Statistical Transparency of Policing Report

Per House Bill 2355 (2017)

December 1, 2021

Revised April 15, 2022

Oregon Criminal Justice Commission

Ken Sanchagrin
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.

Table of Contents

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 third 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 as well as data collection software offered free of charge to all state law enforcement agencies. As of December 2021, the STOP Program has received at least one full year of data from 143 law enforcement agencies in the state and analyses using those data are presented in this report. This is the first stop report to include data from all Tier 1, 2, and 3 agencies.

Table E.1. Descriptive Statistics for Aggregate Year 3 Stop Data

Variable	Tier 1	Tier 2	Tier 3
Traffic Stop	97.8%	97.4%	94.6%
Race/Ethnicity			
Asian/PI	3.4%	2.6%	1.9%
Black	5.3%	3.2%	2.0%
Latinx	14.9%	13.0%	13.1%
Middle Eastern	1.3%	1.0%	0.7%
Native American	0.6%	0.4%	0.7%
White	74.5%	79.9%	81.8%
Gender			
Male	67.5%	64.4%	64.9%
Female	32.4%	35.4%	34.0%
Non-Binary	0.1%	0.1%	1.1%
Age			
Under 21	11.2%	12.8%	13.0%
21-29	25.3%	23.1%	23.0%
30-39	25.3%	24.9%	23.2%
40-49	16.5%	16.8%	16.2%
50 and Older	21.7%	22.5%	24.7%
Stop Disposition			
None	3.1%	8.5%	7.0%
Warning	57.2%	60.0%	68.9%
Citation	37.1%	28.9%	22.1%
Juvenile Summons	0.0%	0.0%	0.1%
Arrest	2.6%	2.6%	1.9%
Search Conducted	2.5%	1.9%	1.4%

Table E.1. reports descriptive statistics for the combined Tier 1, 2, and 3 data, which represents stops made from July 1, 2020 through June 30, 2021. Across all agencies, the vast majority of the reported data were for traffic stops, although the share of pedestrian stops made by Tier 3 agencies was higher than that for their larger counterparts. 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-quarter of Tier 1 stops and close to one-fifth of Tier 2 and Tier 3 stops involved Asian/PI, Black, Latinx, Middle Eastern, or Native American Oregonians. 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 for Tier 1, a little under 70 percent of stops for Tier 2, and three-quarters of stops for Tier 3.

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 (VOD) Analysis. 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 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 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: Oregon State Police. Specifically, results indicated that Oregon State Police had disparities in the Predicted Disposition Analysis with regard to citation patterns involving Asian/PI, Black, Latinx, Middle Eastern, and Native American individuals, with search patterns for Latinx and Native individuals, and with arrest patterns for Native American individuals. The KPT Hit-Rate test indicated a disparity with regard to searches of Middle Eastern individuals. Thus, it is recommended that Oregon State Police be examined in greater detail by STOP Program researchers and receive technical assistance from DPSST.

Tables and Figures

1. Background

This is the third annual report from the Statistical Transparency of Policing (STOP) Program. In 2017, the Oregon Legislature mandated that by July 2020 all Oregon law enforcement agencies were to collect data concerning all officer-initiated traffic and pedestrian stops. The mandate also required that the Oregon Criminal Justice Commission analyze the collected data to determine whether racial disparities exist in the treatment of Oregonians by law enforcement. To implement this mandate, the Legislature required the largest agencies to collect data first, followed by medium and smaller agencies in the intervening years. In December of 2019, the Criminal Justice Commission published its first annual STOP report, which contained data and analyses for the 12 largest law enforcement agencies in the state. In December 2020, the Criminal Justice Commission published its second report, which included an additional 39 mid-sized police agencies. This report builds on the first two by including analyses on an additional 92 small police agencies. The inclusion of these Tier 3 agencies means that this report analyzes stops from 143 law enforcement agencies in the state[1].

This report differs from previous reports by analyzing Tier 1 and Tier 2 agency stops using two years of data when possible, instead of a single year of data. The December 2019 STOP report analyzed stops for Tier 1 agencies using one year of data, from July 1, 2018 through June 30, 2019. The December 2020 STOP report analyzed stops for Tier 1 and Tier 2 agencies using a second singular year of data, from July 1, 2019 through June 30, 2020. This report analyzes stops for Tier 1 and Tier 2 agencies using two years of data, from July 1, 2019 through June 30, 2021, and stops for Tier 3 agencies using one year of data, from July 1, 2020 through June 30, 2021.

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, which mandated that law enforcement agencies develop written policies related to traffic stop 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 Legislature again considered the collection of police stop data. In SB 415 (2001), the Legislature created the Law Enforcement Contacts Policy & Data Review Committee (LECC), which provided for the voluntary collection of stop data by agencies, and for analysis of collected data by the LECC.

Apart from 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 is the first statewide data collection and analysis program focused on traffic and pedestrian stops in Oregon.

[1] For a full list of agencies see Appendix A, and for reporting rates by agency see Appendix B – Data Audit.

HB 2355, which is codified in ORS 131.930 et seq., created a statewide data collection effort for all officer-initiated traffic[2] and pedestrian[3] 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 per Agency	Data Collection Began	Reporting Began
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

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, the STOP Program 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] 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).
[3] 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).

2. Methodological Approach

2.1. Background

The formal examination of police traffic and pedestrian stop data began in the mid-1990s. Advocacy groups have long 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 were filed (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 and 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 mandated the collection of traffic stop data. In states that had yet to require data collection, many local jurisdictions and departments started collecting and analyzing stop data on their own.

During the approximately three decades that stop data have been studied, the majority of analyses have relied on population-based benchmarks. This approach compares the demographic breakdown of stopped individuals to residential census data. Benchmarks are both intuitive and relatively simple to calculate, but the comparisons that result are overly simplistic and often biased or invalid (see 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[4]. The central thrust of these critiques is that 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 (e.g., commuting patterns and tourism), all of which lead to a disjuncture between the residential demographics and those of the driving population[5].

2.2. Oregon STOP Program Analyses

To address the shortcomings of population-based benchmark analyses, researchers and statisticians have developed several statistical approaches that allow for more precise and less biased estimates of disparities in stop data. The STOP Program relies on 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 law enforcement-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. Race/ethnicity could be a factor in each decision to stop, search, cite, and/or arrest an individual. This distinction is critical, because both the data and analytical techniques required to analyze the various decision points found in a single stop differ. STOP Program researchers address each of these decision points separately.

---

[4] See STOP Program Research Brief: Analytical Approaches to Studying Stops Data (October 2018), which can be found at Traffic_Stop_Research_Memo_Final_Draft-10-16-18.pdf (oregon.gov).

[5] 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.

Second, while the statistical tests utilized by the STOP Program represent the gold standard[6] 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 are estimates and some degree of error could influence results, whether stemming from data collection practices, errors in reporting, or the like. The three analyses utilized by the STOP Program are[7]:

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 is 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). This test matches stop data between two groups based 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, stop day of the week, reason for the stop, season of the year, 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 the factors listed above and further controlling for these factors with regression analysis, minority individuals are either cited, searched, or arrested more often than similarly situated white individuals, then there is 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 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 is evidence of a disparity.

2.3. Analytical Sample

2.3.1. Data Elements

A total of 477,964 records were submitted by 143 Tier 1, Tier 2, and Tier 3 agencies during the third 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, which included representatives from law enforcement, community groups, state agencies, and the Oregon Legislature, to formally define the data elements contained in the statute.

---

[6] 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.
[7] More detailed, technical descriptions of these analyses can be found in Appendices D, E, and F.

Date and Time the Stop Occurred. Law enforcement personnel are required to record the date (month/day/year) and time that the stop occurred. The data is further categorized into day of the week and season. 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 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[8].

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, Latinx, 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 represent over 90 percent of all stops, law enforcement provides 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: Serious moving violations; minor moving violations; equipment, cell phone, and seat belt violations; registration and license violations; and "other" violations (e.g., criminal offenses, camping violations)[9].

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 Year 3 data included only 174 juvenile summons (0.04 percent of all dispositions).

---

[8] For instance, 454 Year 3 stops were labeled as traffic stops, but the citation code was ORS 814.070, which refers to a pedestrian improperly proceeding along a highway. These stops were reclassified by CJC researchers as pedestrian stops.

[9] 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).

The STOP Program created four 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. Additionally, variables representing sunrise time and sunset time were made for use in the Veil of Darkness and Predicted Disposition analyses[10]. 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.

In 2019 and 2021, the STOP program added two additional data categories. First, in July 2019, the STOP Program began collecting data on whether the stopped individual was perceived prior to the police stop. This data point is particularly valuable in the Veil of Darkness analysis which relies on the assumption that the race of the driver will be harder for the officer to perceive in darkness. Data on whether the subject, and their race, was perceived prior to the stop enables analysts to test the Veil of Darkness assumption. Second, beginning in December 2020, law enforcement agencies were able to start submitting additional data to the STOP Program on the reason for the most serious stop disposition. Previously, for example, if an officer stopped someone for a moving violation but the stop ended in arrest because of an outstanding warrant, analysts would only be able to see a moving violation ending in arrest. This additional data point allows the STOP program analysts to more accurately account for the reason for the stop disposition. However, data on whether the subject was perceived prior to the stop and most serious reason for stop disposition is submitted voluntarily by STOP agencies. Thus, not all STOP agencies consistently submit these data so these data are not included in this year’s analyses but may be incorporated in the future.

---

[10] Sunrise time and sunset time were also used for analysis conducted for the 2019 and 2020 STOP reports. They were not explicitly listed in this section previously, however their construction is the same as in the past.

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 and agency. This issue was particularly acute compared to previous stop reports due to the inclusion of Tier 3 agencies. Tier 3 agencies have fewer officers than Tier 1 and 2 agencies, and therefore submit a relatively low number of police stops. For example, eight Tier 3 agencies made fewer than 100 stops in Year 3, and three Tier 3 agencies made fewer than 30 stops. In cases where the sample size is too small, statistical analyses cannot be conducted.

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 sample size thresholds in Table 2.3.1.

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[11]
Predicted Disposition	Model convergence[12]
Hit-Rate	Minimum 30 observations per racial/ethnic group analyzed; no cell with less than 5 observations

The sample size issue identified above had a significant impact on the STOP Program research team's ability to conduct analyses on each of the racial/ethnic groups found in the stop database. Tables 2.3.2.a. and b., and Table C.1. in Appendix C report the breakdown by race/ethnicity and agency for all Tier 1, Tier 2, and Tier 3 agencies for stops occurring from July 1, 2020 through June 30, 2021, the most recent year of data collection. In several cases, even with two years of data for Tier 1 and 2 agencies, 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), many of these counts fell under the requisite thresholds.

