439	111	55	6,051
Hillsboro PD	733	639	2,506	413	52	7,499
Marion Co SO	383	339	2,292	152	8	10,638
Medford PD	71	270	627	11	19	4,497
Multnomah Co SO	422	945	1,322	132	50	7,914
Oregon State Police	4,829	5,624	21,875	2,490	1,005	172,037
Portland PB	1,788	5,410	3,044	463	164	20,194
Salem PD	270	341	2,223	49	54	7,442
Washington Co SO	1,715	1,525	5,207	786	404	19,948
Tier 1 Total	**13,044**	**19,707**	**47,029**	**5,443**	**2,205**	**300,202**

Related to the concerns regarding sample size more generally, STOP Program researchers faced similar issues with pedestrian stops. Across all Tier 1 agencies, only 2.8 percent of stops, which represents 11,101 individual encounters, were pedestrian stops. When pedestrian stops are broken down by agency, in nearly all instances it is not possible to estimate models examining these stops on their own. Further, when agency-level pedestrian

9 Wilson, Voorhis, and Morgan (2007: 48) recommend that for regression equations where six or more variables are included in the model, "an absolute minimum of 10 participants per predictor variable is appropriate." While this is the minimum, if possible, they recommend 30 participants per predictor. Further, in instances where the outcome variable is skewed due to the small sizes of minority groups relative to the White group, which is certainly the case in many of the STOP research team's analyses, larger sample sizes are needed. For the analyses in this report, the STOP research team elected to use the 10 participant minimum, which when multiplied by 10 predictor variables sets the minimum number of observation for an individual racial/ethnic group at 100.

10 In some instances, despite having the minimum number of observations required to run a model, the models did not converge when estimated in Stata.

stops are disaggregated by race/ethnicity, the problem only becomes more acute. For instance, only three agencies stopped more than 100 Black pedestrians or 100 Hispanic pedestrians. With regard to Asian/PI, Native American, or Middle Eastern pedestrians, no agency reported more than 50 stops. Indeed, the average number of stops for any of these groups at the agency level was a little less than eight.

Due to these sample size limitations and concerns, statistical models at the state level are conducted for all racial/ethnic groups where possible, while models for agency specific analyses are limited to comparisons between White, Black, and Hispanic individuals, and a combined sample of Black and Hispanic individuals. In the future, the STOP Program team will conduct agency specific analyses of Asian/PI, Native American, and Middle Eastern individuals when possible by combining several years of data. Unfortunately, this is an issue that will not be limited to Tier 1, as it is likely that Tier 2 and Tier 3 agencies will require significant passage of time and data collection before samples large enough for robust analyses are obtained, due to their inherently lower overall stop volumes. Finally, due to the sample size issue discussed, pedestrian and traffic stops were analyzed together in this report for all post-stop outcomes.[11]

A final concern in any analysis is the presence of missing data. Missing data in the context of the STOP Program could come from two sources. First, a data point could be missing because it was never entered. Second, a data point could be submitted in an invalid format which lacks the information necessary to determine where it fits into the STOP Program data schema. In the STOP Program data, missing data attributable to both of these sources were found, although the total amount of missing data in the STOP database is low compared to the overall sample size and is not a threat to the validity of the statistical models from a methodological point of view. Please see Appendix B for more details.

2.4. Threshold for Statistical Significance

To determine if disparities identified in this report warrant additional in-depth analysis and/or technical assistance from the Oregon Department of Public Safety Standards and Training (DPSST), STOP Program researchers reviewed the results of each of the three analyses conducted on the STOP Program data. For each individual analysis, an estimated disparity must meet the 95 percent confidence level for it to be statistically significant.[12] This means, in simple terms, that the STOP Program research team must be at least 95 percent confident that differences or disparities identified by the analyses were not due to random chance. In some cases, confidence in the reported results exceeded the 95 percent confidence threshold.

Beyond the 95 percent confidence threshold for each individual analysis, STOP Program researchers also established a threshold at which identified disparities warrant further investigation and technical assistance from DPSST at the project level. Following best practices and the "gold standard" analyses conducted by the State of Connecticut,[13] for a law enforcement agency to be identified as one requiring further analysis as well as DPSST technical assistance, it must be identified as having a statistically significant disparity in two of the three

11 As the STOP Program database grows, it is likely that robust samples for pedestrian stops will be obtained. Once thresholds are met, these stops will be analyzed separately from traffic stops in future reports.

12 Given that multiple comparisons were made for each test (e.g., Black-White, Hispanic-White, Black/Hispanic-White), Bonferroni adjustments were utilized to adjust for the likelihood of a given test resulting in a false positive. The Bonferroni adjustment differed for each test, and is described in greater detail in Sections 4, 5, and 6, as well as in the corresponding technical appendices.

13 The Connecticut Racial Profiling Prohibition Project is located at http://www.ctrp3.org/

analytical tests performed on the STOP data.[14] The justification for this approach mirrors the reasoning behind the utilization of multiple tests to examine the data acquired for this project. As discussed previously, given that the statistical outputs provided in this report in many instances are estimates which could lead to false positives or false negatives in any single analysis, best practices suggest that caution should be taken when examining and interpreting results from the statistical tests performed.

2.5. Limitations

The data collected by the STOP Program for the State of Oregon represent one of the most robust stop data collection efforts in the United States. While data are collected by some jurisdictions in the vast majority of states, few states can boast a statewide, statutorily mandated data collection effort like the one Oregon will have in the coming years. This robust database and the statistical evaluation of stop data can form the foundation of a transparent dialogue between state leaders and agencies, law enforcement, and the communities law enforcement agencies serve.

In spite of its promise as a means for systematically analyzing statewide data concerning police-citizen interactions, the STOP Program and its associated data and analyses have limitations. First, the statistical analyses can only identify *disparities* in police/citizen interactions. This means that the analyses contained in this report cannot be used either as absolute proof that a law enforcement agency engaged in racially biased conduct or as disproof of racially biased conduct. Further, the results in this report are conducted at the department level because HB 2355 expressly forbids the collection of data that identify either stopped individuals or officers. These analyses, therefore, can only identify systematic disparities across a department. As such, regardless of whether a department is reported to have an identified disparity or not, this report cannot and does not discount or speak to the personal experiences of individuals who have been subjected to biased treatment.

In spite of these limitations, the statistics presented in the following sections demonstrate that after the application of rigorous standards, if multiple disparities are identified for an agency then there is cause for concern, further investigation, and technical assistance. STOP Program researchers have selected highly respected, thoroughly vetted and peer reviewed, cutting-edge analyses. The STOP Program stands behind the significant amount of work that went into the analyses and crafting this report and believes that the results presented herein will provide a first step in starting a dialogue between law enforcement and the citizens of the State of Oregon.

14 The State of Connecticut applies a sliding scale in its analyses, whereby a disparity identified via the Veil of Darkness analysis alone results in an agency being identified for further analysis. For its other analyses, two or more identified disparities results in further analysis. Unlike Connecticut, the Oregon STOP Program treats all three of its analyses as coequal while retaining the two out of three threshold. Importantly, even if the Connecticut threshold was used for the STOP Program analyses, the referral results would be the same.

3. Characteristics of the Tier 1 Stop Data

This report analyzes data collected by the STOP Program for officer-initiated traffic and pedestrian stops occurring in Oregon from July 1, 2018 through June 30, 2019. In total, 396,612 stops were submitted to the STOP Program by the twelve Tier 1 agencies. The number of stops reported by each agency is displayed in **Table 3.1**. There was significant variation in the frequency with which Tier 1 agencies stopped individuals. The Oregon State Police, for instance, made 54.2 percent of stops in the state, the largest number reported by any of the twelve agencies, whereas Medford Police made the fewest stops, accounting for only 1.39 percent of the data reported.

Table 3.1.
Percent and Number of Tier 1 Agency Stops by Stop Type Traffic vs. Pedestrian

Agency	Traffic Stops		Pedestrian Stops		Total Stops	
Beaverton PD	93.8%	19,319	6.2%	1,283	5.19%	20,602
Clackamas Co SO	95.4%	24,516	4.6%	1,175	6.48%	25,691
Eugene PD	77.2%	8,601	22.8%	2,546	2.81%	11,147
Gresham PD	95.9%	9,061	4.1%	388	2.38%	9,449
Hillsboro PD	94.1%	11,167	5.9%	699	2.99%	11,866
Marion Co SO	99.8%	15,318	0.2%	37	3.87%	15,355
Medford PD	68.3%	3,754	31.7%	1,743	1.39%	5,497
Multnomah Co SO	96.1%	10,375	3.9%	423	2.72%	10,798
Oregon State Police	99.5%	213,944	0.5%	1,028	54.20%	214,972
Portland PB	98.4%	30,592	1.6%	493	7.84%	31,085
Salem PD	95.3%	10,043	4.7%	490	2.66%	10,533
Washington Co SO	97.3%	28,821	2.7%	796	7.47%	29,617
Tier 1 Total	**97.2%**	**385,511**	**2.8%**	**11,101**	**100.00%**	**396,612**

Table 3.1 reports the number and percentage of stops by agency broken down by stop type (traffic or pedestrian). Across all of the Tier 1 agencies, only 2.8 percent of stops were of pedestrians. By agency, the frequency with which pedestrian stops were made, as well as the degree to which those stops impacted a department's overall stop profile, varied significantly. Medford Police Department and Eugene Police Department reported the largest percentage of pedestrian stops; Medford Police Department's pedestrian stops accounted for around 32 percent of all their reported contacts and Eugene Police Department's pedestrian stops accounted for approximately 23 percent of their reported contacts. As for raw numbers, Eugene Police Department, Beaverton Police Department, and Medford Police Department made the most pedestrian stops. Marion County Sheriff's Office, the Oregon State Police, and Portland Police Bureau recorded the fewest pedestrian stops.

The demographic breakdowns for traffic and pedestrian stops are reported in **Table 3.2**. For the Tier 1 agencies contained in this report, the vast majority of stops were of White drivers/pedestrians, with Hispanic and Black individuals being the two most frequently stopped minority groups. This pattern held when broken down by traffic versus pedestrian stops, although White individuals made up an even higher proportion of pedestrians. With regard to gender, more males were stopped than females. When examined by stop type, the male/female divide became even more pronounced, as the vast majority of pedestrian stops involved male individuals. Less

Table 3.2.
Aggregate Tier 1 Demographics by Stop Type

Race/Ethnicity	Traffic	Pedestrian	Total
Asian/PI	3.4%	1.4%	3.4%
Black	5.0%	7.1%	5.1%
Hispanic	12.3%	8.0%	12.1%
Middle Eastern	1.4%	0.4%	1.4%
Native American	0.6%	0.7%	0.6%
White	77.3%	82.5%	77.4%
Gender			
Male	66.0%	82.0%	66.5%
Female	33.5%	17.2%	33.1%
Non-Binary	0.5%	0.9%	0.5%
Age			
Under 21	10.1%	9.4%	10.1%
21 – 29	24.5%	20.7%	24.4%
30 – 39	24.6%	31.4%	24.8%
40 - 49	16.7%	18.8%	16.8%
50 and Older	24.0%	19.8%	23.9%

than 1 percent of either traffic or pedestrian stops were of an individual perceived to be non-binary. Most traffic and pedestrian stops of are of individuals perceived to be aged in their thirties, more so for pedestrians.

Figure 3.1 displays the number of traffic and pedestrian stops by the month of the year and racial/ethnic category. The highest number of stops occurred in March and September, while the lowest number of stops occurred in February, October, November, and December. There is little variation in the racial/ethnic makeup of stops over the course of the year. Consistent with the data reported in **Table 3.2**, Tier 1 agencies stopped mostly White and Hispanic individuals, and a smaller proportion of all other races/ethnicities throughout the year.

Figure 3.1.
Traffic and Pedestrian Stops by Month of Year

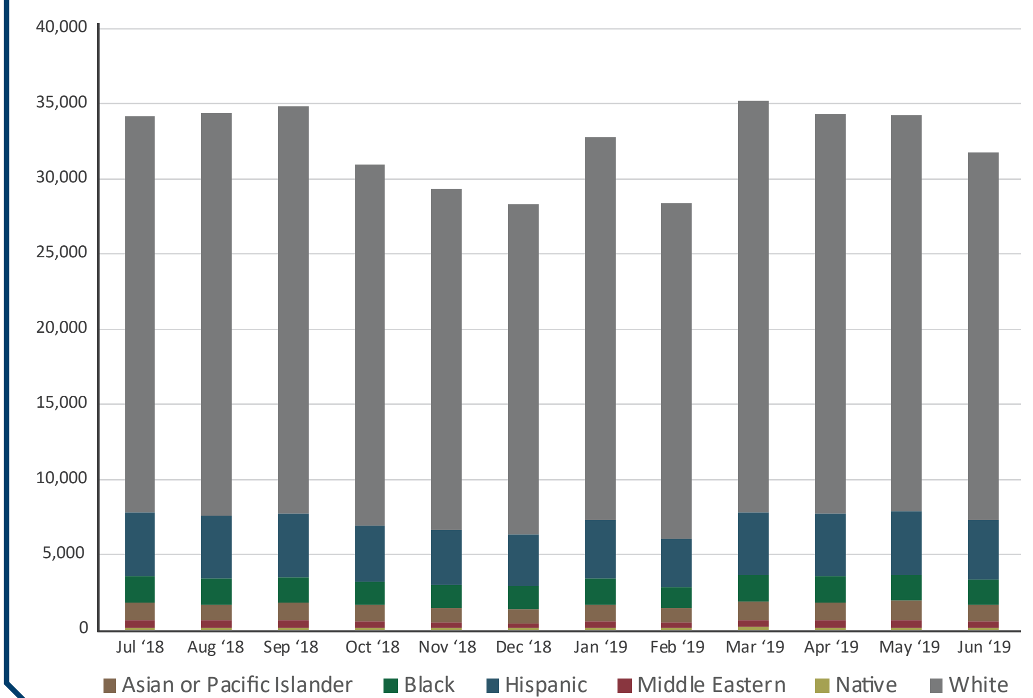


Figure 3.2 depicts traffic and pedestrian stops stratified by time of day. Few stops occur from 3:00 am to 6:00 am. About five times as many stops are reported during evening commuting times, from around 3:00 pm to 7:00 pm. In general, the demographic trends across time of day are constant except during morning commuting times. Specifically, a higher proportion of minority stops are reported from 7:00 am to 9:00 am compared to other times of the day.

Figure 3.2.
Traffic and pedestrian Stops by Time of Day

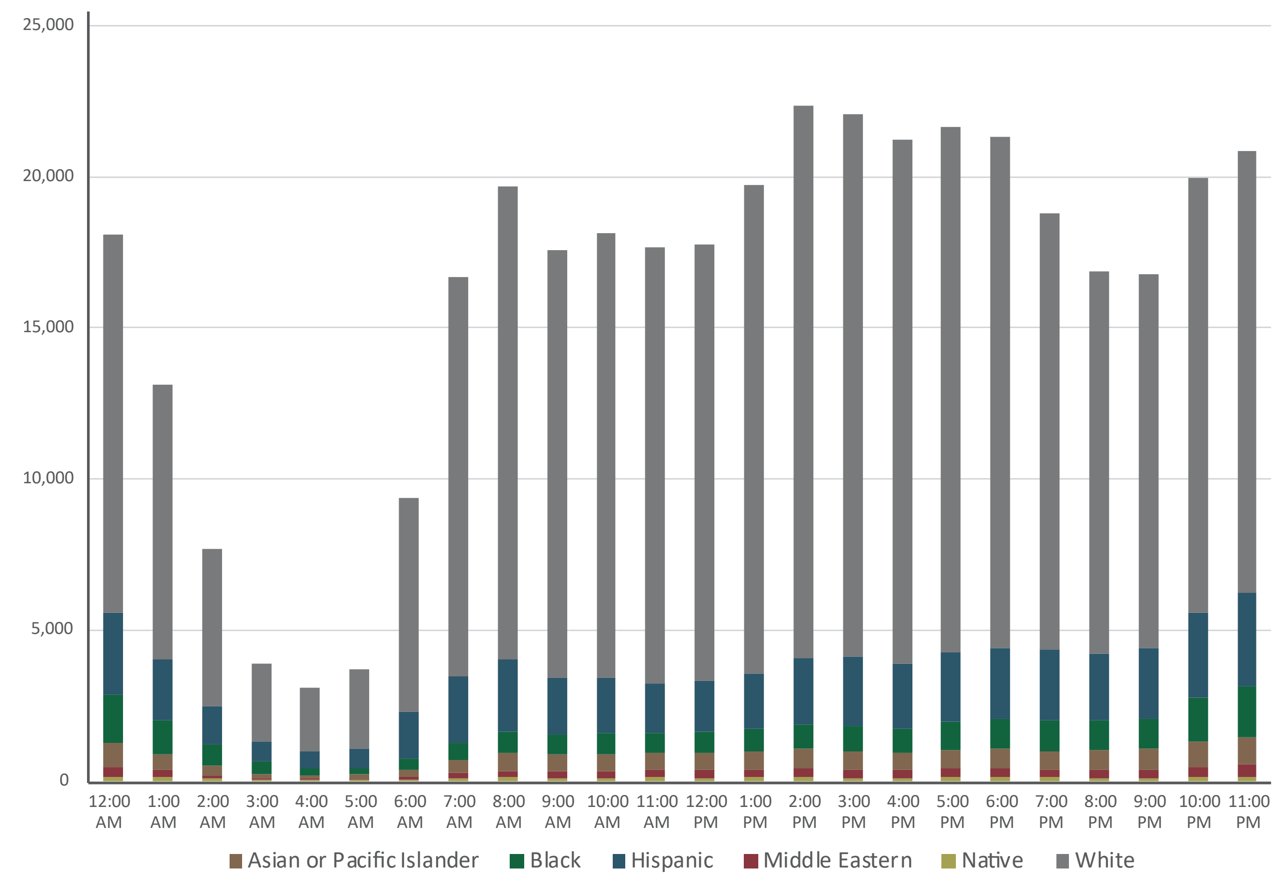


Table 3.3 displays the most serious dispositions reported by law enforcement. In all, most police stops do not result in further action taken against the stopped individual. While less than 3 percent of stopped individuals received no warning or sanction, the most common outcome of a police stop regardless of type was a warning. It is important to note that it is the policy of many agencies to give a warning to everyone who is stopped. With regard to sanctions, 37 percent of stopped individuals were given a citation and less than 3 percent of stops end in an arrest.

Table 3.3.
Aggregate Tier 1 STOP Disposition by Stop Type

Disposition	Traffic	Pedestrian	Total
None	2.4%	17.3%	2.9%
Warning	57.4%	52.6%	57.3%
Citation	37.6%	13.2%	37.0%
Juvenile Summons	0.0%	0.1%	0.0%
Arrest	2.5%	16.7%	2.9%

When disaggregated by stop type, there are some clear differences in stop disposition. Pedestrian stops, for instance, were more likely to result in no action/sanction compared to traffic stops, as nearly one fifth of all pedestrian stops fell into this category. Alternatively, pedestrian stops were much more likely to result in an arrest compared to traffic stops.

Finally, **Table 3.4** provides aggregate Tier 1 data concerning searches. Overall, nearly 3 percent of stops resulted in a search of some type. When examined by stop type, however, there was a stark difference in searches, as pedestrians were searched in over 16 percent of stops while drivers were only searched in 2.5 percent of stops. With regard to the justification for search, overall there was nearly an even split between consent searches and "other" searches, which include frisks, searches based on probable cause, and the like. When broken down by traffic versus pedestrian stops, however, it was clear that searches occurring during traffic stops were more likely to be justified by the driver providing consent to the search. For pedestrian searches, alternatively, consent was provided in less than 50 percent of cases, with the remainder being justified on some other legal basis. Finally, with regard to search success, a breakdown of search findings is also reported. It is important to note that searches can result in findings in multiple categories. For example, an individual could be searched and the officer could find alcohol, drugs, and a weapon. In all, alcohol and drugs were found most often during the reported searches.