Table 2.3.2.a. Race/Ethnicity Reporting for Tier 1 Agencies for All Reported Stops

Agency Name	Asian/PI	Black	Latinx	Middle Eastern	Native American	White
Beaverton PD	659	1,179	2,395	329	69	8,301
Clackamas CO SO	809	1,076	2,509	264	189	16,099
Eugene PD	264	727	893	0	28	11,210
Gresham PD	109	362	449	14	12	1,533
Hillsboro PD	359	355	1,715	170	37	3,834
Marion CO SO	468	420	3,105	171	16	11,464
Medford PD	82	228	1,059	31	5	4,346
Multnomah CO SO	383	1,052	1,287	123	55	6,704
Oregon State Police	3,565	4,220	17,314	1,646	749	100,714
Portland PB	736	2,563	1,587	158	68	9,163
Salem PD	104	204	1,265	28	26	3,440
Washington CO SO	1,089	1,090	4,215	465	142	11,764

[11] 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, larger sample sizes are needed. 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 observations for an individual racial/ethnic group at 100.
[12] All possible racial group and stop outcome models are estimated in Stata. Models that did not converge are not included in the results.

Table 2.3.2.b. Race/Ethnicity Reporting for Tier 2 Agencies for All Reported Stops

Agency Name	Asian/PI	Black	Latinx	Middle Eastern	Native American	White
Albany PD	63	122	532	17	10	4,278
Ashland PD	69	130	197	32	4	2,191
Bend PD	60	77	356	21	11	4,105
Benton CO SO	117	173	400	35	18	4,487
Canby PD	43	37	489	16	9	1,708
Central Point PD	41	60	210	5	1	1,487
Corvallis PD	406	296	632	153	37	6,525
Deschutes CO SO	56	42	187	13	1	2,377
Douglas CO SO	15	28	85	9	0	1,349
Forest Grove PD	100	109	1,158	33	9	3,168
Grants Pass DPS	26	36	154	10	0	1,716
Hermiston PD	29	66	1,768	12	35	2,431
Hood River CO SO	31	9	269	20	0	822
Jackson CO SO	84	97	732	19	4	3,853
Keizer PD	54	74	565	9	0	1,784
Klamath CO SO	20	12	71	2	9	445
Klamath Falls PD	150	111	420	34	5	2,639
Lake Oswego PD	216	198	391	100	46	3,926
Lane CO SO	23	56	110	3	1	1,187
Lebanon PD	3	2	9	0	0	170
Lincoln City PD	31	22	88	9	0	496
Lincoln CO SO	60	31	183	17	10	1,411
Linn CO SO	70	94	329	22	33	4,266
McMinnville PD	19	20	174	7	0	706
Milwaukie PD	89	182	253	39	15	2,019
Newberg-Dundee PD	106	96	552	23	0	3,141
Oregon City PD	136	209	547	59	37	5,118
OHSU PD	10	14	10	5	0	87
Polk CO SO	112	93	581	40	15	2,076
Port of Portland PD	96	187	150	31	4	851
Redmond PD	36	27	248	9	3	1,587
Roseburg PD	36	62	205	28	9	4,353
Springfield PD	59	325	461	1	0	6,850
Tigard PD	199	230	520	120	38	2,709
Tualatin PD	227	279	947	60	8	4,531
UO PD	11	14	7	3	1	256
West Linn PD	147	160	372	106	49	3,179
Woodburn PD	12	13	925	9	0	549
Yamhill CO SO	100	107	750	32	12	3,486
Total Tier 2	**3,162**	**3,900**	**16,037**	**1,163**	**434**	**98,319**

To combat sample size issues, this report includes two years of data in all analysis for Tier 1 and Tier 2 agencies, while only one year of data was available for Tier 3 agencies. In previous reports, models aggregated at STOP Tier level were conducted for all racial/ethnic groups where possible, while models for agency-specific analyses were limited to comparisons between white and Black individuals and white and Latinx individuals. In this third report, no tier level analyses are conducted, however, agency-specific analyses of Asian/PI, Native American, and Middle Eastern individuals are all done when possible.

STOP Program researchers faced similar sample size issues with pedestrian stops. Across all Tier 1, Tier 2, and Tier 3 agencies that submitted data to CJC, only 3.0 percent of stops, which represents 14,141 individual encounters, were pedestrian stops in the third year of data collection. In nearly all instances, models for pedestrian stops could not be estimated on their own. Further, when agency-level pedestrian stops are disaggregated by race/ethnicity, the problem becomes more acute. For instance, only one agency—Eugene Police Department—stopped at least 100 Black pedestrians or Latinx pedestrians in the third year of data collection. No agency reported more than 50 Asian/PI and Middle Eastern pedestrian stops, and only one agency—Pendleton Police Department—reported more than 50 Native American pedestrian stops. Due to these sample size issues, pedestrian and traffic stops were analyzed together in this report for all post-stop outcomes[13].

A final concern is the prevalence of missing data. Resource limitations at some law enforcement agencies with a small number of staff is a challenge for STOP data submission and increases the potential for missing data. These resource and staffing limitations are likely exacerbated by the impacts of the COVID-19 pandemic, with Tier 3 agencies beginning data collection in July 2020 shortly after the pandemic started. 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. Some Tier 3 agencies submitted a partial year of data. 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. Missing data attributable to both of these sources were found.

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 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. This means 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 variation in statistical estimates. In some cases, confidence in the reported results exceeded the 95 percent confidence threshold.

When possible, multiple comparisons were made for each agency test. In situations where multiple tests are employed, all of which may indicate statistical significance, best practices require Bonferroni adjustments to adjust for the likelihood of a given test yielding a false positive result. The Bonferroni adjustment differed for each agency test, contingent on the number of comparisons made. The number of comparisons is detailed in Table 2.4.1. Some agencies had too few stops of Asian/PI, Black, Latinx, Middle Eastern, or Native American individuals to run tests for each group. Therefore, the magnitude of the Bonferroni adjustment may differ by agency, based on the number of tests run for that agency.

Table 2.4.1. Bonferroni Adjustment by Analysis

Analysis	Number of Comparisons per Agency
Veil of Darkness	Up to 5 comparisons
Predicted Disposition	Up to 20 comparisons
Hit-Rate	Up to 5 comparisons

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[14], 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[15]. The justification for this approach mirrors the

[13] 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.
[14] The Connecticut Racial Profiling Prohibition Project is located at http://www.ctrp3.org/.
[15] 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

reasoning behind the utilization of multiple tests to examine the data acquired for this project. As discussed previously, given that the statistical output 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 we 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 majority of states, few states can boast a statewide, statutorily mandated data collection effort like Oregon's. This robust database and the statistical evaluation of stop data can form the foundation of a transparent dialogue between state leaders, government 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 during discretionary stops. 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 police agency 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 law enforcement agency or at a larger level of aggregation. 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.

An additional limitation to the current report is the effect of the ongoing COVID-19 pandemic. In March 2020, many police agencies reduced the number of individuals stopped and some curtailed their discretionary stops entirely. As of June 2021, many agencies continue to have a reduced number of stops compared to pre-pandemic levels. The effects of the COVID-19 pandemic are discussed in more detail in Section 3.2., although it is important to mention that this decrease in stops further exacerbated sample size limitations.

Despite these limitations, the statistical results 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 contribute to the dialogue between law enforcement and Oregonians.

3. Characteristics of Year 3 Stop Data

3.1. General Characteristics

While the analyses contained in Sections 4., 5., and 6. utilize two years of submitted data, this section analyzes data collected by the STOP Program for officer-initiated traffic and pedestrian stops solely for the most recent year, which includes stops made between July 1, 2020 through June 30, 2021. In total,

---

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.

477,964[16] stops were submitted to the STOP Program by 143 Tier 1, Tier 2, and Tier 3 agencies during Year 3. The number of stops reported by each agency is displayed in Tables 3.1.1.a. and b., and Table C.2. in Appendix C. There was significant variation in the frequency with which Tier 1, Tier 2, and Tier 3 agencies stopped individuals. Tier 1 agencies generally made more stops than Tier 2 agencies, which in turn made more stops than Tier 3 agencies, which is consistent with size differences in terms of officers employed. The Oregon State Police, which is the state's largest law enforcement agency, made 131,000 stops in year three, the largest number reported by any one agency and just over a quarter of all stops in the state. At the other end of the continuum, Butte Falls PD made the fewest stops, totaling nine, accounting for less than 0.001 percent of the reported stops in Year 3.

Table 3.1.1.a. Number and Percent of Tier 1 Agency Stops by Stop Type Traffic vs. Pedestrian

Agency Name	Traffic		Pedestrian		Total
Beaverton PD	12,429	96.1%	503	3.9%	12,932
Clackamas CO SO	20,276	96.8%	670	3.2%	20,946
Eugene PD	11,111	84.4%	2,053	15.6%	13,164
Gresham PD	2,477	99.9%	2	0.1%	2,479
Hillsboro PD	6,278	97.0%	192	3.0%	6,470
Marion CO SO	15,593	99.7%	51	0.3%	15,644
Medford PD	5,035	87.5%	718	12.5%	5,753
Multnomah CO SO	9,301	96.8%	303	3.2%	9,604
Oregon State Police	130,276	99.4%	724	0.6%	131,000
Portland PB	14,190	99.4%	85	0.6%	14,275
Salem PD	4,913	96.0%	203	4.0%	5,116
Washington CO SO	18,616	99.2%	149	0.8%	18,765
Tier 1 Total	**250,495**	**97.8%**	**5,653**	**2.2%**	**256,148**

Tables 3.1.1.a. and b. and Table C.2. in Appendix C report the number and percentage of stops by agency broken down by stop type—traffic or pedestrian— and separated by Tier. Stop type has been adjusted as described in Section 2.3.1. By agency and within tier, the frequency with which pedestrian stops were made, as well as the degree to which those stops affected a department's overall stop profile, varied significantly. Across all Tiers, Tier 3 agencies had the highest proportion of pedestrian stops, 5.4 percent, compared to Tier 1's 2.2 percent, and Tier 2's 2.6 percent. This is likely due to the presence of agencies which are small and do not patrol highways or streets. For instance, two Tier 3 agencies, Union Pacific Railroad PD and Portland State University, both reported 100 percent pedestrian stops. Of Tier 1 agencies, Eugene PD and Medford PD made the highest proportion of pedestrian stops, echoing past reports. Of Tier 2 agencies, UO PD had the highest proportion of pedestrian stops; just under a half of all their stops were of pedestrians.