Table 3.4.
Aggregate Tier 1 STOP Search Data

Variable	Traffic	Pedestrian	Total
Search Conducted	2.5%	16.3%	2.9%
Search Justification			
Consent Search	56.5%	46.7%	54.5%
"Other" Search	43.5%	56.3%	45.6%
Successful in Total	45.4%	39.6%	44.5%
Alcohol Found	12.8%	6.0%	11.7%
Drugs Found	27.1%	25.6%	26.9%
Weapons Found	5.2%	5.9%	5.3%
Stolen Property Found	1.6%	3.5%	1.9%
Other Evidence Found	7.7%	7.2%	7.6%
Other Non-Evidence Found	2.1%	3.8%	2.4%

4. Veil of Darkness Analysis

Often referred to as the "gold standard" of statistical analyses examining the initial law enforcement decision to stop an individual,[15] the Veil of Darkness (VOD) analysis compares stops made by law enforcement officers during the day when it is light to those made at night when it is dark to test for disparities when officers can more easily perceive the race/ethnicity of drivers. This VOD analysis is built on the assumption that officers can better detect the race/ethnicity of an individual in daylight as compared to darkness. The chief advantage to this approach is that the analysis does not rely on a benchmark comparison with the estimated driving or residential population to the population of stopped individuals. Rather, the VOD analysis takes advantage of natural variations in daylight over the course of the year to compare minority stops made in daylight to those made in darkness at similar times of the day when commuting patterns should be relatively consistent.

More specifically, the VOD analysis relies on comparing the racial composition of individuals stopped during a combined inter-twilight window, which occurs during morning and evening commute times. The morning twilight window is defined as the earliest start of civil twilight to the latest sunrise, while the evening twilight window is defined as the earliest sunset to the latest end of civil twilight. Visibility during this time will vary throughout the course of the year, which makes it possible to compare stop decisions at the same time of day but in different lighting conditions. For example, the VOD analysis can compare stops made on January 10, 2019 when it was dark at 5:00pm to stops made two months later at the same time on March 10, 2019, when it was still light outside. Given that these two points in time should capture substantially similar driving populations, comparisons made between the race/ethnicity of stopped drivers in the light and darkness will detect whether stops are being made in a disparate fashion when race/ethnicity is visible.

Beyond this central assumption underlying the VOD approach, the analytical test also assumes that driving behavior does not change throughout the year or between daylight and darkness, and that driving patterns have little seasonal variation during the morning and evening commute times. While this assumption is likely too strong and not reflective of actual driving patterns, it can be accounted for statistically by including additional control variables available in the STOP Program database, including: age, gender, reason for stop, day of week, time of day, quarter or season, county stop volume, and agency stop volume.

To accomplish the analysis described above, the VOD approach tests whether the odds of non-White traffic stops during daylight are significantly different from the odds of non-White traffic stops during darkness. In the tables that follow in the next subsection, this difference in odds is presented as an *odds ratio*, which displays the change in odds for non-White stops during daylight compared to darkness. If the odds ratio is not statistically different from 1.0, then the test finds no difference in stops made during daylight and darkness. If the odds ratio is greater than 1.0 and statistically significant, however, the test concludes the odds of non-White drivers being stopped in daylight is significantly higher than in darkness, which is taken as evidence of a racial disparity in stops, after accounting for additional control variables that are available in the stop data. Conversely, if the odds ratio is less than 1.0 and statistically significant, the odds of a non-White driver being stopped in daylight is significantly lower than in darkness. In sum, following best practices, the STOP Program identifies all agencies with disparities above 1.0 that are statistically significant at the 95 percent confidence level in any minority group at the aggregate, Tier 1 level, or for the Black or Hispanic alone groups at the agency level.

4.1. Aggregate Veil of Darkness Analysis for All Tier 1 Agencies

At the statewide level, it is possible to estimate VOD models for all of the non-White groups reported in the stop database. First, **Table 4.1.1** displays the odds ratios for the aggregate Tier 1 VOD models for all non-White

15 *See* Barone et al. (2018).

stopped drivers, including those perceived as Black, Hispanic, Asian/PI, Middle Eastern, and Native American, compared to White stopped drivers. For the full statewide models, all comparisons show no statistically significant differences in the odds of minority stops in daylight compared to darkness. Further, the coefficients for nearly all comparisons are less than 1.0. Taken together, this indicates that disparities were not detected at a statewide level.

Table 4.1.1.
Logistic Regression of Minority Status on Daylight for All Tier 1 Agencies

	Asian/PI	Black	Hispanic	Black/ Hispanic	Middle Eastern	Native American
Tier 1	0.96	0.93	1.05	0.89	0.86	0.92
Tier 1 (no OSP)	0.90	0.90	0.95	1.02	0.81	0.74

NOTES: * p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction with 5 Comparisons)
Logistic regression results include controls for age, gender, reason for stop, day of week, time of day, quarter or season, county stop volume, agency stop volume, and agency fixed effects.

To check the robustness of the statewide finding reported in the first row of **Table 4.1.1**, an additional analysis was conducted. In the data reported by law enforcement for the first year of data collection, stops made by the Oregon State Police accounted for over half of all traffic stops. Due to this imbalance, it is possible that patterns in Oregon State Police stop rates could mask disparities occurring in the other eleven agencies when aggregated to the state level. Due to this concern, a VOD model was estimated for all stops other than those made by the Oregon State Police. The results of this robustness check, reported in the second row of **Table 4.1.1**, demonstrate that even with the Oregon State Police removed from the analysis, there is no evidence of a racial disparity at the state level for non-White drivers.

An additional robustness check for the VOD analysis is possible when a sufficient number of data points are found in a given dataset. As discussed above, one assumption of the VOD model is that driving behavior and patterns does not vary throughout the year. It is possible, however, that patterns in the summer, for instance, are different than those found in the winter. To account for this possibility, the VOD analysis can be restricted to stops made during the 30-day windows before and after the fall and spring Daylight Savings Time (DST) dates. This ensures that comparisons of stops in daylight and darkness are made between dates that are relatively close to one another, which narrows the possibility that variation in driving behavior and driving patterns could impact the results of the statistical test. The results of these models are reported in **Table 4.1.2** for Black, Hispanic, and a combined sample of Black

Table 4.1.2.
Logistic Regression of Minority Status on Daylight for All Tier 1 Agencies Restricted to DST Window

	Black	Hispanic	Combined
Tier 1	0.90	0.89	0.89
Tier 1 (no OSP)	0.85	0.92	0.89

NOTES: * p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction with 3 Comparisons)
DST (Daylight Saving Time) window is +/- 30 days around fall and spring DST dates
Logistic regression results include controls for age, gender, rea-son for stop, day of week, time of day, quarter or season, county stop volume, agency stop volume, and agency fixed effects.

or Hispanic Drivers.[16] Similar to the results presented in **Table 4.1.1** above, the results for the restricted sample show no evidence of a statistically significant difference in stops during the daylight compared to those at night.

4.2. Agency-level Veil of Darkness Analysis

While the aggregate, state level results indicate that there is no evidence of a statistically significant disparity in outcomes for any minority group compared to White individuals, it is still possible that analyses of individual law enforcement agencies could detect disparities at a disaggregated level. As such, VOD models were estimated for all twelve Tier 1 agencies for Black, Hispanic, and the combined sample of Black or Hispanic drivers. At this time, sample sizes were insufficient to estimate models for Asian/PI drivers, Middle Eastern drivers, or Native American drivers by agency. In addition, the sample sizes were insufficient to estimate the VOD models for Black drivers for Marion County Sheriff's Office, Medford Police Department, and Salem Police Department. As described in Section 2, the sample size requirement for the VOD model was at least 100 stops in each race/ethnicity group within the inter-twilight windows.

Table 4.2.1.
Logistic Regression of Minority Status on Daylight by Tier 1 Agency

Agency	Black	Hispanic	Combined
Beaverton PD	0.90	1.19	1.10
Clackamas Co SO	1.34	1.12	1.18
Eugene PD	0.58	1.18	0.83
Gresham PD	0.77	0.62	0.70
Hillsboro PD	1.02	0.87	0.89
Marion Co SO	N/A	0.82	0.92
Medford PD	N/A	0.78	0.64
Multnomah Co SO	0.93	1.00	0.97
Oregon State Police	1.04	0.95	0.96
Portland PB	0.86	0.77	0.83*
Salem PD	N/A	1.18	1.22
Washington Co SO	0.89	0.82	0.84

NOTES: * p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction with 3 Comparisons)
Logistic regression results include controls for age, gender, reason for stop, day of week, time of day, quarter or season, county stop volume, agency stop volume, and agency fixed effects.

Table 4.2.1 reports the agency specific model results for Black, Hispanic, and the combined sample of Black or Hispanic drivers compared to White drivers. While a number of agencies have odds ratios above 1.0, which tends to indicate the possibility of a disparity, no agency had an odds ratio that was above 1.0 and statistically significant. Indeed, only Portland Police Bureau reported a statistically significant outcome of any type, but it was for an odds ratio that was below 1.0, which indicates that for the combined sample of Black and Hispanic individuals, the odds of drivers being stopped were higher in the darkness than in daylight. Thus, taken together, the STOP Program did not find evidence of racial/ethnic disparities in the Veil of Darkness analyses.[18]

---

16 To conduct a VOD analysis limited to the 30-day windows around daylight savings time changes, the sample size must be sufficiently large. Unfortunately, at this time, samples sizes were only sufficient to make comparisons of Black, Hispanic, and Black or Hispanic drivers to White drivers. In future reports, as the number of stopped individuals of other racial/ethnic categories grows, it will be possible to run models of this type.

17 During the process of analyzing the Tier 1 data collected during the first year of reporting, it was determined that there was an error in reporting by the Multnomah County Sheriff's Office. Specifically, the data submitted to the STOP Program contained incorrect information with regard to the time the stop was made. Working with the Multnomah County Sheriff's Office, the STOP Program was able to obtain corrected data for nearly 59 percent of Multnomah County's year one submissions. In all, 44.7 percent of Multnomah County's data were adjusted with regard to the time the stop was made because in a relatively small number of cases the stop time that was originally submitted was similar enough to the revised data that it did not need to be replaced. For the remainder of the data that could not be corrected, the original stop time was used.

18 The p-values for the agency-level Veil of Darkness models were examined to ensure that the rigorous threshold set by the inclusion of the Bonferroni correction did not result in the under identification of agencies. In no instance does the Bonferroni correction lead to an agency not being identified that otherwise would have had a statistically significant finding at 95 percent confidence level absent the Bonferroni correction.

5. Predicted Disposition Analysis

This report presents results from two analyses assessing outcomes occurring after the initial stop decision has been made and an individual has been stopped by law enforcement. The first of these two approaches, the Predicted Disposition analysis, is presented in this section and focuses on the outcomes of stops, including whether stopped individuals were cited, searched, and/or arrested during their encounter with law enforcement.

HB 2355 requires all law enforcement agencies to collect data regarding the disposition of stops. Due to the fact that stops can have multiple dispositions (e.g., an individual could be both cited for a traffic violation *and* arrested for a crime) the STOP Program collects data on the most serious disposition that occurred within a single stop.[19] This means, therefore, that if an individual was stopped for speeding, received a citation, and was subsequently arrested on a preexisting warrant, this individual would be recorded in the stop data as only having been arrested. **Table 5.1** reports the percentages of dispositions broken down by agency and demonstrates that there is considerable variation across the twelve Tier 1 agencies examined in this report with regard to the share of drivers cited, searched, and arrested. While most agencies report similar arrest rates, some agencies use warnings or issue citations more often than other agencies. The Marion County Sheriff's Office, for example, cites a higher share of drivers than all other Tier 1 agencies (82.8 percent), while the next closest rate is about half as high for Salem Police Department (43.2 percent). Similarly, Medford and Eugene Police departments have relatively high search and arrest rates compared to the other Tier 1 agencies.

Law enforcement agencies likely experience variation in the application of these different enforcement outcomes. This variation could be due to time of day, day of the week, and/or the offense that led to the stop, to name a few. During rush hour on a weekday, for instance, if heavy traffic flows prevent drivers from exceeding the speed limit then the likelihood of receiving a citation for speeding would be reduced. Variation could also be attributed to other factors, including age, gender, or season. To account for as many of these differences as possible, a variety of propensity score analyses are employed.

Propensity score methods have a long and well-established history in applied statistics. Here, STOP Program researchers use these methods to answer the question, "holding all else constant, do we find different dispositional outcomes

Table 5.1.[20]
Aggregate Tier 1 and Agency-level Stop Dispositions

Agency	Nothing/ Warning	Citation	Search	Arrest
Beaverton PD	61.0%	32.2%	5.4%	5.6%
Clackamas Co SO	62.0%	36.2%	1.3%	1.7%
Eugene PD	55.4%	36.1%	8.1%	7.5%
Gresham PD	59.1%	35.3%	5.1%	3.8%
Hillsboro PD	75.4%	21.6%	2.5%	2.7%
Marion Co SO	15.7%	82.8%	1.2%	1.5%
Medford PD	64.9%	25.2%	10.1%	8.3%
Multnomah Co SO	76.2%	20.2%	2.8%	3.2%
Oregon State Police	61.3%	36.5%	1.9%	1.9%
Portland PB	50.9%	41.2%	5.2%	5.4%
Salem PD	51.4%	43.2%	5.9%	5.3%
Washington Co SO	67.4%	28.8%	3.2%	3.3%
Tier 1	59.6%	36.9%	2.9%	2.9%
Tier 1 (No OSP)	57.6%	37.4%	4.1%	4.0%

19 See Appendix D for more details on how the STOP research team determines the most serious disposition and the appropriate comparison outcomes for each type of disposition.

20 Counts of each form of disposition differ slightly from the raw data to ensure that each stop is evaluated for the most severe dispositional outcome. See Appendix D for more details.

across racial/ethnic groups?" Propensity score methods use the estimated tendency to be included in the group of interest, or propensity score, to make that group and the comparison group look as similar as possible. This approach enables us to make the White comparison group look identical across all measured factors compared to the non-White group of interest. If all other measured variables (i.e., time of day, day of the week, gender, age, stop reason, stop volume) are identical across the two groups then the remaining difference in outcomes is evidence of a disparity due to racial/ethnic differences (Ridgeway 2006).

Many different propensity score methods have been developed in the statistical literature, but they all have a similar goal of making two groups comparable to one another. The best matching method for a given research program depends on the data available, the sample size, the completeness of the data, and other factors; there is no one-size-fits-all approach. Here the STOP Program employed Inverse Probability Weighted Regression Adjustment. [21]

Table 5.2.
Analyses Completed for Each Agency

Disposition of Interest	Comparison Dispositions[22]	Analysis Groups		
Citation	None or Warning	Black,	Hispanic,	Black or Hispanic
Search	None, Warning, or Citation	Black,	Hispanic,	Black or Hispanic
Arrest	None, Warning, Citation, or Search	Black,	Hispanic,	Black or Hispanic
Citation, Search, or Arrest	None or Warning	Black,	Hispanic,	Black or Hispanic

The current analysis included twelve sub-analyses for each agency: each outcome of citation, search, arrest, or any non-warning disposition across each racial/ethnic group of Black, Hispanic, and a combined sample of Black and Hispanic individuals (**Table 5.2**). With higher sample sizes at the full, statewide level these groups were expanded to include Black, Hispanic, Asian/PI, Middle Eastern, and Native American, and the combined Black or Hispanic sample was omitted. The comparison group was drawn from the group of White stops for the agency in question. Each row of **Table 5.2** describes the three tests conducted for each agency. In row 1, STOP Program researchers tested whether there was a disparity in issuing citations between the Black, Hispanic, and the combined Black or Hispanic group, respectively, and a matched White group[23]. Row 2 does the same for searches, row 3 for arrests, and row 4 describes tests for any Citation, Search, or Arrest disposition.

In each of the twelve tests for each agency, the results compare the actual, observed rate of the dispositional outcome for the minority group with an estimate of the rate of the disposition for the best possible comparable White group. A stylized example of this process is presented in **Figure 5.1**, which displays a hypothetical group of Black or Hispanic individuals on the left and the White group on the right, across age of the stopped individual and time of day of the stop. For the White group, the solid dots represent the best comparison group across these factors, whereas the hollow observations represent observations that are relatively poor comparisons based on the age and time of stop variables. The STOP Program research team created comparison groups across many variables including perceived gender, age, time of day, day of week, reason for stop, season, daylight, agency stop volume, and county stop volume.

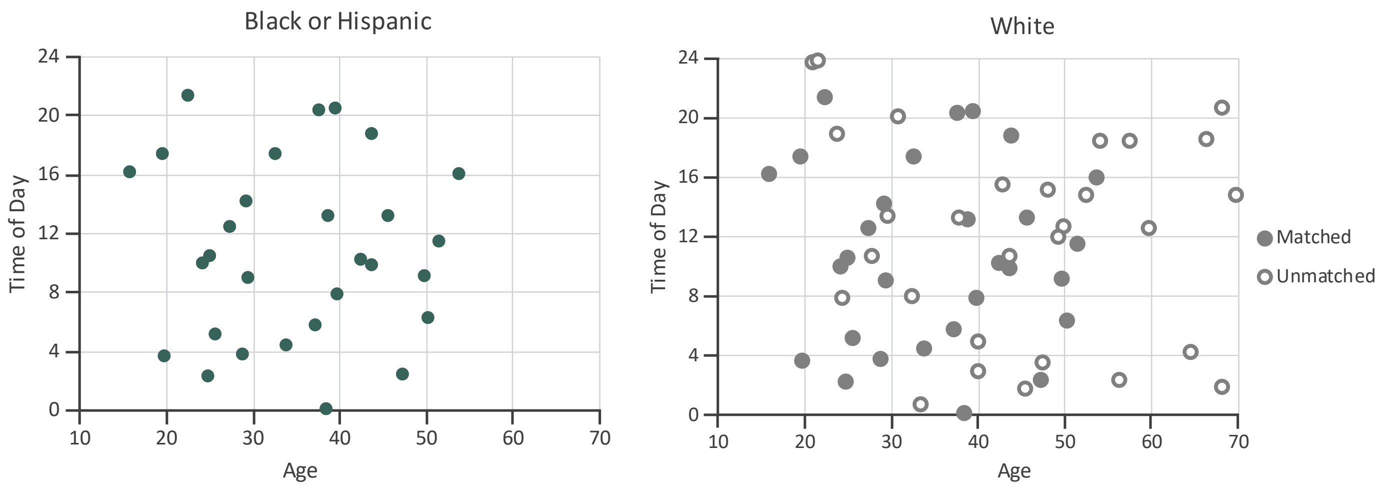
21 IPWRA weights the groups based on the propensity score and then uses these weighted data to estimate the effect of race/ethnicity on dispositional outcomes through regression analysis. For a thorough discussion of this methodology see Appendix D.

22 Discussions with some law enforcement agencies indicated that the inclusion of warnings in the same category as "none" did not necessarily match their conceptualization of warnings as an enforcement device. In the opinion of these agencies, a warning constitutes an action, albeit one that is not as serious as a citation.

23 Each matched White group will differ from the next, since the characteristics of the stops of the group being matched differ.