[16] 748, or 0.16% of these 477,964 stops were not definitively identified as either a pedestrian or traffic stop, so stop totals in Table 3.1.1.a. and b., and Table C.2. do not add up exactly to 477,964.

Table 3.1.1.b. Number and Percent of Tier 2 Agency Stops by Stop Type Traffic vs. Pedestrian

Agency Name	Traffic		Pedestrian		Total
Albany PD	4,860	96.8%	162	3.2%	5,022
Ashland PD	2,380	90.7%	243	9.3%	2,623
Bend PD	4,608	99.5%	23	0.5%	4,631
Benton CO SO	5,208	99.6%	22	0.4%	5,230
Canby PD	2,290	99.5%	12	0.5%	2,302
Central Point PD	1,739	96.3%	67	3.7%	1,806
Corvallis PD	7,998	99.4%	51	0.6%	8,049
Deschutes CO SO	2,675	100.0%	1	0.0%	2,676
Douglas CO SO	1,428	94.9%	76	5.1%	1,504
Forest Grove PD	4,486	98.0%	91	2.0%	4,577
Grants Pass DPS	1,810	93.0%	137	7.0%	1,947
Hermiston PD	4,224	97.2%	121	2.8%	4,345
Hood River CO SO	1,151	100.0%	0	0.0%	1,151
Jackson CO SO	4,774	99.7%	15	0.3%	4,789
Keizer PD	2,485	100.0%	1	0.0%	2,486
Klamath CO SO	536	95.7%	24	4.3%	560
Klamath Falls PD	3,349	99.7%	10	0.3%	3,359
Lake Oswego PD	4,840	99.2%	37	0.8%	4,877
Lane CO SO	1,382	99.5%	7	0.5%	1,389
Lebanon PD	185	100.0%	0	0.0%	185
Lincoln City PD	601	93.0%	45	7.0%	646
Lincoln CO SO	1,705	99.6%	7	0.4%	1,712
Linn CO SO	4,736	98.4%	78	1.6%	4,814
McMinnville PD	925	99.9%	1	0.1%	926
Milwaukie PD	2,452	94.4%	145	5.6%	2,597
Newberg-Dundee PD	3,868	98.7%	50	1.3%	3,918
Oregon City PD	5,698	93.3%	408	6.7%	6,106
OHSU PD	126	98.4%	2	1.6%	128
Polk CO SO	2,903	99.5%	14	0.5%	2,917
Port of Portland PD	1,156	87.4%	167	12.6%	1,323
Redmond PD	1,825	95.5%	85	4.5%	1,910
Roseburg PD	4,516	96.2%	177	3.8%	4,693
Springfield PD	7,285	93.9%	473	6.1%	7,758
Tigard PD	3,623	94.9%	193	5.1%	3,816
Tualatin PD	6,006	99.2%	46	0.8%	6,052
UO PD	157	53.8%	135	46.2%	292
West Linn PD	3,977	99.1%	36	0.9%	4,013
Woodburn PD	1,505	99.8%	3	0.2%	1,508
Yamhill CO SO	4,456	99.2%	34	0.8%	4,490
Total Tier 2	**119,928**	**97.4%**	**3,199**	**2.6%**	**123,127**

The demographic breakdowns for traffic and pedestrian stops are reported in Table 3.1.2. For all agencies contained in this report, the majority of stops were of white drivers/pedestrians, with Latinx 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 a higher proportion of pedestrians across all tiers. With regard to gender, more males were stopped than females. This gender difference is more pronounced in pedestrian stops. Most traffic and pedestrian stops are of individuals perceived to be aged in their thirties, slightly more so for pedestrians, across all tiers. This echoes previous years' data. Tier 3 agencies stopped a higher proportion of older individuals than other tiers.

Table 3.1.2. Aggregate Demographics by Tier and Stop Type

	Tier 1			Tier 2			Tier 3		
	Traffic	**Ped.**	**Total**	**Traffic**	**Ped.**	**Total**	**Traffic**	**Ped.**	**Total**
Race/Ethnicity									
Asian/PI	3.5%	1.4%	3.4%	2.6%	1.3%	2.6%	1.9%	1.4%	1.9%
Black	5.3%	5.0%	5.3%	3.1%	4.5%	3.2%	2.0%	1.6%	2.0%
Latinx	15.1%	8.3%	14.9%	13.2%	7.7%	13.0%	13.5%	5.4%	13.1%
Middle Eastern	1.4%	0.3%	1.3%	1.0%	0.4%	1.0%	0.7%	0.3%	0.7%
Native American	0.6%	0.6%	0.6%	0.3%	0.7%	0.4%	0.6%	2.2%	0.7%
White	74.2%	84.4%	74.5%	79.8%	85.4%	79.9%	81.3%	89.2%	81.8%
Gender									
Male	67.1%	83.4%	67.4%	64.0%	79.4%	64.4%	64.8%	66.0%	64.9%
Female	32.8%	16.4%	32.4%	35.8%	20.3%	35.4%	34.2%	31.9%	34.0%
Nonbinary	0.1%	0.2%	0.1%	0.1%	0.3%	0.1%	1.1%	2.1%	1.1%
Age									
Under 21	11.3%	7.5%	11.2%	12.9%	8.9%	12.8%	13.1%	11.0%	13.0%
21-29	25.4%	20.0%	25.3%	23.2%	19.1%	23.1%	23.1%	20.9%	23.0%
30-39	25.2%	32.0%	25.3%	24.7%	31.0%	24.9%	23.1%	24.4%	23.2%
40-49	16.4%	21.4%	16.5%	16.7%	19.5%	16.8%	16.1%	17.0%	16.2%
50 and Older	21.7%	19.2%	21.7%	22.5%	21.6%	22.5%	24.6%	26.8%	24.7%

Figure 3.1.1. Stops by Month of Year

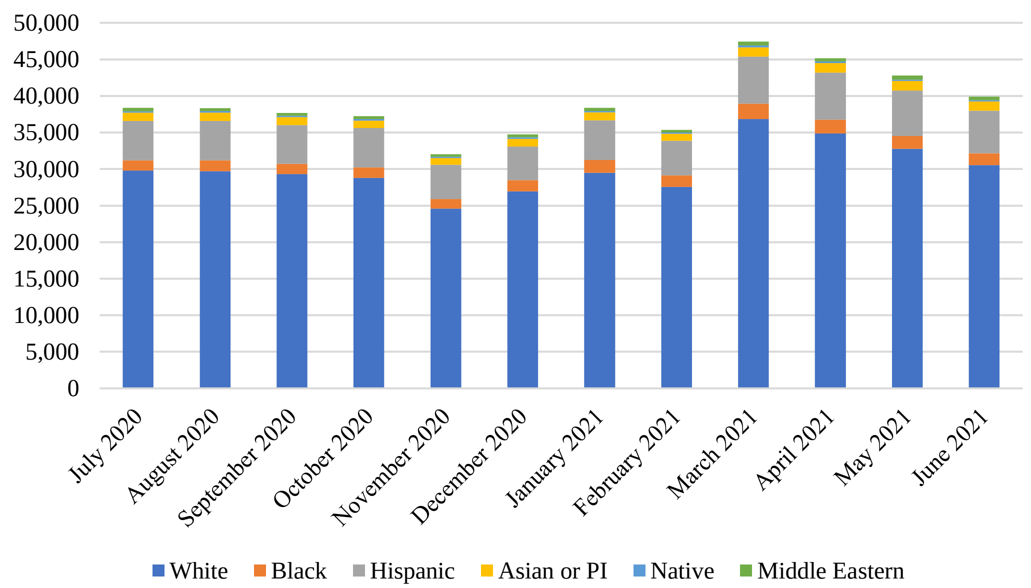


Figure 3.1.1. depicts traffic and pedestrian stops broken out by time of day. Fewer stops occurred from three am through five am. Demographic trends across time of day are generally consistent except during the very early commuting hours, 4-6 am, when Latinx stops make up between 19 and 20 percent of stops. During the rest of the day, Latinx stops make up, on average, 14 percent of stops.

Table 3.1.3. displays the most serious dispositions reported by law enforcement. Most police stops did not result in further action taken against the stopped individual. The most common outcome of a stop

regardless of type or Tier was a warning[17]. Three-quarters of stops by Tier 3 agencies end in no action or a warning, which is a higher proportion than Tier 1 (60 percent) or Tier 2 (69 percent) agencies. Juvenile summons remains a rare outcome as in past reports. Tier 1 agency stops end in arrest more often than Tier 2 or Tier 3 stops. Similar to the Year 2 report, pedestrian stops were more likely to end in an arrest and traffic stops were more likely to end in a citation, regardless of tier.

Table 3.1.3. Stop Disposition by Stop Type and Tier

Disposition	Tier 1			Tier 2			Tier 3		
	Traffic	Ped.	Total	Traffic	Ped.	Total	Traffic	Ped.	Total
None	2.8%	15.3%	3.1%	8.4%	16.8%	8.6%	6.8%	10.8%	7.0%
Warning	57.2%	57.0%	57.2%	60.1%	55.1%	59.9%	68.5%	76.0%	68.9%
Citation	37.6%	13.3%	37.1%	29.3%	12.5%	28.8%	23.0%	6.9%	22.1%
Juv. Summons	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.2%	0.1%
Arrest	2.3%	14.4%	2.6%	2.3%	15.5%	2.6%	1.7%	6.1%	1.9%

Table 3.1.4. provides Year 2 search data, stratified by Tier. Tier 1 agencies conduct searches in 2.5 percent of stops, a higher percentage than Tier 2 and Tier 3. Pedestrians were searched more often than drivers, but searches were less successful. For Tier 1 and Tier 2 agencies, about half of all searches were consent searches. For Tier 3 agencies, consent searches made up less, just under a third of all searches. Echoing previous STOP reports, drugs were the most common form of contraband found in searches, followed by alcohol. Tier 3 agencies found alcohol more often (14.1 percent) during a search than Tier 2 (6.5 percent) or Tier 3 agencies (12.9 percent).