5. Predicted Disposition Analysis

Figure 5.1.
Hypothetical Representation of the Best Possible Comparison White Group



5.1. Aggregate Predicted Disposition Results for All Tier 1 Agencies

Similar to the Veil of Darkness results presented previously, the STOP Program identifies all agencies with disparities in their predicted versus actual outcomes where those differences were statistically significant at the 95 percent confidence level in any minority group at the aggregate, Tier 1 level or for the Black or Hispanic alone groups at the agency level. **Table 5.1.1** reports the results of models examining the actual and predicted disposition rates for all Tier 1 agencies combined. For each outcome, which includes receiving a citation, being searched, being arrested, or falling into any one of these categories, the actual versus predicted outcomes

Table 5.1.1.
Predicted Disposition Analysis Results for All Tier 1 Agencies

Agency	Race/ Ethnicity	Citation		Search		Arrest		Citation, Search, or Arrest	
		Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Tier 1	Asian/PI	--	--	--	--	--	--	--	--
	Black	--	--	6.3%	4.1%	5.0%	4.2%	--	--
	Hispanic	43.0%	37.4%	3.7%	3.0%	3.5%	3.0%	45.4%	39.6%
	Middle Eastern	--	--	--	--	--	--	--	--
	Native American	--	--	4.5%	3.3%	5.0%	3.4%	--	--
Tier 1 No OSP	Asian/PI	--	--	--	--	--	--	--	--
	Black	--	--	7.8%	4.8%	6.0%	4.9%	--	--
	Hispanic	40.7%	37.3%	4.8%	3.9%	4.6%	3.8%	44.0%	40.1%
	Middle Eastern	--	--	--	--	--	--	--	--
	Native American	--	--	--	--	--	--	--	--

are broken down by minority group. Importantly, for this table and for those that follow, only the results that reached statistical significance at the 95 percent confidence level are reported.[24] As shown in **Table 5.1.1**, in the aggregate, there were disparate outcomes for Hispanic individuals with regard to citations, searches, and arrests. With regard to citations, for example, the propensity score models predicted that 37.4 percent of Hispanic individuals would have been cited, while the actual share of Hispanic individuals cited was approximately 5.5 percentage points higher. For Black and Native American drivers there were also disparities in the aggregate, although these disparities were only found for searches and arrests. Among Native Americans, for instance, the propensity score models predicted that searches should have occurred in 3.3 percent of stops, while the actual share of searched Native Americans was 4.5 percent.

5.2. Agency-level Predicted Disposition Results

While the aggregate results for the Tier 1 agencies indicate that there is evidence of a disparity in dispositional outcomes, it is necessary to examine outcomes by individual agency, as the policies of individual agencies and the training received by officers differ from one agency to another. Indeed, as demonstrated by the variation displayed in **Table 5.1** with regard to the breakdown of different dispositions by agency, aggregate rates are not very informative of the actual patterns of disparity found across the state.

Agency-level results are presented in **Table 5.2.1** (only the results that reached statistical significance at the 95 percent confidence level are reported—full results are available upon request). Eight Tier 1 law enforcement agencies report statistically significant disparities for the Predicted Disposition analysis. For five agencies, Hillsboro Police Department, Marion County Sheriff's Office, Multnomah County Sheriff's Office, Oregon State Police, and Salem Police Department, disparities were only detected for citations and/or for the combined measure of all dispositions (i.e., citation or search or arrest). This indicates that it is likely for these agencies that the only relevant disparity is for citations and not the other outcomes. For two agencies, Beaverton Police Department and Portland Police Bureau, disparities were reported for searches and arrests. Finally, Washington County Sheriff's Office was the sole agency for which a disparity was found for a combination of citations and arrests.

In most of the instances where disparities were found, the average gap in the predicted versus the actual outcomes was around 15 to 30 percent. For instance, the gap between Hillsboro Police Department's actual versus predicted outcomes for citations to Hispanic drivers was 36.6 percent (20.2 percent predicted versus 27.6 percent actual citations).[25] When examined in this fashion, the sole outlier was Portland Police Bureau, as the gap between their predicted and actual outcomes for searches was substantially higher than the other Tier 1 agencies. Specifically, for searches of Black individuals, the actual percentage of searches conducted was over twice as high as the predicted percentage (11 percent actual versus 4.8 percent predicted).

24 Each of the twelve tests is compared to a more stringent Bonferroni adjusted level that is conditional on conducting 12 tests at once, at the 95 percent confidence level. See Appendix D for a more thorough discussion of this adjustment and the full table of results.

25 To calculate the percent difference for these comparisons, Δ% = ((*Actual-Predicted*)/*Predicted*)×100.

Table 5.2.1.
Predicted Disparity, by Agency and Disposition - Statistically Significant Results

Agency	Race/ Ethnicity	Citation		Search		Arrest		Citation, Search, or Arrest	
		Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
Beaverton PD	Black	--	--	--	--	--	--	--	--
	Hispanic	--	--	7.0%	5.0%	6.5%	5.0%	39.6%	36.8%
	Combined†	--	--	7.2%	5.3%	6.9%	5.3%	38.8%	36.6%
Clackamas Co SO	Black	--	--	--	--	--	--	--	--
	Hispanic	--	--	--	--	--	--	--	--
	Combined†	--	--	--	--	--	--	--	--
Eugene PD	Black	--	--	--	--	--	--	--	--
	Hispanic	--	--	--	--	--	--	--	--
	Combined†	--	--	--	--	--	--	--	--
Gresham PD	Black	--	--	--	--	--	--	--	--
	Hispanic	--	--	--	--	--	--	--	--
	Combined†	--	--	--	--	--	--	--	--
Hillsboro PD	Black	--	--	--	--	--	--	--	--
	Hispanic	27.6%	20.2%	--	--	--	--	30.1%	22.4%
	Combined†	26.3%	19.8%	3.4%	2.3%	--	--	28.9%	22.1%
Marion Co SO	Black	--	--	--	--	--	--	--	--
	Hispanic	85.9%	83.4%	--	--	--	--	86.2%	83.7%
	Combined†	--	--	--	--	--	--	--	--
Medford PD	Black	--	--	--	--	--	--	--	--
	Hispanic	--	--	--	--	--	--	--	--
	Combined†	--	--	--	--	--	--	--	--
Multnomah Co SO	Black	24.6%	20.1%	--	--	--	--	27.6%	23.3%
	Hispanic	--	--	--	--	--	--	--	--
	Combined†	24.3%	20.8%	--	--	--	--	26.9%	23.8%
Oregon State Police	Black	45.4%	38.1%	--	--	--	--	47.1%	39.8%
	Hispanic	45.6%	37.7%	--	--	--	--	47.1%	39.2%
	Combined†	45.6%	37.7%	2.4%	2.1%	--	--	47.1%	39.3%
Portland PB	Black	--	--	11.0%	4.8%	7.4%	5.5%	--	--
	Hispanic	--	--	5.9%	4.4%	--	--	--	--
	Combined†	--	--	9.1%	4.7%	6.9%	5.5%	--	--
Salem PD	Black	--	--	--	--	--	--	--	--
	Hispanic	51.3%	46.5%	--	--	--	--	54.6%	49.2%
	Combined†	51.3%	46.4%	--	--	--	--	54.7%	49.3%
Washington Co SO	Black	--	--	--	--	--	--	--	--
	Hispanic	32.1%	28.0%	--	--	4.7%	3.7%	35.7%	31.0%
	Combined†	30.5%	27.7%	--	--	--	--	34.3%	30.9%

† A law enforcement agency with a disparity identified solely for the combined Black/Hispanic group does not meet the threshold to be reported as having a disparity.

6. KPT Hit-Rate Analysis

The second analysis conducted for this report examining post-stop outcomes is the KPT Hit-Rate test. Originally developed in the context of economics, various hit-rate models use outcomes as indicators of economic discrimination in areas such as mortgage loan decision making (Becker 1957; 1993). In the past few decades, this approach examining outcomes to identify discrimination has been adapted extensively in analyses of policing, and the most widely used model is the KPT Hit-Rate model developed by Knowles, Persico, and Todd (2001).

The Knowles, Persico, and Todd (KPT) Hit-Rate model examines whether the likelihood of a "successful" police search differs across racial/ethnic groups, where success is defined as finding contraband. The KPT model assumes that officers make the decision to search a person based on visual and other contextual evidence that they are carrying contraband (e.g., location, furtive movements, or odors associated with drugs, to name a few) in order to maximize search success rates. The model also assumes that motorists adjust their decision to carry contraband based on their likelihood of being searched. In the case that a certain group is more likely to carry contraband, officers will search this group more often in order to maximize their hit-rate, and the group, as a whole, will adjust their likelihood to carry contraband downward. Eventually an equilibrium is reached at which hit-rates are the same across all groups. However, if officers are subjecting a group to more frequent searches based on racial bias, then their hit-rate for that group will decrease. If a minority group's hit-rate is less than the White hit-rate, therefore, this indicates that the minority group is "over searched," which is evidence of a disparity. Put simply, if search decisions are based on race/ethnicity-neutral factors, then hit-rates should be similar. If they are substantially dissimilar, then a disparity is identified.

Hit-rates are calculated by dividing the number of searches in which contraband was found by the total number of searches for each racial/ethnic group. The results for non-White groups are then compared to the outcomes for White individuals to determine whether the success rates are similar. Statewide search data were analyzed for disparities between the White baseline group and individuals identified as Black, Hispanic, Asian/PI, Middle Eastern, and Native American. Due to sample size limitations, agency-specific hit-rates were only calculated for White individuals compared to Black individuals, Hispanic individuals, and a combined sample of Black and Hispanic individuals. For certain agencies and racial/ethnic groups, the Hit-Rate analysis was not performed, because to perform these analyses for an agency for a particular racial/ethnic group the agency must have searched at least 30 people of both the minority group and the White group.[26] This protects against statistical anomalies due to low search counts, and aligns with best practices.[27] Finally, chi-square tests of independence with a Bonferroni adjustment were performed for each comparison to determine if observed differences in hit-rates are statistically significant. Following best practices, the STOP Program identifies all agencies with disparities in the KPT Hit-Rate analysis. At the aggregate Tier 1 level, this includes any minority search hit-rate below the White hit-rate and statistically significant at the 95 percent confidence level. For individual agencies, this includes Black or Hispanic alone hit-rates less than the White hit-rate and statistically significant at the 95 percent confidence level. See Appendix E for more detailed technical information about the KPT Hit-Rate model and statistical tests.

26 Black hit-rate comparisons were not completed for four agencies – Clackamas County Sheriff's Office, Hillsboro Police Department, Marion County Sheriff's Department, and Multnomah County Sheriff's Office – due to the agencies having searched fewer than 30 Black citizens over the study period. These agencies' other racial/ethnic group analyses had sufficient samples and were performed.

27 Connecticut Racial Profiling Prohibition Project (2019).

6. KPT Hit-Rate Analysis

6.1 Aggregate KPT Hit-Rate Results for all Tier 1 Agencies

Table 6.1.1 presents KPT Hit-Rate results for all Tier 1 agencies combined. As shown in the table, the hit-rate for the Black group was determined to be statistically significantly lower than the White hit-rate at the 95 percent confidence level, indicating a disparity. There was no evidence that statewide hit-rates for other groups, Hispanic, Asian/PI, Middle Eastern, or Native American, were statistically significantly different from the White hit-rate. As a robustness check, because searches by Oregon State Police make up 36 percent of all searches across Tier 1 agencies, statewide hit-rates were tested both including and removing Oregon State Police searches. Removing Oregon State Police searches did not substantively change the results, as shown in **Table 6.1.2**, although the overall hit-rates fell slightly across the board.

Table 6.1.1.
Tier 1 Hit-Rates and Significance by Race/Ethnicity

Race/Ethnicity	Minority Hit-Rate	White Hit-Rate	p-value (Significance)
Asian/PI	0.436	0.488	p = 0.144
Black	0.299	0.488	p = 0.000***
Hispanic	0.454	0.488	p = 0.012
Middle Eastern	0.474	0.488	p = 0.931
Native American	0.536	0.488	p = 0.404

* p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction with 5 Comparisons)

Table 6.1.2
Tier 1 (no OSP) Hit-Rates and Significance by Race/Ethnicity

Race/Ethnicity	Minority Hit-Rate	White Hit-Rate	p-value (Significance)
Asian/PI	0.378	0.413	p = 0.439
Black	0.257	0.413	p = 0.000***
Hispanic	0.408	0.413	p = 0.791
Middle Eastern	0.353	0.413	p = 0.597
Native American	0.511	0.413	p = 0.236

* p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction with 5 Comparisons)

6.2 Agency-level KPT Hit-Rate Results

While the aggregate results for the Tier 1 agencies indicate that there is evidence of a disparity in outcomes for the Black group, it is necessary to examine outcomes by individual agency, as the policies of individual agencies and the training received by officers likely differs from one agency to another. As discussed above, individual agencies were analyzed only for Black, Hispanic, and combined Black and Hispanic groups. Results for these analyses are presented in **Table 6.2.1**.

Table 6.2.1
Statewide Hit-Rates and Significance by Agency and Race/Ethnicity

Agency	Race/ Ethnicity	Minority Hit-Rate	White Hit-Rate	p-value (Significance)
Beaverton PD	Black	0.432	0.473	0.468
	Hispanic	0.392	0.473	0.038
	Combined†	0.406	0.473	0.044
Clackamas Co SO	Black	0.263	0.348	(no analysis)
	Hispanic	0.293	0.348	0.600
	Combined†	0.283	0.348	0.416
Eugene PD	Black	0.370	0.395	0.825
	Hispanic	0.444	0.395	0.620
	Combined†	0.404	0.395	0.956
Gresham PD	Black	0.385	0.475	0.142
	Hispanic	0.444	0.475	0.741
	Combined†	0.409	0.475	0.200
Hillsboro PD	Black	0.261	0.343	(no analysis)
	Hispanic	0.429	0.343	0.225
	Combined†	0.393	0.343	0.468
Marion Co SO	Black	1.000	0.329	(no analysis)
	Hispanic	0.195	0.329	0.183
	Combined†	0.283	0.329	0.733
Medford PD	Black	0.429	0.409	0.964
	Hispanic	0.500	0.409	0.207
	Combined†	0.475	0.409	0.270
Multnomah Co SO	Black	0.481	0.450	(no analysis)
	Hispanic	0.378	0.450	0.469
	Combined†	0.417	0.450	0.721
Oregon State Police	Black	0.610	0.601	0.896
	Hispanic	0.564	0.601	0.135
	Combined†	0.575	0.601	0.232
Portland PB	Black	0.109	0.255	0.000***
	Hispanic	0.225	0.255	0.474
	Combined†	0.136	0.255	0.000***

28 During the process of analyzing the Tier 1 data collected during the first year of reporting, it was determined that there was an error in reporting by the Portland Police Bureau. Specifically, the data submitted to the STOP Program by the Portland Police Bureau did not contain search findings for any consent searches, which led to artificially low hit-rates. Shortly before this report was finalized, the Portland Police Bureau provided additional data regarding consent searches. Preliminary analyses indicate that the hit-rates increase to the following levels: White to 50.5 percent, Black to 43.1 percent, Hispanic to 47.2 percent, and Black/Hispanic to 44.1 percent. The addition of this data, however, still results in a disparity for searches of Black individuals and for the combined Black/Hispanic group. Thus, the substantive conclusions do not change.

Salem PD	Black	0.382	0.410	0.895
	Hispanic	0.366	0.410	0.391
	Combined	0.369	0.410	0.388
Washington Co SO	Black	0.562	0.601	0.595
	Hispanic	0.604	0.601	1.000
	Combined	0.593	0.601	0.876

* p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction with 5 Comparisons)

As shown in **Table 6.2.1**, all agencies have differences in search success rates between White individuals and the three comparison groups. These differences were relatively small in nearly all cases, and in all but one case the differences reported were not statistically significant. The lack of statistical significance could be attributed to the relatively small sample sizes found across agencies, but it is also important to note that small, statistically insignificant differences in search outcomes are likely to occur due to random chance even in the absence of policies or practices that could lead to disparate treatment of different groups.

While the vast majority of comparisons present no evidence of disparity in KPT Hit-Rate outcomes and demonstrate only small differences in search outcome percentages, a disparity was found for the Portland Police Bureau in their Hit-Rate analyses for both the Black and the combined Black and Hispanic groups. More specifically, for the White-Black hit-rate comparison, the percentage of successful searches for White individuals was 25.5 percent, while the percentage of successful searches for Black individuals was 10.9 percent. Similarly, for the combined Black/Hispanic group, the percentage of successful searches for White individuals was 25.5 percent, while the percentage of successful searches for the combined group was 13.6 percent. The difference reported for the combined Black/Hispanic group is most likely attributable to the fact that Portland Police Bureau's hit-rate for White individuals is over double that reported for Black individuals and not because of a disparity between Hispanic individuals and White individuals, given that hit-rates for these two groups are much closer to one another (25.5 percent for White individuals versus 22.5 percent for Hispanic individuals alone).

To aid in contextualizing these comparisons, **Figures 6.2.1, 6.2.2**, and **6.2.3** present the results reported in **Table 6.2.1** in visual form. In each figure, the White hit-rate occupies the horizontal axis, while the hit-rate for the comparison group is found on the vertical axis. A diagonal line is also included where the hit-rates between the two groups would be exactly equal, and each agency is represented by a colored dot whose size corresponds with the number of searches conducted and included in the analysis. The location of the dot represents the relationship between the White and comparison hit-rates. For each agency, it would be expected that their dot be close to the diagonal line if disparities are not present. Alternatively, the likelihood of identifying a disparity increases (dependent upon sample size) as an agency's dot falls further below the diagonal line (the region below the diagonal line is where a comparison group's hit-rate is less than the White hit-rate).