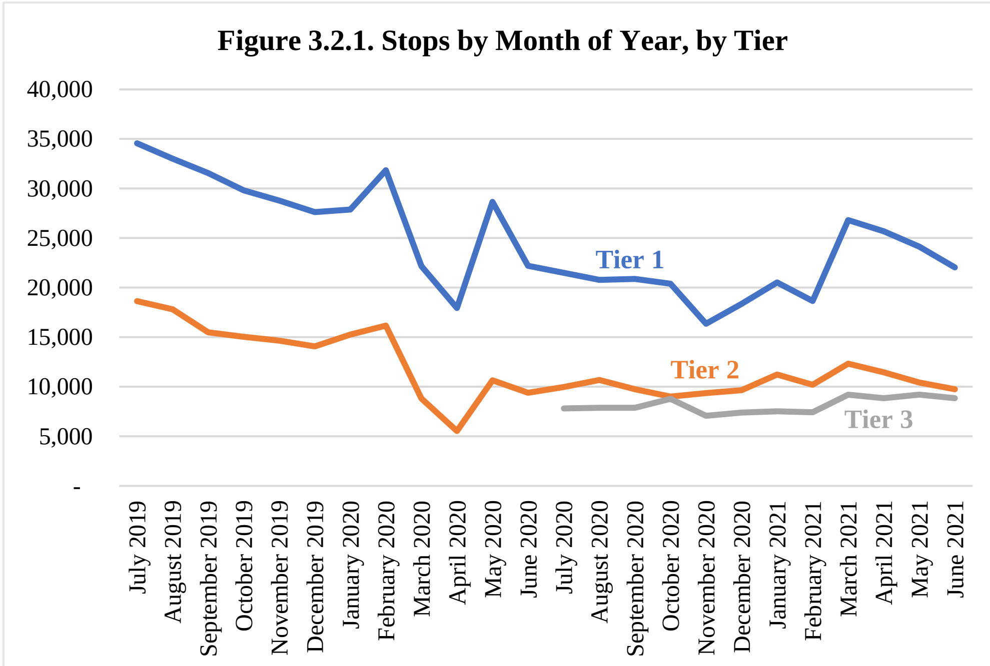
Table 3.1.4. Search Results by Stop Type and Tier

Variable	Tier 1			Tier 2			Tier 3		
	Traf.	Ped.	Total	Traf.	Ped.	Total	Traf.	Ped.	Total
Search Conducted	2.2%	13.3%	2.5%	1.7%	10.0%	1.9%	1.1%	5.5%	1.4%
Reason									
Consent Search	43.5%	44.7%	43.7%	43.1%	57.1%	45.0%	34.2%	19.3%	31.0%
“Other” Search	55.3%	56.5%	56.4%	56.9%	43.0%	55.0%	65.9%	80.7%	69.0%
Percent Successful	55.3%	47.8%	54.2%	40.8%	29.5%	39.2%	52.2%	32.8%	48.0%
Item Seized									
Alcohol Found	13.9%	5.8%	12.9%	7.2%	1.9%	6.5%	16.3%	6.2%	14.1%
Drugs Found	33.2%	28.0%	32.5%	26.8%	19.8%	25.8%	29.8%	21.0%	27.9%
Weapons Found	9.2%	8.0%	9.0%	8.5%	6.0%	8.2%	8.0%	6.6%	7.7%
Stolen Property Found	3.2%	5.0%	3.5%	2.0%	1.9%	2.0%	2.6%	1.4%	2.3%
Other Evidence Found	10.2%	13.0%	10.6%	6.1%	3.5%	5.7%	8.2%	1.4%	6.8%
Other Non-Evidence Found	2.6%	6.0%	3.0%	6.1%	5.0%	5.9%	4.5%	3.8%	4.3%

[17] It is the policy of many agencies to give a warning to everyone who is stopped.

3.2. COVID-19 Data Trends

Figure 3.2.1. displays stops made by Oregon law enforcement agencies from July 2019 through June 2021, stratified by Tier. From July 2019 through June 2020, only Tier 1 and Tier 2 agencies reported stops. In July 2020, Tier 3 agencies started reporting. From February to March 2020, when COVID-19 mitigation efforts were first put in place, Tier 1 stop volume dropped 30 percent and Tier 2 stop volume dropped by a greater percentage, 45 percent. Overall stop volume dropped a further 24 percent in April 2020, before rebounding in May 2020 to 82 percent of the stop volume pre-pandemic.



When Tier 3 agencies began reporting, COVID-19 changes continued to correspond with changing stopping patterns. On November 18, 2020, Governor Brown implemented a two-week freeze statewide which limited restaurants and bars to take-out service only, closed indoor gyms, museums, and theaters, and included other closures[18]. Leading up to these closures, Tier 2 stop volume dropped 8 percent month-over-month in both September and October. Tier 2 agency overall stop volume did not drop further in November 2020 (Figure 3.2.1.). However, Tier 1 and Tier 3 agencies, which had experienced smaller drops or even gains in the two preceding months, each experienced a 19 percent decrease in stop volume in November 2020 compared to the preceding month. In March 2021, COVID-19 vaccines became more widely available to Oregonians. Corresponding to increased vaccine access, all tiers experienced gains in stop volume in March 2021 that generally persist through June 2021. However, for Tier 1 and Tier 2 agencies, which reported stop data before the pandemic, overall stop volume has not returned to pre-pandemic levels. This may be due to ongoing impacts of the COVID-19 pandemic including law enforcement staffing shortages and/or potentially permanent commuting changes. Next year's stop report will evaluate the enduring effect of COVID-19 as employment structures and their associated commuting patterns potentially become more permanent.

[18] State of Oregon: Administration - Executive Order 20-65

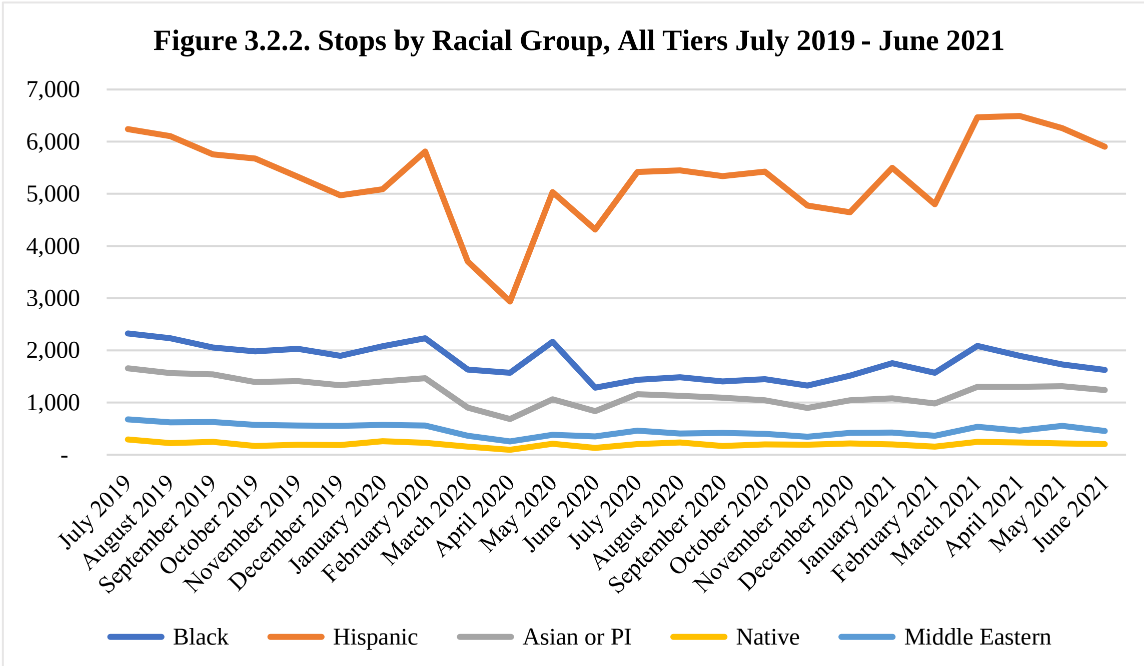
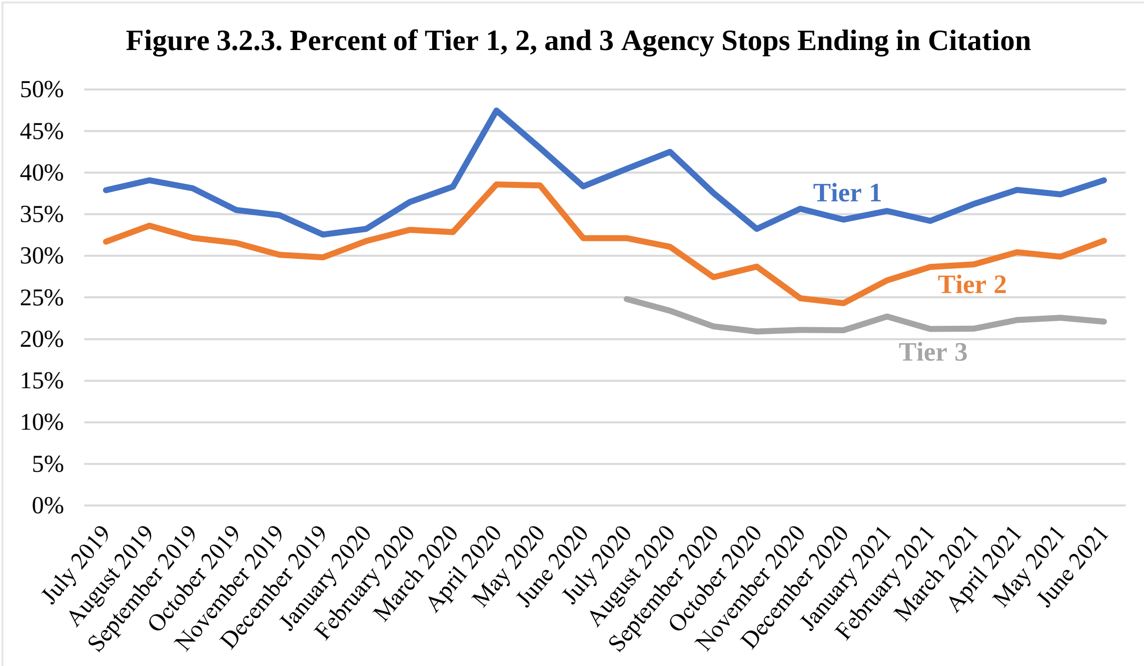


Figure 3.2.2. shows all reported police stops by racial/ethnic group—excluding stops of white individuals—from July 2019 through June 2021[19]. In the December 2020 report, STOP researchers noted that although stops dropped for all racial groups in March and April 2020, stops of Black individuals did not fall as much as other racial/ethnic groups, potentially because white workers were more likely to work from home[20]. This result is partially echoed in later months as the pandemic continued to influence stop volumes. In November 2020, when overall stop volume dropped 14 percent from October, stops of white individuals dropped 15 percent, Latinx individuals dropped 12 percent, Asian/PI individuals dropped 17 percent, Middle Eastern individuals dropped 14 percent, and stops of Black and Native individuals dropped only 9 and 6 percent, respectively. When stop volume increased in March 2021, increases in stops were slightly more uniform across racial/ethnic groups. From February to March 2021, for instance, stops of white individuals increased 34 percent, Black stops increased 33 percent, Latinx stops increased 36 percent, Asian/PI stops increased 25 percent, Middle Eastern stops increased 46 percent, and Native American stops increased the most, 65 percent.

Despite the fluctuations in stop volume around changes in COVID-19 rates and mitigation efforts, arrest and search rates were generally stable. Citation rates however, changed for Tier 1 and Tier 2 agencies as stop volume increased or decreased. Figure 3.2.3. displays citation rates from July 2019 through June 2021 by Tier. Overall stop volume decreased 36 percent in March 2020, and then a further 24 percent in April 2020. For Tier 1 agencies, the percent of stops ending in citation increased from 36 percent in February to 38 percent in March, and further up to 47 percent in April 2020. The proportion of stops ending in citation increased similarly for Tier 2 agencies over the same period. However, when stop volume dropped in November 2020, citation rates remained relatively flat across all tiers. When COVID-19 mitigation efforts were most aggressive in March and April 2020, citation rates increased and stop volume decreased sharply. However, that effect appears to be isolated to those two months in the existing data.

---

[19] White stops make up 78 percent of monthly stops, on average, and largely echo the monthly variation shown in Figure 3.2.1.
[20] From June 2020 article, *Ability to Work From Home: Evidence From Two Surveys and Implications for the Labor Market in the COVID-19 Pandemic*.https://www.bls.gov/opub/mlr/2020/article/ability-to-work-from-home.htm



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[21], 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. The 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, 2021 when it was dark at 5:00pm to stops made two months later at the same time on March 10, 2021, 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

---

[21] *See* Barone et al. (2018).

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 agency level.

4.1. Agency-Level Veil of Darkness Analysis

The following analyses utilized two years of data for Tier 1 and Tier 2 agencies. At the agency level, therefore, it is possible to estimate VOD models for many of the non-white groups reported in the stop database given a sufficient sample size for the first time. First, Table 4.1.1. displays the odds ratios for the Tier 1 and Tier 2 VOD models with at least two comparisons for all non-white stopped drivers, including those perceived as Black, Latinx, Asian/PI, Middle Eastern, and Native American, compared to white stopped drivers. As described in Section 2., the sample size requirement for the VOD model was at least 100 stops in each racial/ethnic group within the inter-twilight windows for the two years of data provided. For the full Tier 1 and Tier 2 models, most comparisons show no statistically significant differences in the odds of minority stops in daylight compared to darkness. For Milwaukie PD, however, the odds of stops for Black drivers in daylight was nearly 2.7 times the odds for white drivers, indicating a statistically significant difference evidencing a disparity in the rate of stopped drivers in daylight compared to darkness[22].

[22] The odds ratio for Gresham PD for Asian/PI drivers (2.46) shows a p-value of 0.03. With the Bonferroni adjustment with three comparisons this is not significant, however without the adjustment the p-value is below the 0.05 threshold. Gresham PD is not identified in the Predicted Disposition or KPT Hit-Rate test, and therefore would not be referred to DPSST if a significant result was found for the Veil of Darkness test.

Table 4.1.1. Logistic Regression of Minority Status on Daylight by Tier 1 or 2 Agency

Agency	Tier	Asian/PI	Black	Latinx	Middle Eastern	Native American
Beaverton PD	1	0.70	1.10	0.91	0.61	--
Clackamas CO SO	1	0.88	1.01	1.10	1.05	0.59
Eugene PD	1	1.10	1.09	1.02	--	--
Gresham PD	1	2.46	1.19	1.01	--	--
Hillsboro PD	1	0.82	0.81	1.14	1.07	--
Marion CO SO	1	0.93	1.73	1.01	--	--
Medford PD	1	--	--	0.95	--	--
Multnomah CO SO	1	1.50	0.98	1.30	--	--
Oregon State Police	1	0.94	1.05	1.01	1.02	1.03
Portland PB	1	0.80	0.94	1.10	0.96	--
Salem PD	1	--	0.98	1.05	--	--
Washington CO SO	1	0.86	0.70	0.94	0.51	1.14
Corvallis PD	2	0.81	1.35	1.21	--	--
Lake Oswego PD	2	1.69	0.90	1.14	--	--
Milwaukie PD	2	--	2.68*	1.58	--	--
Port of Portland PD	2	--	1.34	0.91	--	--
Springfield PD	2	--	0.74	0.52	--	--
Tigard PD	2	0.62	1.26	1.17	--	--
Tualatin PD	2	1.46	0.66	0.72	--	--
West Linn PD	2	1.06	1.52	0.97	--	--

Notes: * p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction based on number of comparisons).
Logistic regression results include controls for age, gender, reason for stop, day of week, time of day, quarter or season, year, county stop volume, and agency stop volume.

Table 4.1.2. reports the Tier 2 agency specific model results for Latinx drivers compared to white drivers for agencies not displayed above. While a number of agencies have odds ratios above 1.0, most agencies show no statistically significant difference in the rate of stopped Latinx drivers in daylight compared to darkness. For Ashland PD, however, the odds of stops for Latinx drivers in daylight was 2.3 times the odds for white drivers, indicating a statistically significant difference evidencing a disparity in the rate of stopped drivers in daylight compared to darkness.

Table 4.1.2. Logistic Regression of Latinx Drivers on Daylight by Tier 2 Agency

Agency	Latinx	Agency	Latinx
Albany PD	1.33	Klamath Falls PD	0.99
Ashland PD	2.33*	Lane CO SO	1.01
Bend PD	1.00	Lincoln City PD	0.76
Benton CO SO	1.16	Lincoln CO SO	0.80
Canby PD	0.94	Linn CO SO	0.71
Central Point PD	1.41	McMinnville PD	0.73
Corvallis PD	1.21	Newberg-Dundee PD	0.74
Deschutes CO SO	1.22	Oregon City PD	0.97
Forest Grove PD	1.18	Polk CO SO	1.05
Grants Pass DPS	0.85	Redmond PD	1.24
Hermiston PD	0.89	Roseburg PD	1.40
Hood River CO SO	0.59	Woodburn PD	1.00
Jackson CO SO	1.25	Yamhill CO SO	1.11
Keizer PD	1.54		

Notes: * p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction based on number of comparisons).
Logistic regression results include controls for age, gender, reason for stop, day of week, time of day, quarter or season, year, county stop volume, and agency stop volume.

Table 4.1.3. Logistic Regression of Minority Status on Daylight by Tier 3 Agency

Agency	Latinx
Crook CO SO	0.57
Hood River PD	0.76
Hubbard PD	0.74
Josephine CO SO	0.78
Milton-Freewater PD	0.76
Morrow CO SO	0.67
Newport PD	1.06
Pendleton PD	0.48
Prineville PD	0.95
Seaside PD	0.82
Umatilla CO SO	1.34
Umatilla PD	1.17

Notes: * p<0.05, ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction based on number of comparisons). Logistic regression results include controls for age, gender, reason for stop, day of week, time of day, quarter or season, county stop volume and agency stop volume.

Table 4.1.3. reports the Tier 3 agency specific model results for Latinx drivers compared to white drivers for agencies with sufficient sample size. As described in Section 2., the Tier 3 agency analyses include only one year of data from July 2020 to June 2021. No agency had an odds ratio that was above 1.0 and statistically significant. While some Tier 3 agencies show a higher odds ratio, the estimate is not statistically significantly different from 1.0 and does not indicate a disparity at this time.

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. The Predicted Disposition analysis focuses on the outcomes of stops, including whether stopped individuals were cited, searched, and/or arrested during their encounter with law enforcement.

HB 2355 required all law enforcement agencies to collect data regarding the disposition of stops. Because 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[23]. 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.

5.1. Description of Predicted Disposition Analysis

Variation in enforcement outcomes could be due to time of day, day of the week, the offense that led to the stop, or one of many other factors. 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 at that time. Variation could also be attributed to other factors, including age, gender, or season. Propensity score analysis is employed here to account for as many of these differences as possible and isolate the effect, if any, that race of the stopped individual has on the disposition of the stop.

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 other factors constant, do we find different dispositional outcomes 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 except for the characteristic in question. 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).

[23] See Appendix E for more details on how the STOP research team determines the most serious disposition and the appropriate comparison outcomes for each type of disposition.

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 of these methods to employ 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[24].

Table 5.1.1. Analyses Completed for Each Agency

Disposition of Interest	Comparison Dispositions	Analysis Groups				
Citation	None or Warning	Asian/PI	Black	Latinx	Mid. Eastern	Native
Search	None, Warning, or Citation	Asian/PI	Black	Latinx	Mid. Eastern	Native
Arrest	None, Warning, Citation, or Search	Asian/PI	Black	Latinx	Mid. Eastern	Native
Citation, Search, or Arrest	None or Warning	Asian/PI	Black	Latinx	Mid. Eastern	Native

The current analysis included twenty sub-analyses for each agency: each outcome of citation, search, arrest, or any non-warning disposition across each racial/ethnic group of Asian/PI, Black, Latinx, Middle Eastern, and Native American individuals (Table 5.2.1.). The comparison group was drawn from the group of white stops for the agency in question. Each row of Table 5.1.1. describes the tests conducted for each agency. In row 1, STOP Program researchers tested whether there was a disparity in issuing citations between each of the racial groups shown in the analysis groups column and a matched white group.[25] Row 2 does the same for searches, row 3 for arrests, and row 4 describes tests for any Citation, Search, or Arrest disposition.

5.2. Predicted Disposition Results

As with the Veil of Darkness analysis in the previous section, the analyses conducted in this section include two years of data for all Tier 1 and Tier 2 agencies. Table 5.2.1. reports agency-level results for all agencies where a statistically significant disparity was found. Nine Tier 1 law enforcement agencies report statistically significant disparities for the Predicted Disposition analysis. For three Tier 1 agencies, Clackamas CO SO, Eugene PD, and Hillsboro PD, disparities were detected only 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 six Tier 1 agencies, Beaverton PD, Marion CO SO, Oregon State Police, Portland PB, Salem PD, and Washington CO SO, disparities were reported for either searches and/or arrests, sometimes in addition to citations.

Where disparities were found, the average gap in the predicted versus the actual disposition rate varied by agency and type of disposition. These differences may be especially apparent between large and small agencies. Larger agencies make more stops and thus have a greater sample size, which leads to more precise statistical tests and a lower threshold for identifying statistically important differences.

[24] Inverse Probability Weighted Regression Adjustment 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 E.
[25] Each matched white group will differ from the next, since the characteristics of the stops of the group being matched differ.