Figure 6.2.1.
Black KPT Hit-Rate Comparison

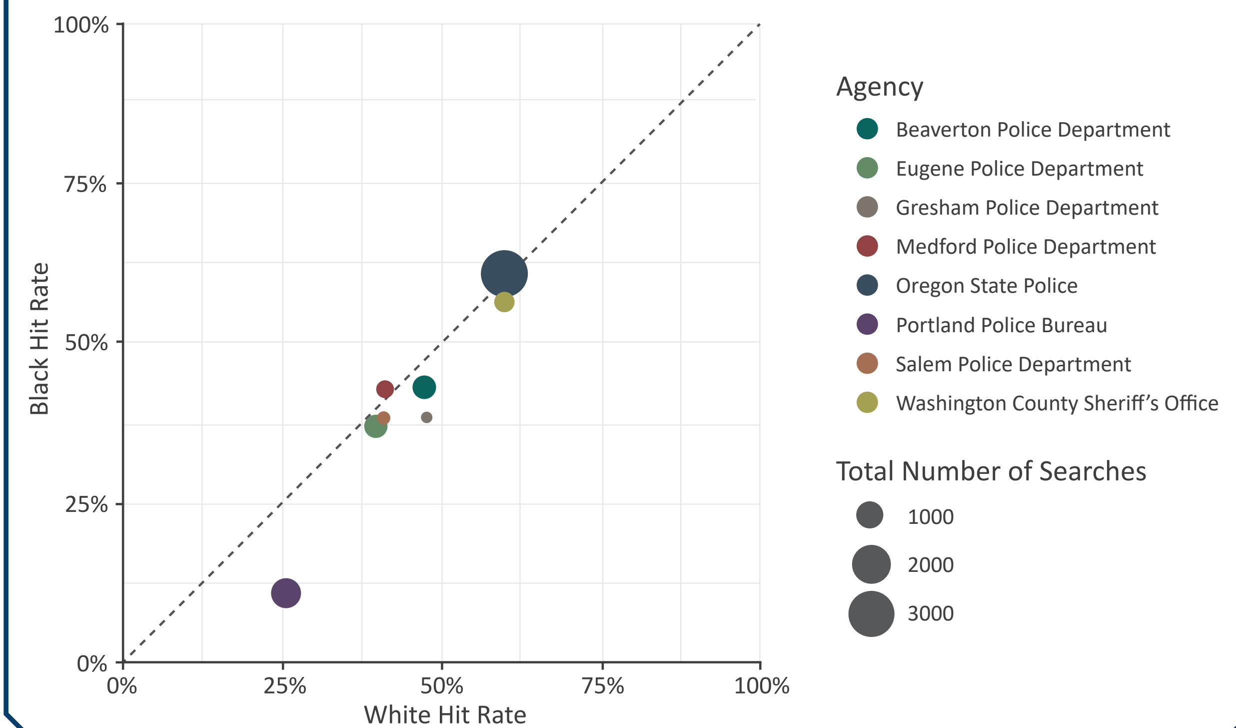


Figure 6.2.2.
Hispanic KPT Hit-Rate Comparison

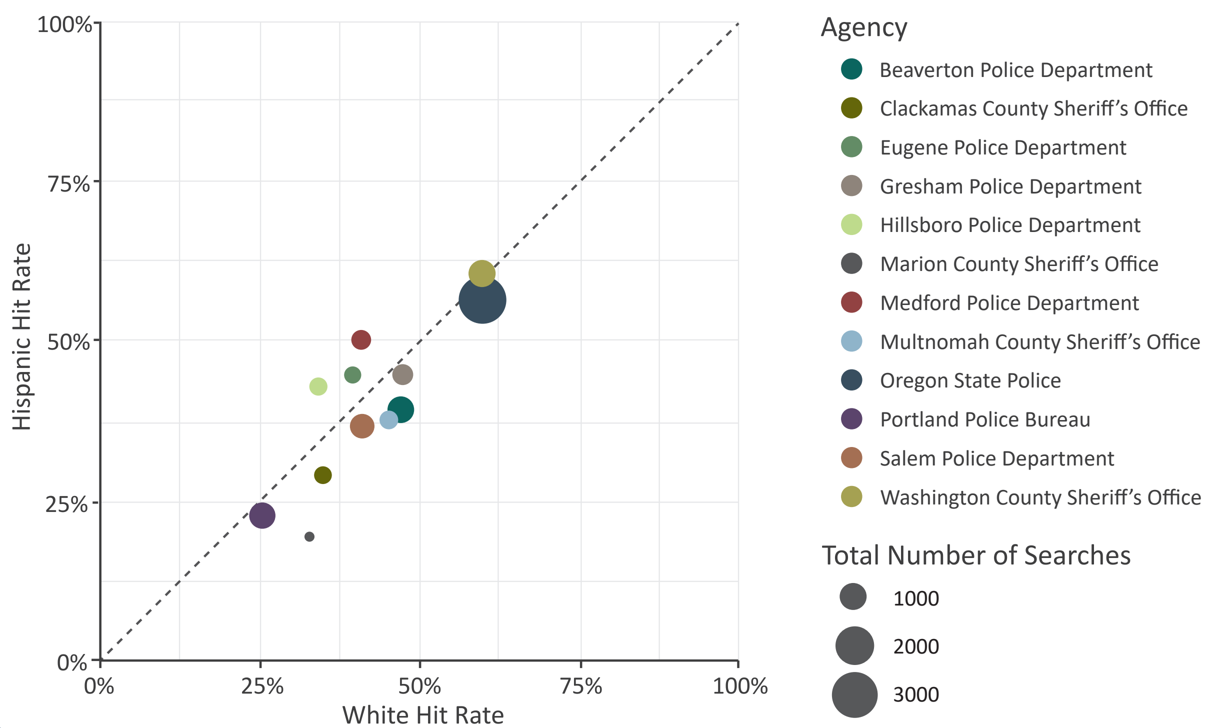
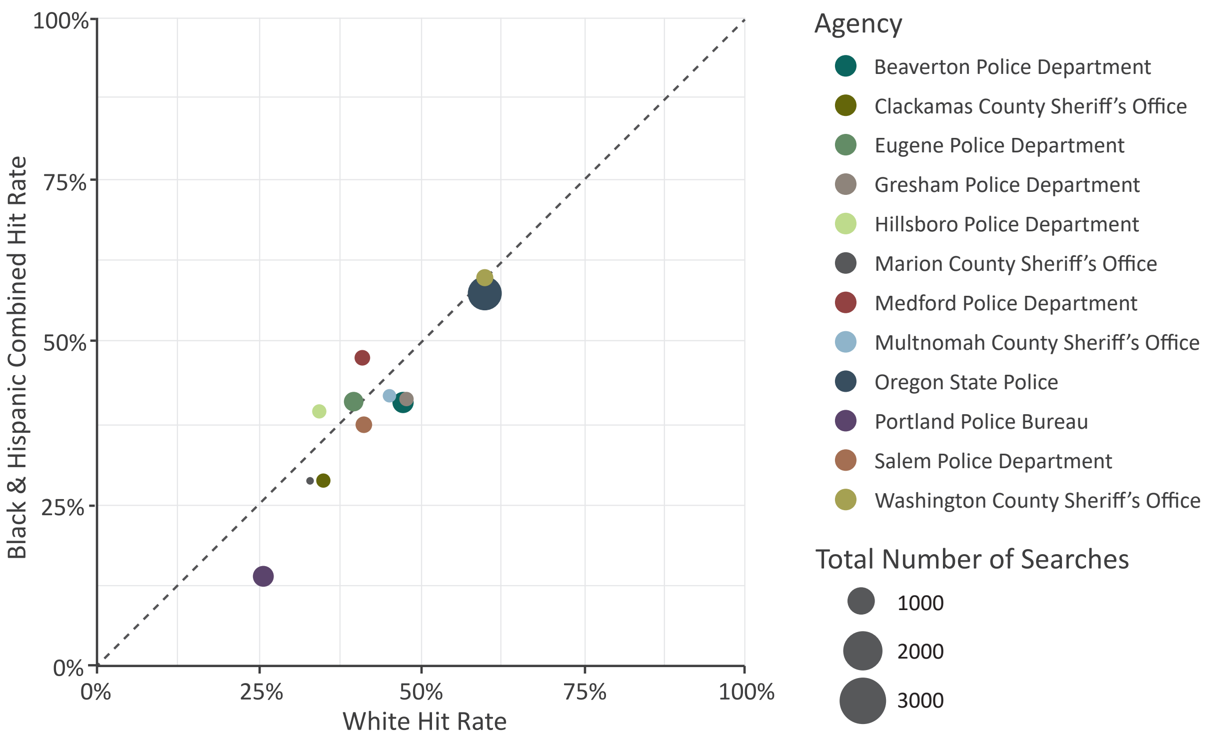


Figure 6.2.3.
Black & Hispanic Combined KPT Hit-Rate Comparison



7. Findings from the 2018-2019 Tier 1 Analysis

7.1 Aggregate Tier 1 Findings for 2018-2019

The results from the analyses presented in this report form an important baseline in the study of traffic and pedestrian stops in Oregon. In all, the Tier 1 data demonstrate that the vast majority of discretionary police-citizen interactions among large law enforcement agencies in Oregon are traffic stops, as over 97 percent of all stops were reported as such in the 2018-2019 data. The breakdown between traffic and pedestrian stops does vary by agency, however, as some Tier 1 law enforcement agencies engage in more pedestrian stops than others, with two agencies, Medford and Eugene, reporting that over 20 percent of their stops were of pedestrians while the remaining agencies reported stop rates similar to the overall state average.

With regard to the demographic characteristics of stopped individuals, the aggregate data indicate that the majority of stops in Oregon were of White drivers or pedestrians. This, in and of itself, is not surprising given the demographic makeup of Oregon as a whole. When disaggregated by traffic versus pedestrian stops, the data indicate that minorities made up a larger share of individuals stopped for traffic violations compared to those stopped as pedestrians. With regard to gender, males were stopped more often than females and non-binary individuals, and this split was greater for pedestrian stops versus traffic stops.

At the aggregate level, Tier 1 agencies reported that stopped individuals either were subject to no further action or merely given a warning in a little over 60 percent of stops. Other outcomes, including receiving a citation or being arrested, varied widely across traffic and pedestrian stops, with nearly 40 percent of all traffic stops resulting in a citation versus slightly more than 13 percent of all pedestrian stops. Similarly, while only 2.5 percent of traffic stops culminated in an arrest, slightly less than 17 percent of pedestrian stops resulted in this most serious of outcomes.

Finally, with regard to searches, approximately 3 percent of all stops resulted in a search of some type. Of those searches, consent was obtained almost 55 percent of the time, while some other legal basis was reported in the remaining 45 percent of cases. Across Tier 1 agencies, almost half of all searches resulted in the discovery of contraband, with alcohol and drugs being the most commonly reported items. When broken down by traffic and pedestrian stops, however, there were some stark differences in search data. Generally, searches of pedestrians occur in a larger share of cases and are successful slightly less often than traffic stop searches (45.4 percent successful for traffic stop searches versus 39.6 percent successful for pedestrian searches).

7.2. Veil of Darkness Findings 2018-2019

One of the few consistent findings reported across the academic and professional literature examining police stop data is that comparisons between stops initiated by law enforcement and residential Census data often leads to invalid, biased results. To examine the decision to stop a driver in a manner that does not rely on benchmarks, STOP Program researchers utilized the Veil of Darkness analysis, which examines stops made in daylight versus darkness surrounding sunrise and sunset. The threshold for identifying disparities was a resulting odds ratio above 1.0 that was statistically significant at the 95 percent confidence level in any minority group at the aggregate Tier 1 level, or for the Black or Hispanic alone groups at the agency level.

The Veil of Darkness Analysis results did not identify any disparities in the decision to stop a driver, either at the state level or for any of the individual Tier 1 agencies. Further, in most of the Veil of Darkness models, the odds ratios were less than 1.0, which indicated that rather than being stopped more often during the daytime when race/ethnicity is easier to discern, non-White drivers were stopped more often in darkness.

7.3. Predicted Disposition Findings 2018-2019

The Predicted Disposition analysis, which relied on balancing samples across racial/ethnic groups to compare similarly situated individuals, was the first of two models used to examine stop outcomes after the decision to stop a driver has been made. For this analysis, STOP Program researchers identified all agencies with disparities in their predicted versus actual dispositional outcomes where those differences were statistically significant at the 95 percent confidence level in any minority group at the aggregate Tier 1 level, or for the Black or Hispanic alone groups at the agency level.

In total, eight Tier 1 agencies were identified as meeting this threshold. With regard to citations, Hillsboro Police Department, Marion County Sheriff's Office, Multnomah County Sheriff's Office, Oregon State Police, Salem Police Department, and Washington County Sheriff's Office were found to have statistically significant differences in the predicted versus actual outcomes. For searches, Beaverton Police Department and Portland Police Bureau were found to have statistically significant disparities. And for arrests, Beaverton Police Department, Portland Police Bureau, and Washington County Sheriff's Office were found to have statistically significant disparities.

7.3.1. Citations

The most common finding of a disparity in this report was for citations, as six Tier 1 agencies were reported to have disparities in citation rates for either Hispanic or Black drivers. The average disparity was a difference in approximately 5.5 percentage points between the predicted and actual rates, with a range of 2.5 to 7.9 percentage points.

Following the estimation of the Predicted Disposition results, STOP Program researchers returned to the stop database to determine whether any additional data could be used to better understand the gap between the actual and predicted citation rates reported in Section 5. Given that HB 2355 (2017) mandates that training and/or technical assistance be offered to agencies with identified disparities, the STOP Program research team wanted to be able to provide their partners at DPSST with as much actionable information as possible so that training and technical assistance could be targeted to the areas of highest need and resources devoted to items that would result in the most immediate, positive changes. In pursuit of this end, STOP Program researchers began to examine the statutory stop reasons found in the STOP database in greater detail.

Using the statutory stop reason data, and in partnership with the Oregon State Police, the STOP Program identified several types of citations for which officer discretion is limited, if not absent completely. According to Oregon State Police policy, for example, all individuals committing the offense of driving with a suspended or revoked driver's license must be cited for this violation. Policies of this kind could be impactful on the outcomes studied in this report, as prior analyses conducted by the Oregon State Police indicate that there is an underlying racial disparity in the share of individuals caught driving while suspended/revoked across different racial/ethnic categories.

To investigate the possibility that departmental policy could drive findings related to citations, STOP Program researchers obtained additional data from the Oregon State Police regarding the reason for the most serious stop disposition. Importantly, while the STOP Program receives data regarding the statutory reason for the stop, it does not gather information concerning the reason for the disposition reported in the data. Using this additional data point, STOP Program researchers were able to both balance on this unique citation reason and conduct robustness checks by leaving out all citations tied to these no/low discretion outcomes. These follow up analyses demonstrated that while a disparity remained for the Oregon State Police for citations, the size of the disparity shrank considerably, by 55 to 71 percent (for a full description of this process, see Appendix F).

The findings with regard to citations are likely influenced, at least in part, by departmental policies regarding citations. While the exact extent of this influence is not yet known and it is unlikely that policies of this kind would explain away the existence of disparities entirely, they represent an important next step in the analysis of stop data in Oregon and a potential area for discussion between law enforcement and the communities these agencies serve.

7.3.2. Searches

With regard to searches, two agencies were identified as having statistically significant disparities for the Black alone or Hispanic alone comparison groups: Beaverton Police Department, and Portland Police Bureau. For Beaverton Police Department, the disparity was for Hispanic drivers, as there was a two percentage point difference between the actual and predicted search rates for these individuals. Portland Police Bureau also had a search disparity for Hispanic individuals of a similar magnitude. Most concerning, however, was the disparity among Black individuals with regard to searches, as the actual search rate for the Portland Police Bureau was more than double the predicted rate obtained from the White comparison group. This represented the widest gap in all of the disparities identified in this analysis.

7.3.3. Arrests

With regard to arrests, three agencies were identified as having statistically significant disparities for the Black alone or Hispanic alone comparison groups: Beaverton Police Department, Portland Police Bureau, and Washington County Sheriff's Office. For Beaverton Police Department and Washington County Sheriff's Office, disparities were identified for Hispanic individuals, while Portland Police Bureau had a disparity for Black individuals.

7.4. KPT Hit-Rate Findings 2018-2019

The second of two analyses examining post stop outcomes was the KPT Hit-Rate analysis, which compared the percentages of successful searches across different racial/ethnic groups. As discussed in greater detail in Section 6, the theoretical idea at the foundation of this test is that if law enforcement personnel apply search criteria or standards equally across race/ethnicity, then similar success rates should be found for all racial/ethnic groups. For this analysis, STOP Program researchers identified all agencies with disparities in their hit-rates where those differences were statistically significant at the 95 percent confidence level in any minority group at the aggregate Tier 1 level, or for the Black or Hispanic alone groups at the agency level.

In this analysis, only Portland Police Bureau was found to have a disparity meeting the above criteria. Specifically, Portland Police Bureau reported successful searches in 25.5 percent of searches involving White individuals but only reported successful searches in 10.9 percent of searches of Black individuals. This disparity represents the largest gap for any agency in this analysis.

7.5. Conclusions and Next Steps

The data contained in this report are intended to be used as a tool for law enforcement, citizens and community members, researchers, Legislators and policy makers, and other stakeholders to focus training and technical assistance on those agencies found to have disparities in outcomes for minority groups. As described previously,

7. Findings from the 2018-2019 Tier 1 Analysis

STOP Program researchers utilized three rigorous statistical analyses, consistent with best practices, to identify disparities in Oregon. The use of these three tests allow the STOP Program researchers to evaluate numerous decision points before and during a stop, while also providing numerous points of analysis in the search for disparate outcomes.

To determine if identified disparities require further analysis and support from the STOP Program and its partners at the Department of Public Safety Standards and Training (DPSST), the following criteria must have been met. (1) An estimated disparity in an individual analysis must have met the 95 percent confidence level for it to be statistically significant. This means, in simple terms, that the STOP Program research team must be at least 95 percent confident that differences or disparities identified by the analyses were not due to random chance. (2) Following best practices and the analyses conducted by the State of Connecticut,[29] for a law enforcement agency to be identified as one requiring further analysis as well as DPSST technical assistance, it must be identified as having a statistically significant disparity in two of the three analytical tests performed on the STOP data.

Based on the above described criteria, it is recommended that Portland Police Bureau be examined in greater detail by STOP Program researchers and receive technical assistance from DPSST. Portland Police Bureau was indicated as having a disparity in both the Predicted Disposition analysis as well as the KPT Hit-Rate analysis with regard to search rates for Black individuals. This consistency across analyses is important, as it indicates that the disparity for searches of Black individuals is robust.

While only the results for Portland Police Bureau rise to the threshold for further analysis set by the STOP Program, a number of other agencies have reported disparities in the Predicted Disposition analysis. The STOP Program has met with nearly all of these agencies and in every case, the agency is willing to engage in further analysis of their data in an effort to isolate the factors leading to the disparate outcomes identified by STOP Program researchers. An ideal typical example of these future partnerships is the work the STOP Program has already begun with the Oregon State Police, as the combination of additional data from law enforcement and the analytical capacity of STOP Program researchers will result in the more precise identification of areas for improvement and/or intervention.

29 The Connecticut Racial Profiling Prohibition Project is located at http://www.ctrp3.org/.

Works Cited

Austin, Peter C. "Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples." Statistics in Medicine 28, no. 25 (2009): 3083–3107. https://doi.org/10.1002/sim.3697.

Austin, Peter C. "Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research." Communications in Statistics - Simulation and Computation 38, no. 6 (September 2009): 1228–34.

Austin, Peter C., and Elizabeth A. Stuart. "Moving towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies." Statistics in Medicine 34, no. 28 (March 2015): 3661–79.

Gary S. Becker, The Economics of Discrimination. Chicago: University of Chicago Press, 1957.

Gary S. Becker, "The Evidence against Banks Doesn't Prove Bias." Business Week, April 19, 1993.

Engel, Robin S., Rob Tillyer, Jennifer Calnon Cherkauskas, and James Frank. 2007. Traffic Stop Data Analysis Study: Year 1 Final Report. November 1, 2007.

Fazzalaro, J., Ken Barone... "Connecticut Racial Profiling Prohibition Project: State of Connecticut Traffic Stop Data Report October 1, 2013 – May 31, 2014." Institute for Municipal and Regional Policy Central Connecticut State University (2014).

Grogger, Jeffrey, and Greg Ridgeway. "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness." Journal of the American Statistical Association 101, no. 475 (2006): 878–87.

"Racial Profiling in Decisions to Search: a Preliminary Analysis Using Propensity-Score Matching." Policing: An International Journal of Police Strategies & Management 36, no. 4 (April 2013). https://doi.org/10.1108/pijpsm-07-2013-0074.

Higgins, George E., Wesley G. Jennings, Kareem L. Jordan, and Shaun L. Gabbidon. 2011. "Racial Profiling in Decisions to Search: A Preliminary Analysis Using Propensity-Score Matching." International Journal of Police Science & Management 13 (4): 336–47.

Higgins, George E., Melissa L. Ricketts, James D. Griffith, and Stephanie A. Jirard. "Race and Juvenile Incarceration: A Propensity Score Matching Examination." American Journal of Criminal Justice 38, no. 1 (2013): 1–12. https://doi.org/10.1007/s12103-012-9162-6.

Niel, Roland, and Christopher Winship. "Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing." Annual Reviews of Criminology 12 (October 2018).

Knowles, John, Nicola Persico, and Petra Todd. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." Journal of Political Economy 109, no. 1 (2001): 203–29. https://doi.org/10.1086/318603.

Persico, Nicola. "Racial Profiling? Detecting Bias Using Statistical Evidence." Annual Review of Economics 1, no. 1 (February 2009): 229–54.

Ridgeway, Greg. "Assessing the Effect of Race Bias in Post-Traffic Stop Outcomes Using Propensity Scores." Journal of Quantitative Criminology 22, no. 1 (March 2006): 1–29.

Works Cited

Simoiu, Camelia, Sam Corbett-Davies, and Sharad Goel. “The Problem of Infra-Marginality in Outcome Tests for Discrimination.” The Annals of Applied Statistics 11, no. 3 (2017): 1193–1216. https://doi.org/10.1214/17-aoas1058.

“Stata Treatment-Effects Reference Manual: Potential Outcomes/Counterfactual Outcomes.” 2013. StataCorp. https://www.stata.com/manuals13/te.pdf.

Stringer, Richard J., and Melanie M. Holland. “It’s Not All Black and White: A Propensity Score Matched, Multilevel Examination of Racial Drug Sentencing Disparities.” Journal of Ethnicity in Criminal Justice 14, no. 4 (2016): 327–47.

“The Stanford Open Policing Project.” openpolicing.stanford.edu. https://openpolicing.stanford.edu/.

Vito, Anthony G., Elizabeth L. Grossi, and George E. Higgins. “The Issue of Racial Profiling in Traffic Stop Citations.” Journal of Contemporary Criminal Justice 33, no. 4 (2017): 431–50.

Wooldridge, Jeffrey M. 2010. Econometric Analysis of Cross Section and Panel Data. Second. MIT Press.

Appendix A – List of Law Enforcement Agencies by Tier

Tier 1

Beaverton PD
Clackamas County SO
Eugene PD
Gresham PD
Hillsboro PD
Marion County SO
Medford PD
Multnomah County SO
Oregon State Police
Portland PB
Salem PD
Washington County SO

Tier 2

Albany PD
Ashland PD
Bend PD
Canby PD
Central Point PD
Clatsop County SO
Corvallis PD
Deschutes County SO
Douglas County SO
Forest Grove PD
Grants Pass DPS
Hood River SO
Jackson County SO
Keizer PD
Klamath County SO
Klamath Falls PD
Lake Oswego PD
Lane County SO
Lebanon PD
Lincoln City PD
Lincoln County SO
Linn County SO
McMinnville PD
Milwaukie PD
Newberg-Dundee PD
Oregon City PD
OHSU PD
Polk County SO
Port of Portland PD
Redmond PD
Roseburg PD
Springfield PD
Tigard PD
Tualatin PD
Umatilla County SO
University of Oregon PD
West Linn PD
Woodburn PD
Yamhill County SO

Appendix A – List of Law Enforcement Agencies by Tier

Tier 3

Astoria PD
Aumsville PD
Baker City PD
Baker County PD
Bandon PD
Black Butte Ranch PD
Boardman PD
Brookings PD
Burns PD
Butte Falls PD
Cannon Beach PD
Carlton PD
Coburg PD
Columbia City PD
Columbia County SO
Condon PD
Coos Bay PD
Coos County SO
Coquille PD
Cottage Grove PD
Crook County SO
Curry County SO
Dallas PD
Eagle Point PD
Enterprise PD
Florence PD
Gearhart PD
Gervais PD
Gilliam County SO
Gladstone PD
Gold Beach PD
Grant County SO
Harney County SO
Hermiston PD
Hines PD
Hood River PD
Hubbard PD
Independence PD
Jacksonville PD
Jefferson County SO
John Day PD
Josephine County SO
Junction City PD
King City PD
La Grande PD
Lake County SO
Lakeview PD
Madras PD
Malheur County SO
Malin PD
Manzanita DPS
Merrill PD
Milton-Freewater PD
Molalla PD
Monmouth PD
Morrow County SO
Mt. Angel PD
Myrtle Creek PD
Newport PD
North Bend PD
North Plains PD
Nyssa PD
Oakridge PD
Ontario PD
Pendleton PD
Philomath PD
Phoenix PD
Pilot Rock PD
Port Orford PD
Portland State University PD
Powers PD
Prineville PD
Rainier PD
Reedsport PD
Rockaway Beach PD
Sandy PD
Scappoose PD
Seaside PD
Sherman County SO
Sherwood PD
Silverton PD
St. Helens PD
Stanfield PD
Stayton PD
Sunriver PD
Sutherlin PD
Sweet Home PD
Talent PD
The Dalles PD
Tillamook County SO
Tillamook PD
Toledo PD
Turner PD
Umatilla PD
Union County SO
Union Pacific Railroad PD
Vernonia PD
Wallowa County SO
Warrenton PD
Wasco County SO
Wheeler County SO
Winston PD
Yamhill PD

Appendix B - Data Audit

Missing data in the STOP Program database can be traced to two sources. First, a value for an individual data point could be truly missing, which means that no information for that particular variable was entered into the STOP database. Second, a data point could be invalid, which means that a value of some type was entered into the STOP database, but it did not conform to the standards of the STOP Program. **Table B.1** presents a breakdown of missing data in these two forms for the STOP Program variables used in the analyses contained in this report.

Table B.1.
Missing Data for STOP Program Variables used in Year 1 Report Analyses

Variable	Description	Analyses Affected	% Missing	% Invalid
age	Age perceived by officer	Veil of Darkness, Predicted Disposition	0.59%	0.00%
agency	Stopping agency	Veil of Darkness, Predicted Disposition, Hit-Rate	0.00%	0.00%
arrest	Physical custody arrest (yes/no)	Predicted Disposition	0.00%	0.00%
cite_cat*	Category of citation (Move/Spd, Ser Move/Spd, Very Ser Move/Sp, Equip Vio/Cell/Seatbelt, Reg/License, Other)	Veil of Darkness	0.00%	0.00%
cite_type	Citation basis for traffic stop (ORS, Mu-nicipal Traffic, Municipal Criminal, County Ordinance)	Veil of Darkness, Predicted Disposition	0.04%	0.00%
county	County in which stop occurred	Veil of Darkness, Predicted Disposition	0.00%	0.00%
disposition	Most severe disposition of stop (none, warning, citation, search, arrest)	Predicted Disposition	0.06%	0.00%
gender	Gender perceived by officer (male, female, non-binary)	Veil of Darkness, Predicted Disposition	0.09%	0.26%
race	Race/ethnicity perceived by officer (Asian/PI, Black, Hispanic, Middle Eastern, Native American, White)	Veil of Darkness, Predicted Disposition, Hit-Rate	0.42%	1.75%
sdate	Date of stop	Veil of Darkness, Predicted Disposition	0.00%	0.00%
search	Whether a discretionary stop occurred (yes/no)	Predicted Disposition, Hit-Rate	0.00%	0.00%
search_f1**	What was found if a search occurred (Nothing, Alcohol, Drugs, Stolen Property, Weapons, Other Evidence, Other non-Evidence)	Hit-Rate	4.04%	0.00%
stime	Time of stop	Veil of Darkness, Predicted Disposition	0.00%	0.00%
stop_type	Type of stop (traffic, pedestrian)	Veil of Darkness, Predicted Disposition	0.00%	0.00%
TOTAL			1.18%	1.95%

*cite_cat is a condensed variable created from the original variable cite_code, which denotes the ORS code for the citation. As not every stop has an ORS code basis, some stops are missing cite_code, but are not included in these missing/invalid counts.
**This missing percentage reflects the percent of search_f1 missing when an entry is expected. In the case that Search= "no", there is not an entry expected for search_f1, so these are not included in the missing percentage in this table.

As shown in **Table B1**, missing data were reported for several, but not all variables used in the STOP Program analyses. For gender and race/ethnicity, there were both missing and invalid values encountered in the database. Missing values, as described above, have no information reported, while invalid values in the context of the STOP Program database took the form of submissions with data points labeled as "U", "O", "Unknown", or "Other" for race/ethnicity and/or gender. As stated, an unknown value for these variables does not conform to STOP Program standards and is thus not considered to be a valid entry in the database.

Importantly, the share of missing data for the demographic variables age, race/ethnicity, and gender was limited to four agencies, with two contributing the largest share of missing values. For both of these agencies, the Oregon State Police and the Marion County Sheriff's Office, STOP Program researchers met with command staff to communicate the concerns raised by the presence of missing data. For the Oregon State Police, the data submitted to the database that were identified for this Appendix all came in the form of invalid "Unknown" values. Through training and directive, the Oregon State Police have been working to remedy this issue and the STOP Program research staff will follow up on the degree to which this issue has improved in its next report.

Missing data concerns for the Marion County Sheriff's Office were different than those described above for the Oregon State Police. For Marion County Sheriff's Office, data were truly missing, meaning that they were not entered for a number of cases. Through meetings with Marion County Sheriff's Office command staff, it was discovered that this was due to a software error which has since been fixed. Similar to the Oregon State Police, STOP Program researchers will follow up on the degree to which this issue has improved in its next report.

On the whole, the data submitted by the Tier 1 agencies have met the standards of the STOP Program and do not raise significant concerns with the research staff. While it is important to strive for 100 percent reporting in all cases, rarely does reporting meet this goal in the first year of data collection. While few states publish data regarding error rates in reporting, compared to those that have reported such information, Oregon appears to be ahead of the curve. [30]

30 In its inaugural report, the State of Arizona, for example, reported an overall error rate of 14.1 percent. *Please see* Engel et al. (2007).

The Veil of Darkness (VOD) analysis was first developed by Grogger and Ridgeway (2006) for analyzing stop data for racial/ethnic disparities and is based on the basic assumption that officers can better detect a driver's race during daylight hours as compared to darkness. Specifically, relying on variations in daylight throughout the year, the VOD test compares the racial composition of stops in daylight to those in darkness during a combined inter-twilight window, which occurs during morning and evening commute times. The primary advantage of the test is that it does not rely on a benchmark comparison of either the estimated driving population or the residential population. Further, it is a widely accepted technique (often referred to as the "gold standard"), does not suffer from benchmarking issues, and when deployed via a multivariate analysis provides a strong test of racial disparities (Fazzalaro and Barone 2014).

The Veil of Darkness analysis relies on two primary assumptions. The first is that in darkness, it is more difficult for officers to determine the race/ethnicity of an individual they intend to stop. Second, the analysis also assumes that driving population is consistent throughout the year, between daylight and darkness, and between the morning and evening commutes. If these assumptions hold, it is possible to model the differences in stops between light and dark using a logistic regression that takes the following form:

$$ln\left(\frac{(P(m|\delta)}{1\text{-}P(m|\delta)}\right)=\alpha+\delta+\gamma+\omega+\varepsilon$$

where m represents the treatment of a minority group relative to the White majority group, δ is a binary indicator representing daylight, γ is a vector of coefficients, including controls for time of day, day of the week, season, and agency and county stop volume, and ω is a vector of coefficients representing the demographic characteristics of the stopped individual as well as the reason for the stop.[31] Importantly, the inclusion of controls for time of day, day of the week, and season ensure that the model meets the second assumption regarding the consistency of the driving population throughout the year.

A key factor in the specification of the VOD model is identifying the appropriate periods of daylight and darkness for the analysis. Following Grogger and Ridgeway (2006), the STOP Program analyzes stops that occur within the combined inter-twilight window. The combined inter-twilight window is created from the Oregon traffic stop data from July 1, 2018 to June 30, 2019. Every traffic stop is defined to have occurred in daylight or darkness based on the date, time, and location of the stop. Astronomical data from the United States Naval Observatory (USNO) is used to determine the sunrise, sunset, and start and end of civil twilight. If the location of the stop has been geo-coded, then those coordinates are used to determine the sunrise, sunset, and civil twilight window for that exact location. If the stop has not been geo-coded due to limitations with location data, the centroid of the city is used. If the city information is unavailable, then the centroid of the county is used.

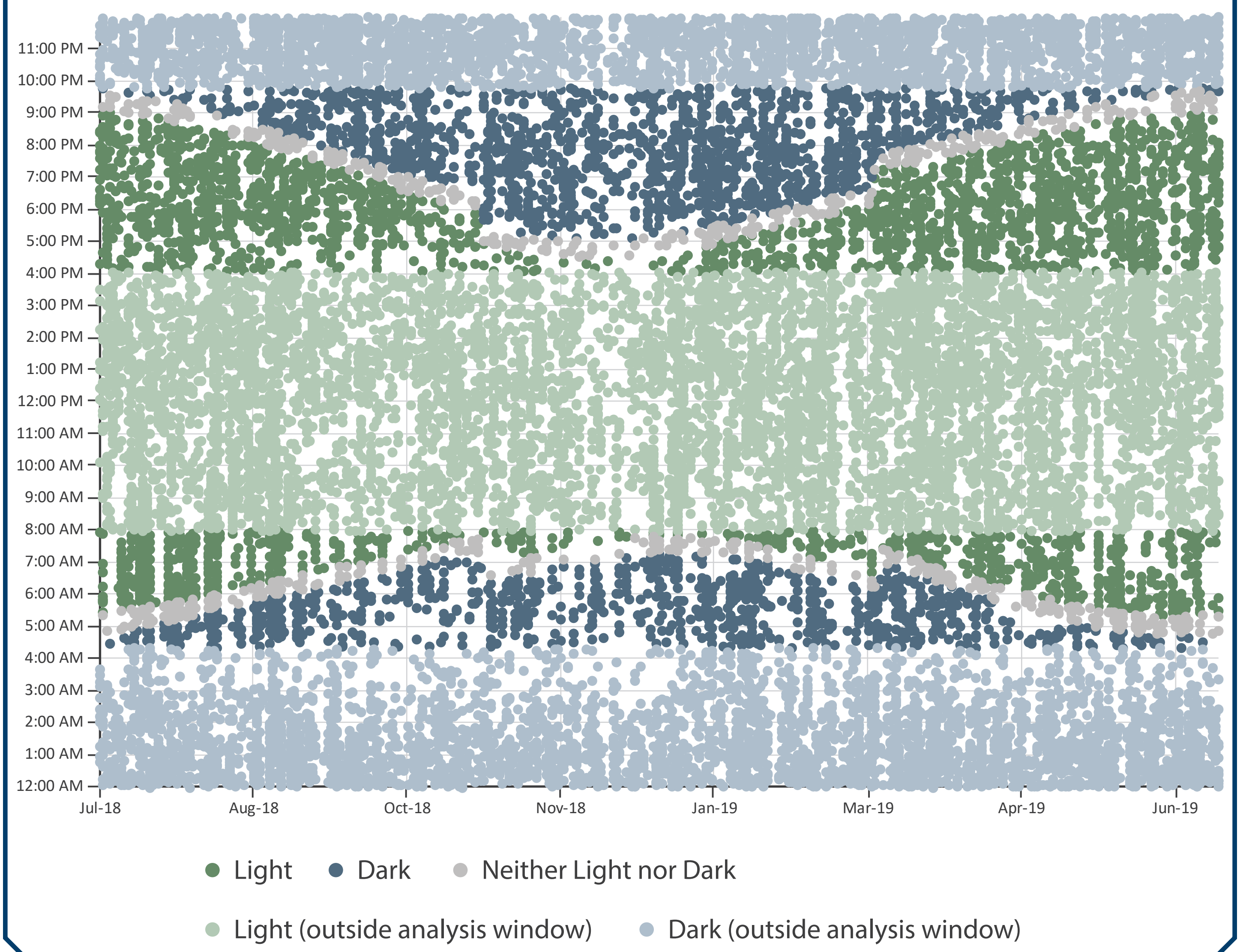
The dawn inter-twilight period is defined as the earliest start of civil twilight to the latest sunrise. The earliest start of civil twilight is 4:21am in Wallowa County, and the latest sunrise is 7:59am in Clatsop County. Stops that occur in the daily morning twilight window (approximately 30 minutes between the start of civil twilight and the sunrise) are removed since it is neither light nor dark during this time period. Conversely, the dusk twilight window is defined as the earliest sunset to the latest end of civil twilight. The earliest sunset is 4:05pm in Wallowa County,

31 The covariates included in the models were age, gender, reason for the stop, day of week, time of day, quarter or season, county stop volume, agency stop volume. Time of day is modeled as a control variable for morning and evening stops, as well as a spline with three degrees of freedom within each twilight window. Alternative time of day controls were tested and did not change the results. In models examining aggregate Tier 1 outcomes, agency fixed effects were included as well.

and the latest end of civil twilight is 9:48pm in Clatsop County. Stops that occur in the daily evening twilight window (approximately 30 minutes between sunset and the end of civil twilight) are similarly removed since it is neither light nor dark during this time period. Adjustments have been made to account for daylight savings time (DST) in November 2018 and March 2019. In addition, most of Malheur County is on Mountain Standard Time (MST) and the stops in Malheur County have been adjusted to account for this time zone.

Figure C.1 displays the traffic stops for one Tier 1 agency by date and time of day. The gradual variation in the sunrise and sunset times shows the shift in daylight and darkness for traffic stops. The morning and evening inter-twilight windows are displayed, and stops during these times are included in the VOD analysis.

Figure C.1.
Statewide Distribution of Traffic Stops by Time and Day



The log odds that result from the Veil of Darkness logistic regression model were then converted to odds ratios. Thus the model tests whether the odds of non-White traffic stops during daylight are significantly different from the odds of non-White traffic stops during darkness. The VOD approach tests whether the odds ratio is statistically significantly different from 1.0. If the odds ratio is not statistically different from 1.0, then the test finds no difference in stops made during daylight and darkness. If the odds ratio is greater than 1.0 and statistically significant, however, the test concludes the odds of non-White drivers being stopped in daylight is significantly higher than in darkness, which is taken as evidence of a racial disparity in stops, after accounting for additional control variables that are available in the stop data. Conversely, if the odds ratio is less than 1.0 and statistically significant, the odds of a non-White driver being stopped in daylight is significantly lower than in darkness. The logistic regression modeling was compiled using SAS software and utilizing the procedure logistic function[32].

Detailed Veil of Darkness Results

The following tables display the VOD Results for the statewide and agency specific models. For each model the sample size, coefficient, standard error, p-value, odds ratio, and 95 percent confidence interval of the odds ratio are displayed. The statewide results in the main report include a Bonferroni correction for 5 comparisons. The statewide results in the DST window, and agency specific results, include a Bonferroni correction for 3 comparisons.

Table C.1. All Tier 1 Agencies VOD Results

Race/Ethnicity	Sample Size	Coefficient	Standard Error	P-Value	Odds Ratio	95% CI	
Asian/PI	101,604	0.024	0.029	0.415	1.048	0.903	1.217
Black	102,855	-0.022	0.025	0.367	0.957	0.843	1.086
Hispanic	112,611	-0.035	0.016	0.028	0.933	0.860	1.012
Middle Eastern	98,720	-0.059	0.045	0.191	0.889	0.705	1.121
Native American	97,747	-0.077	0.071	0.278	0.858	0.597	1.234
Asian/PI (no OSP)	36,971	-0.025	0.038	0.504	0.950	0.781	1.157
Black (no OSP)	38,469	-0.054	0.031	0.079	0.898	0.767	1.051
Hispanic (no OSP)	42,003	-0.053	0.024	0.025	0.899	0.796	1.016
Middle Eastern (no OSP)	35,385	0.010	0.065	0.878	1.020	0.729	1.428
Native American (no OSP)	64,908	-0.106	0.104	0.310	0.809	0.473	1.385

Table C.2. All Tier 1 Agencies Restricted to DST Window VOD Results

Race/Ethnicity	Sample Size	Coefficient	Standard Error	P-Value	Odds Ratio	95% CI	
Black	32,685	-0.054	0.046	0.242	0.898	0.721	1.119
Hispanic	35,674	-0.061	0.030	0.047	0.886	0.766	1.025
Combined Sample	37,580	-0.060	0.026	0.023	0.887	0.781	1.006
Black (no OSP)	12,362	-0.081	0.582	0.165	0.851	0.644	1.124
Hispanic (no OSP)	13,407	-0.039	0.048	0.414	0.924	0.735	1.163
Combined Sample (no OSP)	14,701	-0.059	0.039	0.134	0.889	0.737	1.073

32 SAS software, Version 9.4 of the SAS System for X64_8PRO Windows. Copyright © 2002-2012 SAS Institute Inc., Cary, NC, USA.