Table 5.2.1. Predicted Disparity by Agency and Disposition (only statistically significant results displayed)

Agency	Race/Ethnicity	Citation		Search*		Arrest		Citation, Search, or Arrest	
		Actual	Pred.	Actual	Pred.	Actual	Pred.	Actual	Pred.
Albany PD	Latinx	47.8%	41.6%	--	--	--	--	41.6%	37.1%
	Native	--	--	--	--	--	--	64.5%	37.2%
Beaverton PD	Asian/PI	44.4%	40.2%	--	--	--	--	--	--
	Latinx	--	--	6.1%	4.6%	--	--	43.6%	41.2%
Brookings PD	Latinx	22.3%	12.1%	--	--	--	--	22.3%	12.2%
Canby PD	Latinx	22.8%	17.8%	--	--	--	--	26.0%	21.1%
Clackamas CO SO	Black	32.0%	28.2%	--	--	--	--	34.9%	31.1%
	Latinx	33.3%	30.3%	--	--	--	--	35.5%	32.9%
Eugene PD	Latinx	40.1%	34.0%	--	--	--	--	42.1%	37.3%
Forest Grove PD	Latinx	38.0%	29.2%	--	--	--	--	40.0%	31.1%
Gervais PD	Latinx	87.8%	72.4%	--	--	--	--	87.8%	72.6%
Gilliam CO SO	Latinx	70.5%	60.9%	--	--	--	--	71.5%	60.9%
Hermiston PD	Latinx	32.8%	26.5%	--	--	--	--	34.4%	27.8%
Hillsboro PD	Latinx	35.3%	27.2%	--	--	--	--	37.5%	29.2%
Hubbard PD	Latinx	23.1%	17.0%	--	--	--	--	25.9%	20.0%
Klamath Falls PD	Asian/PI	--	--	8.9%	3.6%	7.6%	2.2%	--	--
Lincoln CO SO	Native	63.0%	24.1%	--	--	--	--	63.0%	25.3%
Linn CO SO	Asian/PI	56.8%	43.2%	--	--	--	--	--	--
	Latinx	49.8%	41.1%	--	--	--	--	44.1%	37.4%
Malheur CO SO	Latinx	46.3%	29.3%	--	--	--	--	46.3%	29.3%
Marion CO SO	Asian/PI	84.9%	81.0%	--	--	--	--	85.2%	81.2%
	Latinx	80.1%	76.6%	3.6%	2.3%	3.3%	2.1%	80.8%	77.2%
	Mideast	--	--	--	--	--	--	88.4%	83.5%
McMinnville PD	Latinx	33.3%	23.8%	--	--	--	--	34.3%	25.2%
Newberg-Dundee PD	Latinx	30.3%	23.6%	--	--	--	--	31.1%	24.2%
Newport PD	Latinx	33.9%	22.9%	--	--	--	--	36.6%	24.3%
Oregon State Police	Asian/PI	40.2%	37.0%	--	--	--	--	40.9%	38.2%
	Black	43.9%	38.1%	--	--	--	--	45.3%	39.6%
	Latinx	45.5%	37.7%	2.4%	1.9%	--	--	47.2%	39.1%
	Mideast	39.2%	35.9%	--	--	--	--	39.9%	37.2%
	Native	42.4%	37.3%	5.0%	2.7%	4.4%	2.7%	45.6%	39.4%
Pendleton PD	Latinx	24.6%	15.9%	--	--	--	--	31.9%	23.2%
	Native	36.1%	18.9%	24.3%	17.1%	25.7%	17.0%	52.9%	33.2%
Polk CO SO	Latinx	28.2%	21.2%	--	--	--	--	31.0%	23.9%
Port of Portland PD	Black	20.4%	13.9%	--	--	--	--	23.1%	16.3%
Portland PB	Black	--	--	6.8%	3.9%	5.9%	4.4%	--	--
Redmond PD	Latinx	36.0%	28.8%	--	--	--	--	--	--
Roseburg PD	Black	66.7%	58.2%	--	--	--	--	--	--
Salem PD	Latinx	62.6%	59.1%	9.0%	7.2%	7.2%	5.0%	65.1%	61.0%
Tigard PD	Latinx	37.5%	26.3%	--	--	--	--	39.0%	28.4%
	Mideast	32.5%	23.3%	--	--	--	--	--	--
	Native	68.5%	36.1%	--	--	--	--	68.9%	37.4%
Tillamook CO SO	Asian/PI	63.4%	42.7%	--	--	--	--	--	--
	Latinx	54.8%	42.9%	--	--	--	--	55.1%	43.7%
Tualatin PD	Latinx	49.1%	43.7%	--	--	--	--	50.9%	45.5%
Umatilla PD	Latinx	25.4%	21.4%	--	--	--	--	26.8%	22.3%
Washington CO SO	Latinx	25.9%	21.9%	3.4%	2.4%	4.0%	3.2%	29.4%	24.7%
West Linn PD	Latinx	25.5%	19.5%	--	--	--	--	26.0%	19.9%
Woodburn PD	Latinx	43.5%	34.6%	4.0%	2.0%	--	--	45.9%	36.2%
Yamhill CO SO	Latinx	28.8%	23.7%	--	--	--	--	29.8%	24.8%

*The search rates presented were corrected in the revised report released April 2022

As described in Section 3, Tier 2 agencies have far fewer stops than Tier 1 agencies. Combined with the already relatively low minority populations in the state, and especially outside of major metro areas, many of the predicted disposition analyses for the Tier 2 agencies did not have sufficient sample sizes to complete the analysis. That said, of the analyses that were completed, Albany PD, Canby PD, Forest Grove PD, Hermiston PD, Lincoln CO SO, Linn CO SO, McMinnville PD, Newberg-Dundee PD, Polk CO SO, Port of Portland PD, Redmond PD, Roseburg PD, Tigard PD, Tualatin PD, West Linn PD, Woodburn PD, and Yamhill CO SO had statistically significant disparities indicated for one or more of the analysis groups for citations and/or any disposition. Significant disparities in searches were found for Klamath Falls PD for Asian/PI individuals and Woodburn PD for Latinx individuals. In addition, the findings indicate that stops of Asian/PI individuals for the Klamath Falls PD were significantly more likely to result in arrests than for white individuals.

Sample size issues were even more pronounced for Tier 3 agencies. However, the following nine Tier 3 agencies were identified as having significant disparities in citations and any disposition for one of the analysis groups: Brookings PD, Gervais PD, Gilliam CO SO, Hubbard PD, Malheur CO SO, Newport PD, Pendleton PD, Tillamook CO SO, and Umatilla PD. In addition, Pendleton PD also had statistically significant disparities in searches and arrests of Native American individuals as compared to white individuals.

As indicated elsewhere in this report, limited sample sizes were a significant barrier to estimation for many of the Tier 2 and 3 agencies. In such cases, a lack of an indicated disparity should not be interpreted as proof-positive that there is no disparity for these groups in these jurisdictions. STOP analysts in these instances were unable to estimate the models with current data limitations. In future iterations of this report Tier 3 agencies will include more than one year of stop data in the analysis, similar to Tier 1 and Tier 2 agencies for the current report.

6. KPT Hit-Rate Analysis

The second analysis conducted 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, Becker 1993). In the past few decades, this approach to 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. Agency level search data were analyzed for disparities between the white baseline group and individuals identified as Black, Latinx, Asian/PI, Middle Eastern, and Native American. For certain agencies and racial/ethnic groups, the Hit-Rate analysis was unable to be 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. This protects against statistical anomalies due to low search counts, and aligns with best practices.[26] Finally, chi-square tests of independence with a Bonferroni adjustment

[26] Connecticut Racial Profiling Prohibition Project (2019).

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. For individual agencies, this includes minority group hit-rates less than the white hit-rate and statistically significant at the 95 percent confidence level. See Appendix F for more detailed technical information about the KPT Hit-Rate model and statistical tests.

6.1. Agency-level KPT Hit-Rate Results

As in the previous two sections, analyses in this section utilized two years of data for all Tier 1 and Tier 2 agencies. In this report, the KPT Hit-Rate analysis was performed for each agency for up to 5 racial/ethnic groups (Black, Latinx, Asian/PI, Middle Eastern, and/or Native) depending upon sample size. Results for these analyses are presented in Table 6.1.1. below.

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

Agency	Race/Ethnicity	Minority Hit-Rate	White Hit-Rate	Significance?
Albany PD	Latinx	42.2%	37.5%	
Beaverton PD	Black	53.3%	61.6%	
	Latinx	63.7%	61.6%	
	Asian/PI	57.1%	61.6%	
Canby PD	Latinx	68.4%	73.6%	
Clackamas CO SO	Black	49.1%	48.7%	
	Latinx	38.8%	48.7%	
Eugene PD	Black	40.0%	43.6%	
	Latinx	41.9%	43.6%	
Forest Grove PD	Latinx	41.9%	29.8%	
Gresham PD	Black	38.7%	50.0%	
	Latinx	50.8%	50.0%	
Hermiston PD	Latinx	56.3%	50.0%	
Hillsboro PD	Latinx	40.3%	43.9%	
Klamath Falls PD	Latinx	38.7%	40.8%	
Marion CO SO	Latinx	14.4%	18.4%	
Medford PD	Black	47.1%	35.4%	
	Latinx	44.3%	35.4%	
Multnomah CO SO	Black	53.9%	55.5%	
	Latinx	46.4%	55.5%	
Oregon State Police	Black	75.0%	65.4%	
	Latinx	66.5%	65.4%	
	Asian/PI	73.0%	65.4%	
	Middle Eastern	37.5%	65.4%	**
	Native	66.3%	65.4%	
Pendleton PD	Latinx	51.6%	32.2%	
	Native	41.5%	32.2%	
Polk CO SO	Latinx	51.9%	62.5%	
Portland PB	Black	37.9%	42.2%	
	Latinx	39.6%	42.2%	
	Asian/PI	47.3%	42.2%	
Salem PD	Black	35.3%	48.0%	
	Latinx	43.0%	48.0%	
Springfield PD	Black	37.9%	41.6%	
	Latinx	37.8%	41.6%	
Washington CO SO	Black	61.4%	62.5%	
	Latinx	71.6%	62.5%	

Notes: * p<0.05. ** p<0.01, *** p<0.001 (Statistical Significance includes a Bonferroni Correction by agency with the number of comparisons shown).

As shown in Table 6.1.1., all agencies have differences in search success rates between white individuals and the comparison groups. These differences in nearly all cases were relatively small, and in all but one case the differences reported were not statistically significant.[27] The lack of statistical significance could be attributed to the relatively small sample sizes found across agencies (particularly for Tier 2 and 3 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 Oregon State Police in their Hit-Rate analysis for the Middle Eastern group. The Hit-Rate analyses for Oregon State Police for other groups (Black, Latinx, Asian/PI, and Native) were not significant. More specifically, for the white-Middle Eastern hit-rate comparison, the percentage of successful searches for white individuals was 65.4 percent, while the percentage of successful searches for Middle Eastern individuals was only 37.5 percent. This difference is significant at the 99% confidence level, indicating a disparity.