Table C.3
Black vs. White Traffic Stops, Agency VOD Results

Agency	Sample Size	Coefficient	Standard Error	P-Value	Odds Ratio	95% CI	
Beaverton PD	3,912	-0.052	0.098	0.597	0.902	0.565	1.439
Clackamas Co SO	5,711	0.147	0.121	0.226	1.341	0.751	2.393
Eugene PD	2,175	-0.276	0.185	0.135	0.575	0.237	1.395
Gresham PD	2,118	-0.131	0.112	0.243	0.770	0.450	1.317
Hillsboro PD	2,433	0.009	0.140	0.947	1.019	0.521	1.992
Marion Co SO	1,862	Insufficient sample size					
Medford PD	959	Insufficient sample size					
Multnomah Co SO	2,488	-0.037	0.131	0.777	0.928	0.495	1.741
Oregon State Police	64,386	0.019	0.042	0.653	1.039	0.848	1.273
Portland PB	8,686	-0.074	0.046	0.106	0.863	0.693	1.074
Salem PD	1,895	Insufficient sample size					
Washington Co SO	6,094	-0.057	0.097	0.555	0.892	0.559	1.421

Table C.4.
Hispanic vs. White Traffic Stops, Agency VOD Results

Agency	Sample Size	Coefficient	Standard Error	P-Value	Odds Ratio	95% CI	
Beaverton PD	4,436	0.086	0.067	0.201	1.187	0.861	1.635
Clackamas Co SO	6,213	0.055	0.072	0.449	1.116	0.789	1.579
Eugene PD	2,208	0.081	0.165	0.625	1.175	0.533	2.590
Gresham PD	2,210	-0.236	0.102	0.020	0.624	0.383	1.014
Hillsboro PD	3,056	-0.070	0.079	0.373	0.869	0.596	1.267
Marion Co SO	2,360	-0.101	0.101	0.320	0.817	0.503	1.328
Medford PD	1,067	-0.123	0.210	0.558	0.782	0.287	2.134
Multnomah Co SO	2,667	0.002	0.101	0.983	1.004	0.620	1.626
Oregon State Police	70,608	-0.027	0.022	0.206	0.947	0.854	1.050
Portland PB	7,944	-0.132	0.057	0.021	0.767	0.583	1.009
Salem PD	2,410	0.081	0.096	0.398	1.177	0.742	1.865
Washington Co SO	7,281	-0.097	0.053	0.068	0.824	0.639	1.062

Table C.5.
Black & Hispanic Combined vs. White Traffic Stops, Agency VOD Results

Agency	Sample Size	Coefficient	Standard Error	P-Value	Odds Ratio	95% CI	
Beaverton PD	4,801	0.048	0.059	0.413	1.101	0.831	1.458
Clackamas Co SO	6,449	0.081	0.640	0.207	1.175	0.865	1.596
Eugene PD	2,322	-0.091	0.127	0.471	0.833	0.454	1.528
Gresham PD	2,556	-0.181	0.082	0.027	0.697	0.470	1.031
Hillsboro PD	3,247	-0.058	0.073	0.423	0.890	0.629	1.260
Marion Co SO	2,404	-0.040	0.099	0.686	0.923	0.576	1.480
Medford PD	1,109	-0.226	0.188	0.230	0.636	0.258	1.566
Multnomah Co SO	2,889	-0.016	0.084	0.853	0.969	0.696	1.349
Oregon State Police	72,508	-0.018	0.020	0.355	0.964	0.877	1.060
Portland PB	9,759	-0.097	0.039	0.013	0.825	0.685	0.992
Salem PD	2,495	0.098	0.092	0.285	1.217	0.784	1.889
Washington Co SO	7,659	-0.088	0.049	0.072	0.839	0.664	1.060

Propensity score methods are a family of statistical methods for drawing causal inference about treatment effects in situations where randomized control trials are not feasible. Randomized control trials ensure that treatment assignment is independent of all covariates. Without this randomization, confounders may bias the estimated treatment effects. Confounding variables are a major hurdle to estimating effects in real-world settings and balancing based on the propensity to receive treatment (i.e., propensity score) is one way to mitigate this bias in non-experimental settings. In general, propensity score techniques aim to balance the characteristics (or confounding variables) of the treatment and control groups. This allows an unbiased comparison between those two groups for the outcome variable of interest, as there are no observed differences between the two groups. These methods are frequently employed in the analysis of disparities in criminal justice settings (Higgins et al. 2011; 2013; Ridgeway 2006; Stringer and Holland 2016; Vito, Grossi, and Higgins 2017).

Propensity score methods measure the characteristics of the "treatment" and "control" groups and then weight one or both of these groups based on measured characteristics so that the two groups look as similar as possible. The resulting groups are said to be "balanced" if they are statistically similar across measured confounding variables following the balancing procedure. If all confounding variables are measured and balanced then the difference in the average outcomes between the treatment and control groups is an unbiased measure of the average treatment effect. Similarly, if unmeasured confounding variables are closely correlated with the balanced confounding variables and thus are also likely to be balanced, then the average treatment effect is unbalanced. Some methods, as employed in the current analysis, go a step further and incorporate regression analysis as an additional controlling method after the balancing process.

There are several different forms of propensity score estimators. Here the researchers employ Inverse Probability Weighted Regression Adjustment (IPWRA) using the Stata statistical package[33]. The method has the following steps:

1. The treatment equation is estimated including potentially confounding variables. The dependent variable is a binary treatment variable and a logistic-type of model is estimated.

2. The predicted treatment values from the estimates in step 1 are stored.

3. Inverse probability weights (IPW) are created for each observation.[34]

    a. For treated observations, IPW =1

    b. For control observations, IPW = $\dfrac{(\textit{propensity score})}{1\text{-}(\textit{propensity score})}$

4. The outcome equation is estimated using the weights created in step 3, including all covariates that are theoretically relevant predictors of the outcome variable.

One advantage of the IPWRA estimator relative to other propensity score estimators is that it benefits from the Double Robust property by estimating the regression equation after the balancing procedure: If either the treatment equation or the outcome equation is correctly specified then the estimator is unbiased. Put alternatively, the estimates from IPWRA estimation are robust to misspecification errors in either the treatment or outcome equation. Two-stage propensity score estimators such as IPWRA balance for important covariates at both the treatment selection and outcome stages of estimation.[35]

33 StataCorp. 2013. Stata: Release 13. Statistical Software. College Station, TX: StataCorp LP.

34 These differ whether the estimated is the Average Treatment Effect (ATE) or the Average Treatment Effect on the Treated (ATET). Here we are estimating the ATET. (Austin and Stuart 2015)

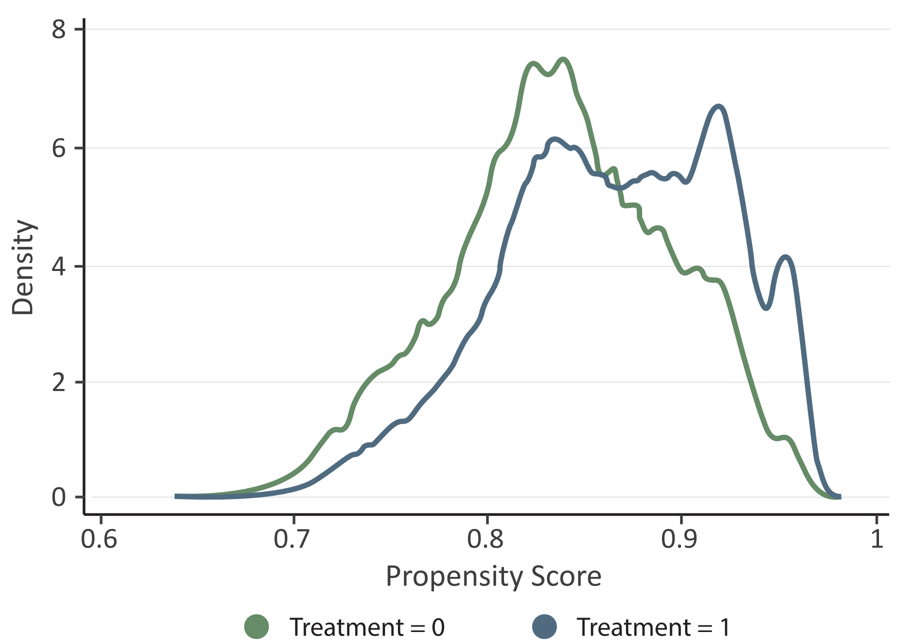
35 For a thorough discussion of IPWRA methods see (Wooldridge 2010, Chapter 21.3.4)

Assumptions

There are a few assumptions that must hold in order for propensity score estimators to be unbiased. The first is the conditional independence assumption[36], which states that the outcome variable is conditionally independent of the treatment. This means that if researchers include all relevant confounding variables in estimating the treatment equation, i.e., the treatment equation is properly specified, and these variables are balanced across the two groups following match selection, then the outcomes are conditionally independent of the treatment. In order for this assumption to hold, changes in any unobserved variables that have an effect on the outcome variable must not also have an effect on the treatment variable. This assumption is a theoretical consideration that is not possible to directly test, as a variable may be correlated with both treatment and outcome but may be a spurious correlation. The analyst may, however, ensure that all the measured confounding variables are equally represented in both the treatment and control groups and thus that the confounding variables are not the drivers of remaining variance in treatments and outcomes.

The second main assumption is the overlap assumption, whereby the range of estimated propensity scores for the treated group must overlap with those of control group observations. If an observation is not within this range then it is omitted from the sample as it is impossible to form a valid match from the comparison group. This idea is best represented with a pre-balance propensity score distribution graph, as seen in the examples below. Figure 2 shows that for most values of the propensity score (horizontal axis) there is an observation for both the treated (treatment=1) and untreated (=0) groups, but also that at the upper and lower ends there are treated observations that do not have a comparable observation in the untreated group. To satisfy this assumption for this example these observations with extreme propensity scores would be dropped.

Figure D.1.
Overlap Example



With a limited range of covariates, including mostly categorical variables, and the large sample sizes with this set of Tier 1 agencies, each analysis completed here had no omitted observations because of a violation of the overlap assumption.[37]

Finally is the Stable Unit Treatment Value Assumption (SUTVA), which is similar in concept to the independent and identically distributed (i.i.d.) assumption, but specific to the treatment assignment setting. SUTVA requires

36 This assumption is also referred to as the unconfoundedness assumption.

37 Omitted treatment variables per analysis are not presented here due to the high number of analyses conducted.

that any given unit's treatment assignment does not have a causal relationship with another observation's treatment assignment. This assumption would be violated in this case if, for example, the stop of a Hispanic individual causes another Hispanic individual to be stopped. There may be clustering of stops by race/ethnicity group based on policing strategies, but this assumption is not likely to be violated in this case as the race of a stopped individual does not directly impact the race of subsequently stopped individuals.[38]

Estimation

If the above assumptions hold then estimation proceeds. The *teffects ipwra* command is used in Stata to estimate these models. First the "treatment" equation is estimated. The treatment variables in this case are indicator variables for each of

1. Officer perception of race/ethnicity: = 1 if Black or Hispanic, = 0 if White
2. Officer perception of race/ethnicity: = 1 if Black, = 0 if White
3. Officer perception of race/ethnicity: = 1 if Hispanic, = 0 if White

For the statewide models, a broader set of treatment variables is available because of the higher sample size:

1. Officer perception of race/ethnicity: = 1 if Black, = 0 if White
2. Officer perception of race/ethnicity: = 1 if Hispanic, = 0 if White
3. Officer perception of race/ethnicity: = 1 if Asian/PI, = 0 if White
4. Officer perception of race/ethnicity: = 1 if Middle Eastern, = 0 if White
5. Officer perception of race/ethnicity: = 1 if Native American, = 0 if White

The standard language of treatment/control used with the IPWRA methodology is ill-suited to this STOP analysis. The current analysis weighs the two groups under each sub-analysis across all observed covariates, rather than giving one group a treatment, but not the other. This method makes it so that the only perceptible difference between the two groups are the race/ethnicity of those two groups, but race/ethnicity does not conform to this "treatment" description. This language is preserved simply to remain consistent with the relevant literature.

The following confounding variables are balanced across the groups:

1. Female indicator, 1 = if female, 0 = if any other
2. Age category indicators for each of <21, 21-24, 25-29, 30-39, 40-49, 50+
3. Season indicators for each of Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec
4. Daylight indicator = 1 if stop happened after sunrise and before sunset, = 0 otherwise
5. Time of stop indicators for each of 12am-5am, 5am-10am, 10am-3pm, 3pm-8pm, 8pm-12am
6. Citation category indicators for each of Moving/Speeding; Serious Moving/Speeding; Very Serious Moving/Speeding; Equipment, Cell, or Seatbelt; Registration/License; Other

38 The Stata handbook provides a good description of these assumptions, and the counterfactual model that underlies all matching methods. ("Stata Treatment-Effects Reference Manual: Potential Outcomes/Counterfactual Outcomes" 2013).

7. Day of week indicators

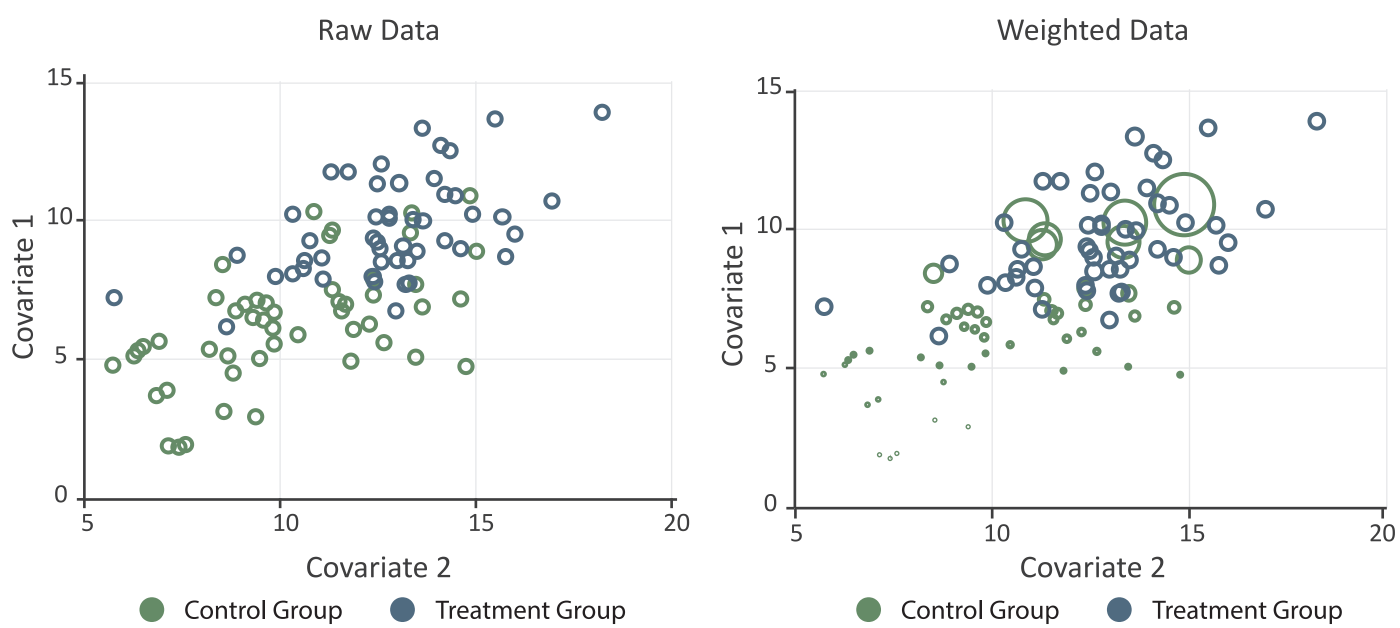
8. Agency stop volume = $$\frac{\textit{Total \# of stops by agency on day of stop}}{\textit{Maximum \# of daily stops by } \textbf{\textit{agency}} \textit{ over year of analysis}}$$

9. County stop volume = $$\frac{\textit{Total \# of stops by agency on day of stop}}{\textit{Maximum \# of daily stops by } \textbf{\textit{county}} \textit{ over year of analysis}}$$

The first step of the analysis uses a probit model to estimate the propensity of being in the treatment group based on the covariates listed above. Overlap of propensity scores is evaluated and any non-overlapping observations are removed from the sample. Inverse Probability Weights (IPWs) are estimated for each observation based on the propensity scores. For the treatment group in an ATET framework these weights are equal to 1. For the control group the weight is equal to $p/(1-p)$, where p is the propensity score (see footnote 31). In effect, this process gives more weight to control observations that have a higher propensity score (i.e., are more similar to treated observations) and treated observations that have a lower propensity score (i.e., are more similar to control observations).

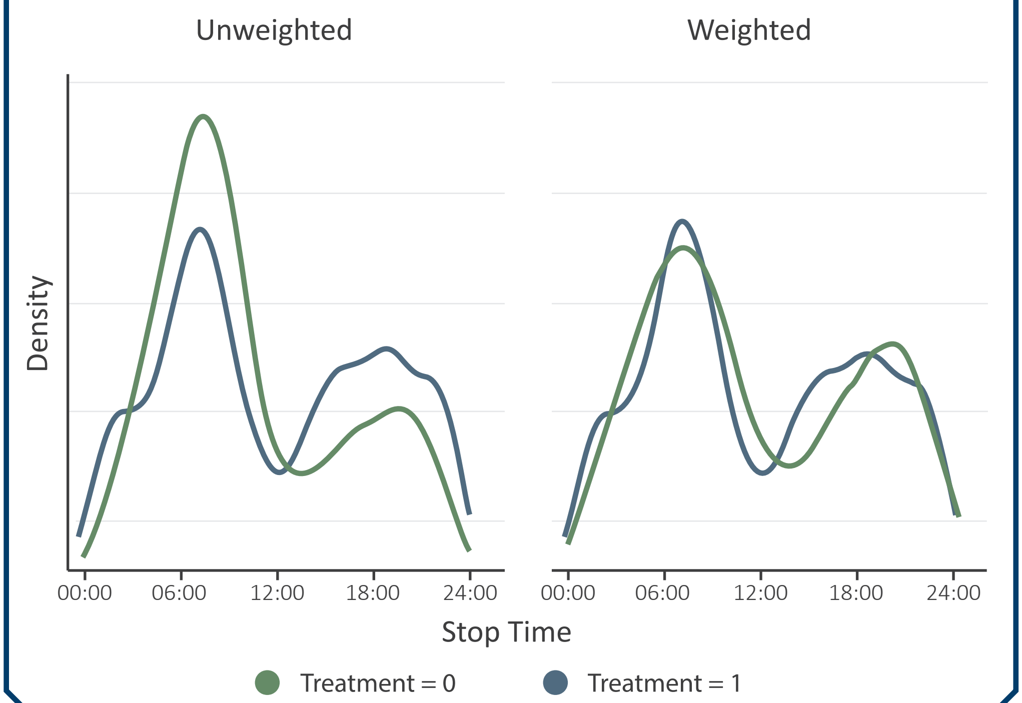
A hypothetical example application of IPWs is in **Figure D.2** below. The two graphs each represent control and treatment group observations and their respective values for each of two covariates. While there is some overlap between the groups in this example, the treatment (light gray) group tends to have higher values of both variables. In the Raw Data (unweighted) we can see that the two groups are not directly comparable. After calculating IPWs for ATET these weights are applied to the two groups and represented by the size of the circles in the Weighted Data graph. The treatment group remains the same here since the weights = 1, but the importance or weight of control group observations are adjusted. The observations that are closer to the treatment group observations are given a large weight, while those that are not are given a small weight. The weighted control group, as a whole, has observations that are much closer to those of the treatment group than the raw control group.

Figure D.2.
Weighting Example



Balance is then measured based on the standardized difference[39] in means and the variance ratio[40] between the treatment and control groups for each of the raw data set and the inverse probability weighted data set. If the resulting standardized difference in the weighted data set is close to zero and the variance ratio is close to 1 for each variable for the weighted data then the sample is said to be balanced. Balance was evaluated in every data subset by agency and strong balance was achieved in every instance, e.g., the standardized differences were always close to zero (usually within .01 of 0, always within 0.05) and the variance ratios were always close to one (usually within .01 of 1, always within 0.05) (Austin 2009a; 2009b). In every case, the data sets were relatively well balanced in the initial, raw data sets, but became more balanced through the weighting process. This balance can also be evaluated graphically for each variable. **Figure D.3** is an example of one of these variables for one agency. The Unweighted chart displays the distribution of stop time for each of the treated group and the untreated group. The Weighted chart displays these same distributions with the IPWs applied. The distributions of the two groups more closely resemble each other in the weighted graph than in the unweighted graph, so we can say that these groups are more balanced when incorporating the IPWs.

Figure D.3.
Confounding Variable Balance Example



Outcome equations are then estimated for each of the treatment variables across four sets of outcomes:

1. = 0 if a warning/none disposition is observed, = 1 otherwise
2. = 1 if a citation disposition is observed, = 0 if warning/none outcome is observed
3. = 1 if a search disposition is observed, = 0 if a citation or warning/none outcome is observed
4. = 1 if an arrest disposition is observed, = 0 otherwise

In the next step, probit models with the inverse probability weights applied and robust standard errors are estimated for each of the treatment and control groups. Predicted outcomes are stored for each observation and their average yields the potential outcome mean for the control group. The comparison between this mean and the actual average of the treatment group yields the Average Treatment Effect on the Treated (ATET), the

39 The standardized difference of variable x is: $\delta_x = \dfrac{\mu_x(t=1) - \mu_x(t=0)}{\sqrt{\dfrac{\sigma_x^2(t=1) - \sigma_x^2(t=0)}{2}}}$

40 The variance ratio is simply the variance of the treated group divided by the variance of the control group.

main estimate of interest in these models. This estimate is slightly different from the Average Treatment Effect as it focuses specifically on the effect on the treated group rather than the population as a whole. In this case, the estimates may be interpreted as the average difference in predicted probability of the outcome if the treated group (Black/Hispanic/Black or Hispanic) had identical characteristics to the control group, except had a race/ ethnicity = White [41].

Limitations

As with any statistical analysis, there are potential shortcomings of IPWRA analysis that may hinder the validity of the results. In this case, the largest concerns are the data limitations that result in the omission of some confounding variables that may be theoretically relevant. Comparable analyses of bias in police stops in other localities have controlled for additional confounding variables not included here, including police officer identifiers, make/model/year of vehicle, and location of the stop. Other variables may influence officer decision criteria, but are rarely included in the comparable analyses in other states due to data availability challenges. These variables include economic characteristics of the driver (i.e., employment status, income, etc.) and information on the driving population from which drivers are stopped. This later variable poses significant estimation challenges as it requires several assumptions regarding directions, populations, time of travel, and frequencies of commuters and tourists at each location in the road system. Without significant preliminary data about these factors any estimation of the driving population is likely to incorporate a significant amount of bias to any effect estimates built on top of these estimates.

Many of these variables are not described in the statutes establishing Oregon's STOP data tracking system (e.g., make/model). Other variables, such as geographic location of the stop, are highly varied in quality and format across these Oregon agencies. Some Oregon agencies provide precise longitude and latitude of the traffic stop via automatic logging in the cellphone app, other agencies allow officers to enter nearest intersections or mile markers, and others require no location to be entered by their officers. Due to this lack of uniformity in reporting, the STOP Project research team could not include location information for some agencies with high quality location information while also conducting uniform analyses across all of the Tier 1 agencies.

The omission of important confounding variables leads to the low Pseudo-R2s in the results and also drives the high amount of balance found in the raw data. In each sub-analysis the balancing procedure leads to greater confounder balance than in the raw data, but the groups were not egregiously unbalanced in the raw data. A high number of the confounders are binary indicator variables, which makes it easier to form very close matches and leads to less imbalance in the raw data, but this also shows that these variables may be imprecisely measured.

Results

The threshold for identifying an effect as significant is 95 percent confidence interval or a p-value of 0.05 or less. For each agency, however, we are conducting 12 tests (20 for statewide). A Bonferroni adjustment[42] is warranted in this situation to adjust for the likelihood of a given test resulting in a false positive. The appropriate threshold for each agency is thus, p-value $\leq 0.05/12=0.0041$ (statewide p-value $\leq 0.05/20=0.0025$). Statistically significant disparities are indicated by a * next to the average treatment effect on the treated (ATET) in the tables below.

41 Conversely, the ATE is predicts these differences for both the treated group and for the untreated group and averages all these differences. Thus, it estimates the difference in predicted probabilities for both the White group and the Black/Hispanic/Black-Hispanic groups and averages across all observations.

42 Weisstein, Eric W. "Bonferroni Correction." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/BonferroniCorrection.html

Table D.1.1
Statewide Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Asian	Citation	0.3720	-0.0072	0.0041	-1.78	0.075	-0.0196	0.0051
Asian	Search	0.0168	-0.0101	0.0011	-8.84	0.000	-0.0136	-0.0066
Asian	Arrest	0.0184	-0.0098	0.0012	-8.36	0.000	-0.0134	-0.0063
Asian	Citation, Search, or Arrest	0.3863	-0.0151	0.0041	-3.68	0.000	-0.0275	-0.0027
Black	Citation	0.3609	-0.0044	0.0036	-1.21	0.227	-0.0154	0.0066
Black	Search	0.0631	0.0226	0.0018	12.53	0.000	0.0171	0.0281
Black	Arrest	0.0500	0.0082	0.0016	5.06	0.000	0.0033	0.0131
Black	Citation, Search, or Arrest	0.4090	0.0089	0.0036	2.44	0.015	-0.0021	0.0199
Hispanic	Citation	0.4300	0.0559	0.0024	23.28	0.000	0.0486	0.0631
Hispanic	Search	0.0365	0.0064	0.0009	6.94	0.000	0.0036	0.0092
Hispanic	Arrest	0.0345	0.0045	0.0009	5.09	0.000	0.0018	0.0072
Hispanic	Citation, Search, or Arrest	0.4540	0.0578	0.0024	24.16	0.000	0.0506	0.0650
Middle Eastern	Citation	0.3416	-0.0220	0.0062	-3.57	0.000	-0.0406	-0.0034
Middle Eastern	Search	0.0106	-0.0163	0.0014	-11.33	0.000	-0.0206	-0.0119
Middle Eastern	Arrest	0.0085	-0.0187	0.0013	-14.21	0.000	-0.0226	-0.0147
Middle Eastern	Citation, Search, or Arrest	0.3495	-0.0366	0.0062	-5.91	0.000	-0.0554	-0.0179
Native Am.	Citation	0.3614	0.0181	0.0102	1.77	0.076	-0.0128	0.0491
Native Am.	Search	0.0450	0.0121	0.0042	2.85	0.004	-0.0007	0.0249
Native Am.	Arrest	0.0500	0.0160	0.0044	3.61	0.000	0.0026	0.0294
Native Am.	Citation, Search, or Arrest	0.3977	0.0282	0.0102	2.76	0.006	-0.0027	0.0591

Table D.1.2
Statewide Results No OSP

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Asian	Citation	0.3486	-0.0240	0.0049	-4.86	0.000	-0.0389	-0.0091
Asian	Search	0.0191	-0.0132	0.0016	-8.49	0.000	-0.0179	-0.0085
Asian	Arrest	0.0226	-0.0113	0.0017	-6.83	0.000	-0.0163	-0.0063
Asian	Citation, Search, or Arrest	0.3667	-0.0335	0.0050	-6.67	0.000	-0.0486	-0.0183
Black	Citation	0.3214	-0.0366	0.0042	-8.65	0.000	-0.0494	-0.0238
Black	Search	0.0778	0.0302	0.0024	12.67	0.000	0.0230	0.0374
Black	Arrest	0.0601	0.0111	0.0021	5.28	0.000	0.0048	0.0175
Black	Citation, Search, or Arrest	0.3843	-0.0164	0.0043	-3.82	0.000	-0.0294	-0.0034
Hispanic	Citation	0.4068	0.0337	0.0032	10.46	0.000	0.0240	0.0434
Hispanic	Search	0.0479	0.0093	0.0015	6.38	0.000	0.0049	0.0138
Hispanic	Arrest	0.0459	0.0077	0.0014	5.42	0.000	0.0034	0.0119
Hispanic	Citation, Search, or Arrest	0.4397	0.0389	0.0032	11.97	0.000	0.0291	0.0487
Middle Eastern	Citation	0.3143	-0.0238	0.0078	-3.05	0.002	-0.0475	-0.0002
Middle Eastern	Search	Failed to converge						
Middle Eastern	Arrest	Failed to converge						
Middle Eastern	Citation, Search, or Arrest	0.3226	-0.0461	0.0079	-5.81	0.000	-0.0701	-0.0221
Native Am.	Citation	0.3471	0.0013	0.0134	0.10	0.921	-0.0391	0.0418
Native Am.	Search	Failed to converge						
Native Am.	Arrest	0.0467	0.0066	0.0057	1.16	0.247	-0.0106	0.0238
Native Am.	Citation, Search, or Arrest	0.3808	0.0048	0.0136	0.35	0.724	-0.0364	0.0460

Table D.1.3
Beaverton PD Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.3065	-0.0046	0.0115	-0.400	0.689	-0.0377	0.0285
Black	Search	0.0769	0.0172	0.0070	2.460	0.014	-0.0029	0.0373
Black	Arrest	0.0773	0.0165	0.0068	2.430	0.015	-0.0030	0.0359
Black	Citation, Search, or Arrest	0.3720	0.0103	0.0120	0.860	0.391	-0.0242	0.0449
Hispanic	Citation	0.3426	0.0138	0.0087	1.590	0.113	-0.0112	0.0389
Hispanic	Search	0.0696	0.0194	0.0048	4.070	0.000	0.0057	0.033
Hispanic	Arrest	0.0648	0.0151	0.0045	3.310	0.001	0.0020	0.0281
Hispanic	Citation, Search, or Arrest	0.3961	0.0280	0.0089	3.130	0.002	0.0023	0.0536
Black-Hispanic	Citation	0.3313	0.0081	0.0074	1.090	0.276	-0.0133	0.0295
Black-Hispanic	Search	0.0719	0.0186	0.0042	4.480	0.000	0.0067	0.0306
Black-Hispanic	Arrest	0.0687	0.0156	0.0040	3.880	0.000	0.0041	0.0271
Black-Hispanic	Citation, Search, or Arrest	0.3885	0.0225	0.0077	2.930	0.003	0.0004	0.0445

Table D.1.4
Clackamas Co SO Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.3440	0.0165	0.0136	1.210	0.225	-0.0225	0.0555
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.3631	0.0213	0.0139	1.530	0.125	-0.0185	0.0611
Hispanic	Citation	0.3737	0.0147	0.0092	1.590	0.113	-0.0119	0.0412
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.3880	0.0153	0.0094	1.630	0.102	-0.0116	0.0422
Black-Hispanic	Citation	0.3650	0.0152	0.0079	1.920	0.055	-0.0075	0.0378
Black-Hispanic	Search	Failed to converge						
Black-Hispanic	Arrest	0.0236	0.0061	0.0027	2.220	0.027	-0.0018	0.0140
Black-Hispanic	Citation, Search, or Arrest	0.3807	0.0171	0.0080	2.130	0.033	-0.0059	0.0402

Table D.1.5
Eugene PD Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.3841	0.0272	0.0200	1.360	0.174	-0.0302	0.0845
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.4394	0.0172	0.0197	0.870	0.383	-0.0393	0.0737
Hispanic	Citation	0.3839	0.0217	0.0210	1.040	0.300	-0.0384	0.0819
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.4365	0.0166	0.0206	0.810	0.419	-0.0425	0.0758
Black-Hispanic	Citation	0.3840	0.0250	0.0150	1.670	0.095	-0.0180	0.068
Black-Hispanic	Search	0.0817	-0.0052	0.0084	-0.620	0.536	-0.0294	0.019
Black-Hispanic	Arrest	0.0787	-0.0010	0.0082	-0.120	0.908	-0.0246	0.0227
Black-Hispanic	Citation, Search, or Arrest	0.4380	0.0173	0.0147	1.170	0.240	-0.0250	0.0596

Table D.1.6
Gresham PD Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.2882	-0.0497	0.0134	-3.700	0.000	-0.0882	-0.0112
Black	Search	0.0770	0.0199	0.0080	2.490	0.013	-0.0030	0.0429
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.3471	-0.0331	0.0139	-2.390	0.017	-0.0729	0.0067
Hispanic	Citation	0.3804	0.0173	0.0133	1.300	0.195	-0.0210	0.0556
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.4114	0.0157	0.0136	1.150	0.250	-0.0234	0.0548
Black-Hispanic	Citation	0.3361	-0.0147	0.0104	-1.410	0.160	-0.0446	0.0153
Black-Hispanic	Search	0.0634	0.0096	0.0057	1.700	0.089	-0.0066	0.0259
Black-Hispanic	Arrest	0.0391	-0.0006	0.0045	-0.140	0.892	-0.0136	0.0124
Black-Hispanic	Citation, Search, or Arrest	0.3800	-0.0078	0.0107	-0.730	0.468	-0.0385	0.0229

Table D.1.7
Hillsboro PD Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.2117	0.0258	0.0169	1.530	0.126	-0.0226	0.0742
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.2426	0.0304	0.0173	1.750	0.079	-0.0193	0.0802
Hispanic	Citation	0.2758	0.0740	0.0101	7.290	0.000	0.0449	0.1031
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.3012	0.0771	0.0104	7.450	0.000	0.0474	0.1068
Black-Hispanic	Citation	0.2628	0.0644	0.0092	7.000	0.000	0.0380	0.0908
Black-Hispanic	Search	0.0339	0.0109	0.0037	2.960	0.003	0.0003	0.0214
Black-Hispanic	Arrest	0.0302	0.0050	0.0035	1.410	0.159	-0.0052	0.0151
Black-Hispanic	Citation, Search, or Arrest	0.2893	0.0677	0.0094	7.200	0.000	0.0407	0.0947

Table D.1.8
Marion Co SO Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	Failed to converge						
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.8478	-0.0132	0.0187	-0.710	0.478	-0.0668	0.0403
Hispanic	Citation	0.8586	0.0251	0.0078	3.210	0.001	0.0027	0.0476
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.8616	0.0251	0.0078	3.210	0.001	0.0027	0.0476
Black-Hispanic	Citation	0.8568	0.0205	0.0075	2.750	0.006	-0.0009	0.0420
Black-Hispanic	Search	0.0173	0.0033	0.0028	1.160	0.245	-0.0048	0.0114
Black-Hispanic	Arrest	0.0206	0.0039	0.0030	1.280	0.202	-0.0048	0.0125
Black-Hispanic	Citation, Search, or Arrest	0.8598	0.0205	0.0075	2.750	0.006	-0.0009	0.0419

Table D.1.9
Medford PD Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	Insufficient sample size						
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.3519	0.0192	0.0294	0.650	0.513	-0.0651	0.1035
Hispanic	Citation	0.2809	-0.0016	0.0194	-0.080	0.935	-0.0571	0.0540
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.3509	0.0071	0.0203	0.350	0.727	-0.0513	0.0655
Black-Hispanic	Citation	0.2734	0.0067	0.0166	0.400	0.687	-0.0410	0.0544
Black-Hispanic	Search	0.1137	0.0086	0.0114	0.760	0.448	-0.0240	0.0412
Black-Hispanic	Arrest	0.0847	0.0049	0.0098	0.490	0.621	-0.0233	0.0330
Black-Hispanic	Citation, Search, or Arrest	0.3512	0.0113	0.0174	0.650	0.516	-0.0386	0.0612

Table D.1.10
Multnomah Co SO Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.2464	0.0459	0.0146	3.140	0.002	0.004	0.0878
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.2759	0.0431	0.0151	2.860	0.004	-0.0002	0.0865
Hispanic	Citation	0.2407	0.0279	0.0126	2.220	0.027	-0.0082	0.0641
Hispanic	Search	Insufficient sample size						
Hispanic	Arrest	Insufficient sample size						
Hispanic	Citation, Search, or Arrest	0.2647	0.0231	0.0129	1.790	0.074	-0.0140	0.0602
Black-Hispanic	Citation	0.2430	0.0356	0.0101	3.530	0.000	0.0066	0.0647
Black-Hispanic	Search	0.0322	0.0015	0.0043	0.350	0.729	-0.0108	0.0137
Black-Hispanic	Arrest	0.0308	-0.0034	0.0042	-0.810	0.418	-0.0153	0.0086
Black-Hispanic	Citation, Search, or Arrest	0.2694	0.0317	0.0104	3.040	0.002	0.0018	0.0616

Table D.1.11
Oregon State Police Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.4541	0.0729	0.0065	11.230	0.000	0.0543	0.0915
Black	Search	0.0264	0.0039	0.0021	1.890	0.058	-0.0020	0.0099
Black	Arrest	0.0245	0.0008	0.0020	0.380	0.702	-0.0049	0.0064
Black	Citation, Search, or Arrest	0.4710	0.0732	0.0064	11.460	0.000	0.0548	0.0915
Hispanic	Citation	0.4560	0.0793	0.0035	22.990	0.000	0.0694	0.0892
Hispanic	Search	0.0234	0.0029	0.0011	2.740	0.006	-0.0001	0.0059
Hispanic	Arrest	0.0213	0.0001	0.0010	0.110	0.914	-0.0027	0.0029
Hispanic	Citation, Search, or Arrest	0.4707	0.0787	0.0034	23.040	0.000	0.0689	0.0885
Black-Hispanic	Citation	0.4556	0.0781	0.0031	24.930	0.000	0.0691	0.0871
Black-Hispanic	Search	0.0240	0.0031	0.0010	3.220	0.001	0.0003	0.0059
Black-Hispanic	Arrest	0.0220	0.0002	0.0009	0.250	0.799	-0.0024	0.0028
Black-Hispanic	Citation, Search, or Arrest	0.4707	0.0776	0.0031	25.070	0.000	0.0687	0.0865

Table D.1.12
Portland PB Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.3268	-0.0847	0.0073	-11.660	0.000	-0.1055	-0.0638
Black	Search	0.1095	0.0611	0.0045	13.440	0.000	0.0480	0.0741
Black	Arrest	0.0735	0.0183	0.0038	4.770	0.000	0.0073	0.0292
Black	Citation, Search, or Arrest	0.4181	-0.0377	0.0074	-5.110	0.000	-0.0589	-0.0166
Hispanic	Citation	0.4515	-0.0099	0.0092	-1.070	0.283	-0.0364	0.0166
Hispanic	Search	0.0588	0.0144	0.0045	3.230	0.001	0.0016	0.0273
Hispanic	Arrest	0.0610	0.0070	0.0045	1.560	0.118	-0.0058	0.0198
Hispanic	Citation, Search, or Arrest	0.4995	-0.0005	0.0092	-0.050	0.961	-0.0269	0.0260
Black-Hispanic	Citation	0.3733	-0.0569	0.0062	-9.160	0.000	-0.0747	-0.0391
Black-Hispanic	Search	0.0912	0.0442	0.0035	12.590	0.000	0.0341	0.0543
Black-Hispanic	Arrest	0.0690	0.0141	0.0031	4.500	0.000	0.0051	0.0232
Black-Hispanic	Citation, Search, or Arrest	0.4475	-0.0244	0.0063	-3.890	0.000	-0.0424	-0.0064