7. Findings from 2021 Analysis

7.1. Aggregate Findings

Similar to the data reported in the first and second annual STOP Report, in all, the STOP data demonstrates that the vast majority of discretionary police-citizen interactions in Oregon are traffic stops. The breakdown between traffic and pedestrian stops does vary by both agency as well as tier, however, as some law enforcement agencies engage in more pedestrian stops than others and Tier 3 agencies, on average, logged more pedestrian stops proportionally than Tier 1 and 2 agencies.

With regard to the demographic characteristics of stopped individuals, the aggregate data continue to 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.

Law enforcement agencies reported that stopped individuals either were subject to no further action or merely given a warning in over 60 percent of stops for Tier 1, just under 70 percent of stops for Tier 2, and three-quarters of stops for Tier 3 agencies. Other outcomes, including receiving a citation or being arrested, varied widely across traffic and pedestrian stops, as pedestrian stops were more likely to end in an arrest and traffic stops were more likely to end in a citation, regardless of tier.

Finally, for searches, Tier 1 agencies conduct searches in 2.5 percent of stops, a higher percentage than Tier 2 and Tier 3. Of those searches, consent was obtained around half of the time for Tier 1 and 2 agencies and around a third of the time for Tier 3 agencies, while some other legal basis was reported in the remaining cases. Upwards of 40 percent of all searches were successful. Like in past reports, alcohol and drugs were the most commonly found items from a search.

[27] There was also one agency for which the difference between the white hit-rate and the Latinx hit-rate was significant, however the white hit-rate was lower than the Latinx hit-rate. This is not indicative of a racial/ethnic disparity for Latinx Oregonians, and therefore is not noted.

7.2. Veil of Darkness Findings 2021

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 again 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 for any minority group at the agency level.

In this analysis, two Tier 2 agencies were found to have a disparity in the rate of stopped minority drivers in daylight versus darkness compared to white drivers. Milwaukie PD shows the odds of stops for Black drivers in daylight was nearly 2.7 times the odds of white drivers, and Ashland PD shows the odds of stops for Latinx drivers in daylight was 2.3 times the odds of white drivers.

7.3. Predicted Disposition Findings 2021

The Predicted Disposition analysis, which relies 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 statistically significant disparities in their predicted versus actual dispositional outcomes for Asian/PI, Black, Latinx, Middle Eastern, and Native American groups, respectively.

In total, nine Tier 1 agencies, eighteen Tier 2 agencies, and nine Tier 3 agencies were identified as meeting this threshold. For Tier 1 agencies this included: Beaverton PD, Clackamas CO SO, Eugene PD, Hillsboro PD, Marion CO SO, Oregon State Police, Portland PB, Salem PD, and Washington CO SO. Among Tier 2 agencies, Albany PD, Canby PD, Forest Grove PD, Hermiston PD, Klamath Falls PD, Lincoln CO SO, Linn CO SO, McMinnville PD, Newberg-Dundee PD, Polk CO SO, Port of Portland PD, Redmond PD, Roseburg PD, Tigard PD, Tualatin PD, West Linn PD, Woodburn PD, and Yamhill CO SO were identified. For Tier 3 agencies, Brookings PD, Gervais PD, Gilliam CO SO, Hubbard PD, Malheur CO SO, Newport PD, Pendleton PD, Tillamook CO SO, and Umatilla PD were identified.

The most common dispositional outcome identified with disparate outcomes was citations, which, in general, is a much more common outcome than searches and arrests. Similarly, the group most often identified for disparate outcomes was Latinx, for which there were generally more stops relative to other non-white groups. 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 fully explain away the existence of disparities in all cases, these policies and controlling for them in future analyses represent an important next step in the analysis of stop data in Oregon.[28] There were, however, additional findings with regard to searches and arrests. Beaverton PD, Marion CO SO, Oregon State Police, Salem PD, Washington CO SO, and Woodburn PD were identified for searches of Latinx individuals. Marion CO SO, Salem PD, and Washington CO SO were indicated for arrests of Latinx individuals. Oregon State Police was identified for searches and arrests of Native American individuals. Portland PB was identified for searches and arrests of Black individuals. Klamath Falls PD was identified for searches and arrests of Asian/PI individuals. Finally, Pendleton PD was identified for searches and arrests of Native American individuals. Notably, many analyses for several agencies could not be estimated due to low sample sizes, especially for smaller

[28] For an example of the effect these policies can have on STOP Program analyses, please see Appendix E of the 2019 Statistical Transparency of Policing Report (https://www.oregon.gov/cjc/CJC%20Document%20Library/STOP_Report_Final.pdf)

agencies. In these situations we cannot detect the presence of a disparity with current data limitations. No analyses could be completed for several smaller agencies.[29]

7.4. KPT Hit-Rate Findings 2021

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 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 agency level.

In this analysis, only Oregon State Police was found to have a disparity meeting the above criteria. Specifically, Oregon State Police reported successful searches in 65.4 percent of searches involving white individuals but only reported successful searches in 37.5 percent of searches of Middle Eastern individuals.

7.5. Conclusions

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, 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 be 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 STOP Program researcher 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, 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 criteria described above, it is recommended that Oregon State Police be examined in greater detail by STOP Program researchers and receive technical assistance from DPSST. Oregon State Police was indicated as having a disparity in the Predicted Disposition analysis with regard to its citations of Black, Latinx, Asian/PI, Middle Eastern, and Native individuals; with regard to searches for Latinx and Native individuals; and with regard to arrests of Native individuals. In addition, the KPT Hit-Rate analysis identified a disparity with regard to searches of Middle Eastern individuals.

Regardless of whether an agency is officially referred to DPSST by this report or not, the CJC urges each agency to scrutinize the full set of results for their agency, found in the STOP Agency Summaries document on the CJC website[30]. While most agencies are not referred to DPSST in this analysis, that does not necessarily mean that the results for all those agencies should be ignored or are not close to the

[29] Full results, including for tests that could not be completed, are available upon request.
[30] https://www.oregon.gov/cjc/CJC%20Document%20Library/STOP_Agency_Summaries_2021_FINAL.pdf

threshold of identification. All agencies and/or interested stakeholders should contact the CJC should they require technical assistance in interpreting specific statistical results.

7.6. Next Steps and Future Work

The third annual STOP Program report includes data from 143 Oregon law enforcement agencies and is the first report to include data for all Tier 1, 2, and 3 agencies. This achievement culminates a multi-year effort to create a statewide data collection system for all officer-initiated traffic and pedestrian stops that are not associated with calls for service. While this is a significant milestone, the STOP program has encountered several challenges, including the impacts of the ongoing COVID-19 pandemic, resource limitations for some agencies, and the ongoing challenge of sample size limitations to conduct the rigorous statistical analyses necessary to identify disparities. However, as the first statewide STOP Program report, the rigorous analyses conducted do include results for nearly 100 agencies across the state and contribute to dialogues between law enforcement agencies and the communities they serve.

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.

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

Table A.1. Tier 1 Agencies

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

Table A.2. Tier 2 Agencies

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

Table A.3. Tier 3 Agencies

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

Please note that the CJC did not receive data from Burns PD, Curry County SO, Gearhart PD, Gold Beach PD, Harney County SO, Hines PD, Lake County SO, Port Orford PD, Powers PD, Scappoose PD, or St. Helens PD.

Appendix B – Data Audit

This third STOP report uses data with a higher frequency of missingness than in previous STOP reports. This missingness manifests in overall rates of missingness within variables (displayed in Table B.1.) and as stops missing altogether. Additionally, some data used in this report look atypical compared to overall rates. The STOP Program team has worked to the extent possible to correct for incorrect and improbable data before the release of this report but some challenges remain.

Table B.1. Missing Data for STOP Program Variables used in Year 3 Report Analyses

Variable	Description	Analyses Affected	% Missing
Age	Age as perceived by officer	Veil of Darkness, Predicted Disposition	0.8%
agency	Stopping agency	Veil of Darkness, Predicted Disposition, KPT Hit-Rate	0.0%
arrest	Physical custody arrest (yes/no)	Predicted Disposition	3.0%
CiteCat*	Category of citation (Move/Spd, Ser Move/Spd, Very Ser Move/Sp, Equip Vio/Cell/Seatbelt, Reg/License, Other)	Veil of Darkness	1.9%
cite_type	Citation basis for traffic stop (ORS, Municipal Traffic, Municipal Criminal, County Ordinance)	Veil of Darkness, Predicted Disposition	9.5%
county	County in which stop occurred	Veil of Darkness, Predicted Disposition	0.0%
disposition	Most severe disposition of stop (none, warning, citation, search, arrest)	Predicted Disposition	1.2%
gender	Gender perceived by officer (male, female, non-binary)	Veil of Darkness, Predicted Disposition	0.3%
race	Race/ethnicity perceived by officer (Asian/PI, Black, Latinx, Middle Eastern, Native American, white)	Veil of Darkness, Predicted Disposition, Hit-Rate	1.0%
sdate	Date of stop. Converted into day of the week, season, and time sun rises and sets of the day of the stop.	Veil of Darkness, Predicted Disposition	0.0%
search	Whether a discretionary stop occurred (yes/no)	Predicted Disposition, KPT Hit-Rate	0.2%
search_f1**	What was found if a search occurred (Nothing, Alcohol, Drugs, Stolen Property, Weapons, Other Evidence, Other non-Evidence)	KPT Hit-Rate	2.6%
search_t1**	Search type	KPT Hit-Rate	0.0%
stime	Time of stop. Converted into time categories (12-5 am, 5-10 am, 10 am-3 pm, 3-8 pm, and 8 pm-12 am)	Veil of Darkness, Predicted Disposition	0.0%
stop_type	Type of stop (traffic, pedestrian)	Veil of Darkness	0.2%

*CiteCat is a condensed variable created from the original variables cite_code and cite_text, which denote the ORS code and text description, respectively, for the citation. Not every stop ends in citation, so the percent missing reflects only those stops that end in citation that are missing CiteCat.

**These missing percentage reflects the percent of missing when an entry is likely expected. In the case that Search= "no", there is not an entry expected, so these are not included in the missing percentage in this table.