Table D.1.13
Salem PD Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.5143	0.0568	0.0261	2.180	0.029	-0.0181	0.1317
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.5526	0.0580	0.0254	2.290	0.022	-0.0148	0.1308
Hispanic	Citation	0.5126	0.0478	0.0117	4.070	0.000	0.0141	0.0815
Hispanic	Search	0.0740	0.0138	0.0063	2.200	0.028	-0.0042	0.0318
Hispanic	Arrest	0.0662	0.0131	0.0058	2.250	0.025	-0.0037	0.0299
Hispanic	Citation, Search, or Arrest	0.5457	0.0534	0.0115	4.650	0.000	0.0204	0.0864
Black-Hispanic	Citation	0.5128	0.0490	0.0112	4.390	0.000	0.0170	0.0810
Black-Hispanic	Search	0.0775	0.0157	0.0060	2.610	0.009	-0.0016	0.0331
Black-Hispanic	Arrest	0.0679	0.0136	0.0056	2.420	0.016	-0.0026	0.0297
Black-Hispanic	Citation, Search, or Arrest	0.5467	0.0539	0.0109	4.940	0.000	0.0226	0.0852

Table D.1.14
Washington Co SO Results

Race/Ethnicity	Outcome	Group Mean	ATET	Robust S.E.s	z-score	p-value	95% CI	
Black	Citation	0.2503	-0.0186	0.0115	-1.620	0.105	-0.0516	0.0144
Black	Search	Insufficient sample size						
Black	Arrest	Insufficient sample size						
Black	Citation, Search, or Arrest	0.2951	-0.0122	0.0120	-1.010	0.311	-0.0468	0.0223
Hispanic	Citation	0.3210	0.0416	0.0072	5.790	0.000	0.0210	0.0622
Hispanic	Search	0.0443	0.0059	0.0032	1.860	0.063	-0.0032	0.015
Hispanic	Arrest	0.0468	0.0095	0.0032	2.980	0.003	0.0004	0.0186
Hispanic	Citation, Search, or Arrest	0.3575	0.0476	0.0074	6.480	0.000	0.0265	0.0687
Black-Hispanic	Citation	0.3051	0.0280	0.0064	4.350	0.000	0.0095	0.0464
Black-Hispanic	Search	0.0461	0.0060	0.0030	2.020	0.044	-0.0025	0.0144
Black-Hispanic	Arrest	0.0469	0.0077	0.0029	2.640	0.008	-0.0007	0.0161
Black-Hispanic	Citation, Search, or Arrest	0.3434	0.0341	0.0066	5.140	0.000	0.0150	0.0531

Appendix E – KPT Hit-Rate Analysis Technical Appendix

Model and Assumptions

The hit-rate analyses performed in this report are based on the model presented by Knowles, Persico, and Todd (2001) which details how police and citizens act surrounding searches. In this model, police officers are assumed to make the decision to search someone based on their perception of the likelihood that the person will have contraband in their possession, while also accounting for the economic "cost" of a search. In the case that the cost of searching members of different groups is the same, we expect officers to search the group that they perceive to be more likely to possess contraband. Similarly, this model assumes that citizens make the decision to carry contraband based on their perception of the likelihood that they will be caught with contraband. If a particular group is more likely to carry contraband, they will be searched more often by police. As a group, they will respond by reducing their likelihood to carry contraband in order to reduce their risk of being caught. In this way, any differences in groups' likelihoods to carry contraband and to be searched by police should tend toward an equilibrium. At equilibrium we expect that the hit-rate (the rate at which searches are "successful", or result in finding contraband) should be equal across groups, whereas unequal hit-rates indicate disparate search practices.

The Knowles, Persico, and Todd (KPT) Hit-Rate Model assesses whether police are participating in racial/ethnic discrimination by oversearching members of a particular group. If a group is "over searched" (searched more often than necessary to maintain the abovementioned equilibrium), then the hit-rate for that group will be lower than that of a baseline group. In our case, if a minority racial/ethnic group is "over searched", then the hit-rate for that group will be lower than that of Whites, perhaps indicating what Becker calls "a taste for discrimination" (1957) (a phrase coined to describe economic discrimination) in officers conducting searches.

Hit-Rate and Significance Calculation

The hit-rate for a group is simply a proportion. The total number of searches of a group is represented by s and the number of searches of that group which result in finding contraband is represented by *f*:

$$\text{KPT Hit-Rate} = \frac{f}{s}$$

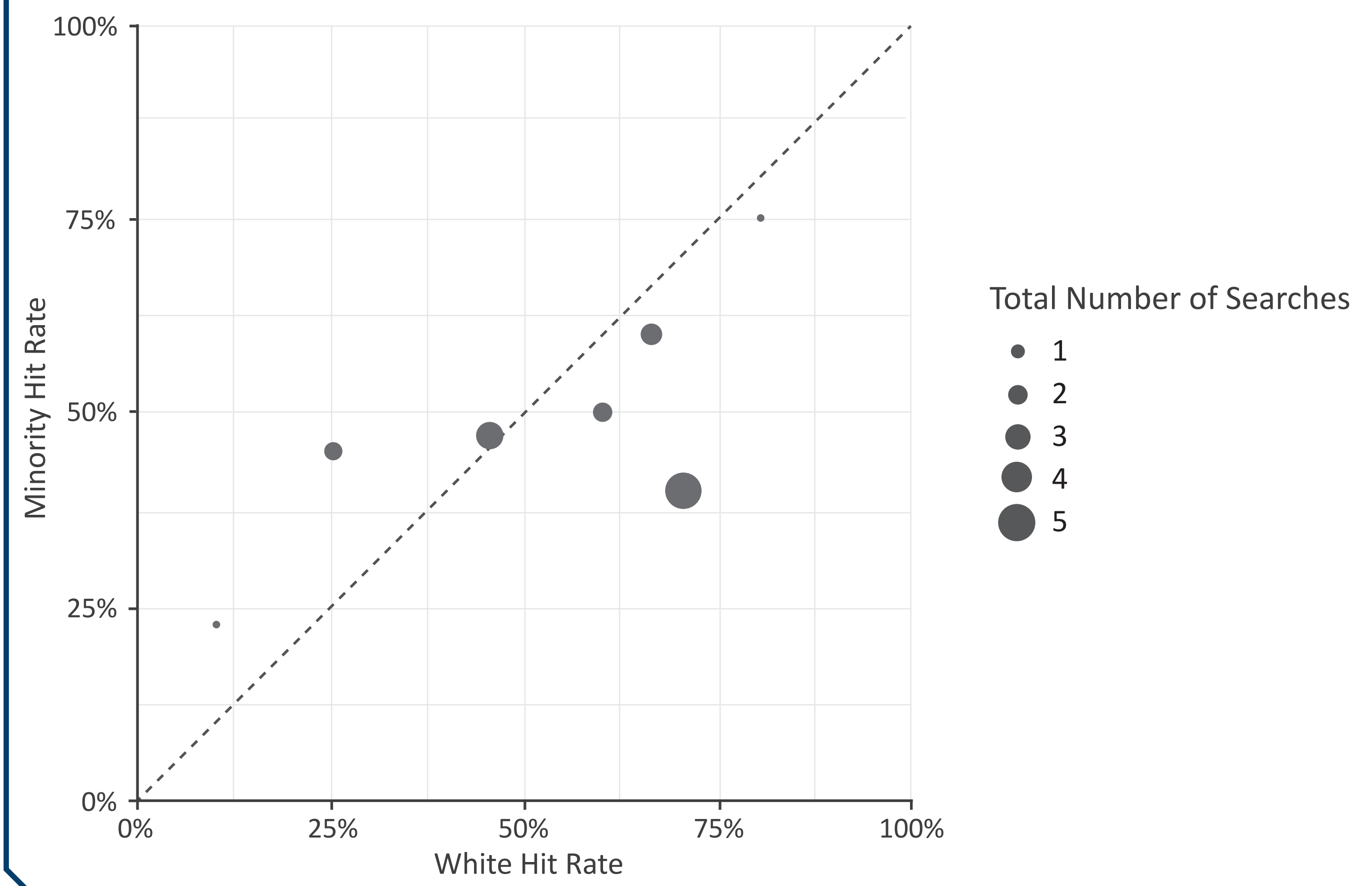
After calculating hit-rates by agency for each racial/ethnic group, chi-square tests of independence were performed in order to determine whether differences in the hit-rates were statistically significant. Yates's continuity correction for the chi-squared test was used to mitigate the test's tendency to produce low p-values due to the discrete nature of the data. However, no substantive difference arose between the results when performed with or without the continuity correction. A confidence level of 95 percent with a Bonferroni correction for multiple testing determined significance. Each agency's White hit-rate was compared to each of three groups (Black, Hispanic, and combined Black/Hispanic), so a Bonferroni corrected p-value of 0.05/3 = 0.017 or lower was considered indicative of a statistically significant difference between minority and White hit-rates. For statewide results, a Bonferroni corrected p-value of 0.05/5 = 0.01 was used, as the statewide White hit-rate was compared to each of five groups (Black, Hispanic, Asian/PI, Middle Eastern, and Native American). Hit-Rate analyses and accompanying statistical tests were performed with the statistical software R[43].

43 R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL **https://www.R-project.org/**.

Visualization

Figure E1 is a stylized example of the figures that appear in the body of the report and on the accompanying data dashboards found online.[44] Each agency is a single point, with the location of the point indicating the relationship between White and comparison group hit-rates for that agency. The size of the point indicates the total number of searches (White and minority) analyzed in the significance test. Agencies below the diagonal line have a lower minority hit-rate than White hit-rate, which could be indicative of disparate searching practices. An agency falling above or below the line does not, alone, indicate a significant difference between White and minority hit-rates. Significance is noted in the text and tables accompanying figures in the body of the report. It is possible for an agency below the line to have no statistically significant difference between White and minority hit-rates, while an agency "closer" to the line may be found to have a significant difference. The power of the chi-square test to determine significance is dependent upon the sample size (in this case, number of searches). An agency with fewer searches may be "further" below the line than an agency with more searches, yet fail to be statistically significant. Hit-Rate analysis figures are created using the ggplot2 package in R[45].

Figure E1.
Traffic and Pedestrian Stops by Month of Year



44 Inspiration for these figures came from the display used by the Stanford Open Policing Project, which is located at https://openpolicing.stanford.edu/.
45 H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Limitations

One important assumption of the KPT Hit-Rate model is that all searches included in the analysis are discretionary. Some searches, such as those made incident to arrest, are non-discretionary, meaning that there is no individual choice (discretion) in the officer's decision to conduct the search. This type of search is not representative of officers' motivations and cannot be used to determine any patterns of behavior. In the STOP Program training that all officers complete prior to submitting data for this study, officers are informed that non-discretionary searches should not be included in the data. This means that when a stop results in an officer arresting someone, although they will always do a "pat-down" to ensure safety at the time of arrest, we should not always see a search recorded for the stop (as these pat-downs are non-discretionary searches). In some cases, the data seem to show records of searches incident to arrest, however it is not possible to distinguish these "mistakes" from true records of discretionary searches. Accordingly, STOP Program researchers chose to take all data at face value – that is, if a search was recorded, it is included in the KPT Hit-Rate analysis as a discretionary search.

Similarly, consent searches can be less representative of officers' decision-making than non-discretionary, non-consent searches. As a robustness check, STOP Program researchers performed the KPT Hit-Rate analyses after removing all consent searches (which make up over half of all searches reported for Tier 1 agencies in the first year of data collection). This resulted in a dramatic reduction in sample sizes, increase in the number of agencies for which analyses could not be performed, and decrease in the power of the Hit-Rate test for those that could be performed. In future years when more discretionary, non-consent searches have been reported, the STOP Program will present Hit-Rate analyses both including and excluding consent searches, however in this first year of analysis, the sample sizes are simply too low to perform reliable analysis without the inclusion of consent searches.

A possible methodological limitation of the hit-rate test is the problem of infra-marginality (Simoiu, 2017). Infra-marginality is best explained by example. Suppose that group A has some portion of members that carry contraband 55 percent of the time (while all other members of the group carry contraband less than 50 percent of the time). Suppose also that group B has some portion of members that instead carry contraband 75 percent of the time (while all other members of the group carry contraband less than 50 percent of the time). If an officer only searches every person (regardless of group) who has over a 50 percent chance of carrying contraband, then group A will have a lower hit-rate. In the hit-rate test, this would appear to indicate discrimination against group A, despite the true "group-neutral" manner of the officer's search decisions. While this is one of the widest criticisms of the KPT Hit-Rate test, Persico (of Knowles, Persico, and Todd) independently addressed the criticism of this limitation in a follow up paper. Persico (2009) argues that inframarginality is alleviated by the allowance in the model for searched groups to respond to search intensity (by lowering their propensity to carry contraband when searched more frequently). This is consistent with KPT's initial assertion that subgroups, as well as larger racial/ethnic groups, should act similarly to larger groups in that they adjust their propensity to carry contraband according to their likelihood of being searched.

Appendix F – Predicted Disposition Follow Up with Oregon State Police

Following the drafting of results from the analyses of STOP Program data, the STOP Program team reached out to the Oregon State Police (OSP) to discuss a disparity STOP Program researchers discovered for OSP with regard to its citation practices (see Section 5 of this report). Through discussions, OSP identified a possible cause of this disparity, as OSP has a long standing internal policy mandating that certain types of violations must result in a citation (see Department of State Police Policy 502.15, below). Importantly, if this policy contributes to the disparity in citation outcomes, then it would identify a portion of the disparity that should not be tied to discretionary decisions made by officers in the field.

To test the hypothesis that these non-discretionary citations could impact disparities discovered for citations via the Predicted Disposition Analysis, OSP provided the STOP Program with an amended full year data set which contained a flag denoting when a citation was given for three types of mandatory citation offenses related to 502.15(III)(A)(1)(a)(5): (i) driving while uninsured, (ii) driving with a suspended license, and (iii) driving with no license.

The STOP research team included this data in the baseline Predicted Disposition Analysis in two ways: first, the flag indicator variable was included in the balancing operation and in the outcome regression equation, which ensures that the comparison groups have equal rates of these citation reasons and then controls for any remaining variation through regression. Second, STOP Program researchers omitted all results that included one of these citations from the baseline data set and reran the baseline analysis. Results of these approaches for the citation analysis are reported below in **Table F.1**.

Table F.1.
Oregon State Police Predicted Disposition

Agency	Race/Ethnicity	Citation Actual	Citation Predicted	Δ
Baseline Analysis	Black	45.4%	38.1%*	7.3
	Hispanic	45.6%	37.7%*	7.9
	Black or Hispanic	45.6%	37.8%*	7.8
Balancing on Mandatory Cite Outcomes	Black	45.4%	42.1%*	3.3
	Hispanic	45.6%	43.3%*	2.3
	Black or Hispanic	45.6%	41.8%*	3.8
Omitting Mandatory Cite Observations	Black	32.3%	30.7%	1.6
	Hispanic	33.9%	31.0%*	2.9
	Black or Hispanic	33.6%	30.9%*	2.7

* Indicates a statistically significant difference with the Bonferroni correction

As reported in Section 5 of this report, OSP has a statistically significant disparity for both Black and Hispanic individuals with regard to citations. These results are reported in **Table F.1** in the first three rows designed as the "Baseline Analysis," and report that while Black individuals should have been cited 38.1 percent of the time, they were actually cited 45.4 percent of the time (Δ = 7.3 percentage points). Similarly, Hispanic individuals were predicted to have been cited 37.7 percent of the time but were actually cited in 45.6 percent of stops (Δ = 7.9 percentage points).

The results from the additional analyses are found in the lower rows of Table F1. When the flag for driving while uninsured, driving with a suspended license, and driving with no license was included in the propensity score analysis, the disparities for OSP dropped substantially (see rows 4 through 6). For instance, while the difference in the predicted versus actual citation rates in the baseline data for Black individuals was 7.3 percentage points, the difference in the actual versus predicted citation rates in the "balanced" data was only 3.3 percentage points, which represents a 50 percent reduction in the disparity reported for this group. Similarly, for Hispanics, the difference in the baseline model was 7.9 percentage points but was 2.3 percentage points for the "balanced" model, a 71 percent reduction in the reported disparity.

In the final three rows of **Table F.1**, results are presented when these stops were removed from the analysis completely. In this case the gaps between the actual and predicted citation rates were small, although they could not be compared to the baseline analysis because of the differences in the sample after these mandatory citation stops were removed.

It is important to note that these analyses are preliminary and that future work should be done to include the other offenses included in OSP Policy 502.15. The inclusion of this information, however, demonstrates that to fully understand disparate outcomes in police-citizen interactions, additional data is often needed. These analyses also demonstrate the complexity of investigating police-citizen interactions. Indeed, in this specific case, the data demonstrates that disparities cannot and should not always be attributed to potential bias on the part of officers, as other factors can lead to disparate outcomes beyond individual officers' control.

DEPARTMENT OF STATE POLICE

ENFORCEMENT

Effective Date: August 3, 2015	Supersedes Date: August 7, 2000	Policy Number: **502.15**
Reference/Laws/Statutory Authority: ORS 153		
Applies to: ☐ All Personnel ☐ All Management ☑ Sworn Personnel ☐ Non Sworn Personnel ☐ Other		
Issuing Authority: Superintendent of State Police		No. Pages **2**

I. PURPOSE

The purpose of this policy it to provide some guidelines to sworn employees when considering what level of enforcement action should be taken during a violator contact.

II. POLICY

Laws are not to be neglected, but in the successful execution of any well planned enforcement program, special emphasis must be directed toward the enforcement of certain parts of a particular code, when and where the need has been made apparent. This is not to be construed to mean that immunity is extended to violators of other sections of the same act, nor to persons who offend in some other manner.

A policy of strict enforcement does not necessarily imply intolerance and should not be so rigid as to deny to the officer discretion in the enforcement of violations.

III. PROCEDURE

A. A person who commits any of the following traffic offenses will be cited or arrested.

1. DUII (Arrests)

   a) In the event of very low breathe test readings, and where there is no indication of narcotic or dangerous drug use, consideration should be given to release on authority of magistrate, prosecutor, or jail recognizance officer. No arbitrary blood alcohol level will be set or suggested, rather, the sworn employee's good judgment under the circumstances existing at the time should be the controlling factor.

2. Eluding a sworn employee
3. Reckless driving
4. Fail to Perform the Duties of a Driver
5. Driving while suspended or revoked
6. Speed Racing

B. The enforcement action taken for other traffic offenses not specifically covered shall be left to the discretion of the sworn employee, dependent upon the circumstances of each individual offense.

C. Employees shall take into account the reasonable convenience of the violator and court when fixing time and place for court appearance.

D. The person being cited should be briefly advised of the violation charged, and the instructions on the summons should be explained to the offender.

E. Once a citation has been issued, the citation may not be dismissed or disposed of until the issuing sworn employee notifies their supervisor of the circumstances for dismissal or disposal of the citation, and receives approval if such action is appropriate.

F. The use of the warning provides the Department with the ability to tabulate statistics that are used to detect the number of vehicles and/or drivers using the highways who have committed specific violations. These activities are beneficial to the Department for statistical and legislative purposes, and it is important that they are properly documented in a written format.

IV. DEFINITIONS

None

V. RULES

A. All sworn employees shall be familiar with ORS 153 in regards to enforcement of offenses.