Table B.1. displays the overall rates of missingness for variables used in STOP analyses for Year 3 data, however these rates vary widely between agencies. For example, Age is missing for less than 1% of Year 3 stops, 3,641 stops. Of those 3,641 stops, 85% are from three agencies. For two of these agencies, these stops with missing age represent 0.8 and 1.1 percent of their total stops in Year 3, respectively. The third agency experienced a technical challenge while submitting data in Year 3 and submitted stops directly to CJC for 748 stops that were missing from their vendor-submitted data. The 748 stops that were submitted separately were missing most identifiers of stopped individuals but corrected the overall count of the agency's stops. This agency has since implemented technical enhancements so stops occurring in future years are unlikely to be missing data. This one agency's stops also make up almost all the stops that are missing either search or stop_type in Year 3.

Information on physical custody arrest is missing for 3 percent of stops, 14,292 stops, in Year 3 data. 95 percent of these stops that are missing arrest are stops made by one Tier 1 agency during March 2021. Some agencies have much higher arrest rates than average. For six Tier 3 agencies, 50 percent or more of their stops end in arrest. On average, 2 percent of stops made by Tier 3 agencies end in arrest. STOP Program researchers will work with these seven agencies in the upcoming year to better understand and correct, when necessary, this data going forward.

Some missing data is concentrated within single agencies or areas. CiteCat is unknown for 1.9 percent of stops, 6,640 stops. Of these 6,640 stops with Unknown CiteCat, 73 percent are stops from one Tier 2 agency. Cite_type is missing for 9.5 percent of stops in year 3, however 78 percent of these stops that are missing cite_type are from one county. Disposition is missing for 1.2 percent of stops and 94 percent of the stops that are missing disposition occur in two counties. 55 percent, 2,792 stops that are missing race were made by one large Tier 1 agency. However, 2,792 represents only 2 percent of this agency's stops in Year 3. Search_f1 is missing for 263 searches in Year 3. 235, or 90 percent of the searches that are missing search_f1 are from one agency. The concentration of missingness in specific agencies or counties points to a technical problem that the STOP Program team will work with these agencies to correct going forward.

Gender is missing rarely in Year 3 data, 0.34 percent, however some agencies have more non-binary stops than most other agencies. For four agencies, non-binary individuals make up nine percent of more of the agency's stops. For all other agencies, 5 percent or fewer of their stops are of non-binary individuals.

Tables B.2. and B.3. display agencies that did not report data for some or all of Year 3. All the agencies listed in Table B.2. and B.3. are Tier 3 agencies which would have been reporting for the first time in Year 3. All Tier 3 agencies were required to begin reporting in July 2020, while facing the challenges associated with COVID-19 that likely exacerbated resource and staffing challenges.

CJC did not receive data from eleven Tier 3 agencies. These agencies are listed in Table B.2. No analysis was done for these agencies. Curry County SO, Gearhart PD, Gold Beach PD, and Port Orford PD did not submit due to technological challenges with their data vendor.

Table B.2. Agencies Which Did Not Submit Year 3 Data
Burns PD
Curry County SO
Gearhardt PD
Gold Beach PD
Harney County SO
Hines PD
Lake County SO
Port Orford PD
Powers PD
Scappoose PD
St Helens PD

Some Tier 3 agencies did not report stops in every month of Year 3. Table B.3. displays agencies where there was one or more months during which no stops were reported. CJC did all analyses when possible for these agencies. Some agencies reported no stops because they truly made no stops during these months. For example, CJC was able to verify with Butte Falls, Union Pacific Railroad PD, and Vernonia PD that their stop count was correct. However, some of the agencies listed in Table B.3. likely had stops which occurred during these months that they did not report to CJC. OSU PD is a new agency as of January 1, 2021, and thus only submitted data for January to June 2021. This agency is not listed in Table B.3.

Table B.3. Agencies Which Reported No Stops in One or More Months

Agency	Months with No Stops Reported
Boardman PD	July 2020-February 2021
Brookings PD	July-August 2020
Butte Falls PD	July 2020, December 2020-June 2021
Columbia City PD	January 2021
Coos Bay PD	July-August 2020
Coquille PD	July-August 2020
Cottage Grove PD	July 2020-January 2021
Florence PD	July-August 2020
Gervais PD	May 2021
Grant CO SO	July-August 2020, December 2020, January 2021
Jacksonville PD	August 2020-June 2021
John Day PD	December 2020-February 2021
Junction City PD	July-August 2020
Malheur CO SO	July-August 2020, November 2020-February 2021
Malin PD	July 2020
Myrtle Point PD	June 2021
Nyssa PD	July-August 2020, November 2020-February 2021
Oakridge PD	July 2020
Ontario PD	July-August 2020, November 2020-February 2021
Pilot Rock PD	May 2021
PSU CPS	August 2020-June 2021
Rainier PD	July 2020-December 2020
Sweet Home PD	July-August 2020
Union Pacific Railroad PD	November-December 2020
Vernonia PD	November 2020
Wallowa CO SO	July-December 2020

Appendix C – Stop Characteristics for Tier 3 Agencies

Table C.1. Race/Ethnicity Reporting for Tier 3 Agencies for All Reported Stops

Agency	Asian/PI	Black	Latinx	Middle Eastern	Native American	White
Astoria PD	47	60	146	11	6	2,664
Aumsville PD	20	9	113	2	0	655
Baker City PD	7	25	64	1	4	929
Baker CO SO	16	24	50	11	0	547
Bandon PD	29	13	52	11	0	618
Black Butte Ranch PD	11	7	31	7	0	424
Boardman PD	0	3	34	0	0	47
Brookings PD	40	28	166	11	4	1,554
Butte Falls PD	0	0	0	0	0	9
Cannon Beach PD	63	43	164	45	5	1,536
Carlton PD	7	0	21	5	0	196
Clatsop CO SO	27	34	128	12	1	1,569
Coburg PD	14	21	74	16	0	671
Columbia City PD	1	2	7	0	0	100
Columbia CO SO	26	29	62	11	4	1,728
Coos Bay PD	25	32	75	9	8	2,415
Coos CO SO	11	6	69	3	4	851
Coquille PD	9	13	43	3	3	1,141
Cottage Grove PD	2	1	13	2	0	188
Crook CO SO	20	46	245	11	11	2,410
Dallas PD	20	21	125	4	0	927
Eagle Point PD	14	23	145	5	1	1,086
Enterprise PD	1	3	7	0	2	107
Florence PD	15	2	17	1	1	818
Gervais PD	13	7	148	5	0	571
Gilliam CO SO	41	46	221	25	0	1,427
Gladstone PD	72	145	323	40	24	2,359
Grant CO SO	0	1	3	0	0	59
Hood River PD	48	36	531	17	24	1,263
Hubbard PD	34	26	853	9	0	881
Independence PD	20	30	244	11	4	893
Jacksonville PD	1	1	1	0	0	13
Jefferson CO SO	29	9	116	6	2	730
John Day PD	1	0	6	0	0	179
Josephine CO SO	59	54	303	20	4	2,586
Junction City PD	5	17	55	1	0	541
La Grande PD	38	25	41	1	0	782
Madras PD	7	8	156	3	40	376
Malheur CO SO	2	3	80	0	2	356
Malin PD	1	2	53	2	0	99
Manzanita DPS	21	6	23	10	0	280
Merrill PD	9	6	38	2	0	99
Milton-Freewater PD	12	17	312	5	5	747
Molalla PD	12	28	144	10	2	1,608
Monmouth PD	26	53	157	10	4	689
Morrow CO SO	14	49	590	8	12	1,200
Mt. Angel PD	13	6	166	6	0	375
Myrtle Creek PD	13	15	47	2	0	1,112
Myrtle Point PD	7	4	7	2	1	136
Newport PD	58	38	265	17	9	1,439
North Bend PD	12	18	62	8	7	890

(Table C.1. continued on next page)

Nyssa PD	0	4	52	0	0	136
Oakridge PD	10	8	11	3	0	132
Ontario PD	4	10	93	0	0	407
Pendleton PD	44	70	307	7	276	2,397
Philomath PD	59	41	125	21	0	1,492
Phoenix PD	17	14	143	2	0	523
Pilot Rock PD	1	0	3	0	1	85
PSU CPS	0	2	1	0	0	10
Prineville PD	27	45	287	8	10	3,248
Rainier PD	2	5	14	0	0	218
Reedsport PD	14	12	22	5	0	264
Rockaway PD	12	2	23	5	0	322
Rogue River PD	8	13	48	2	0	292
Sandy PD	37	27	139	17	15	1,164
Seaside PD	69	66	271	39	4	2,196
Sherman CO SO	43	31	209	19	4	741
Sherwood PD	101	100	286	22	8	2,471
Silverton PD	11	10	134	3	0	694
Stanfield PD	14	29	335	24	8	934
Stayton PD	7	12	120	4	0	1,001
Sunriver PD	24	4	61	7	0	1,141
Sutherlin PD	14	14	87	3	0	1,061
Sweet Home PD	0	2	6	0	0	228
Talent PD	18	28	64	6	0	649
The Dalles PD	29	22	228	4	35	1,043
Tillamook CO SO	41	15	158	14	1	1,014
Tillamook PD	15	4	90	6	2	469
Toledo PD	31	35	124	0	26	1,770
Turner PD	3	1	7	0	0	112
Umatilla CO SO	7	16	304	10	3	748
Umatilla PD	19	65	1,637	9	12	1,993
Union CO SO	24	21	67	6	2	664
Union Pacific Railroad PD	0	1	12	0	0	50
Vernonia PD	0	0	2	0	0	127
Wallowa CO SO	3	1	0	0	0	42
Warrenton PD	21	12	79	6	1	1,263
Wasco CO SO	13	11	105	3	15	540
Wheeler CO SO	14	5	21	8	10	646
Winston PD	10	16	41	2	1	1,147
Yamhill PD	27	20	129	14	0	715
OSU PD	7	5	12	0	1	88
Total Tier 3	**1,793**	**1,894**	**12,653**	**680**	**629**	**79,117**

Table C.2. Percent and Number of Tier 3 Agency Stops by Stop Type Traffic vs. Pedestrian

Agency	Traffic		Pedestrian		Total
Astoria PD	2,933	100.0%	1	0.0%	2,934
Aumsville PD	782	97.9%	17	2.1%	799
Baker City PD	1,019	98.9%	11	1.1%	1,030
Baker CO SO	642	99.1%	6	0.9%	648
Bandon PD	710	98.2%	13	1.8%	723
Black Butte Ranch PD	479	99.8%	1	0.2%	480
Boardman PD	85	100.0%	0	0.0%	85
Brookings PD	1,803	100.0%	0	0.0%	1,803
Butte Falls PD	9	100.0%	0	0.0%	9
Cannon Beach PD	1,566	84.4%	290	15.6%	1,856
Carlton PD	218	95.2%	11	4.8%	229
Clatsop CO SO	1,769	99.9%	2	0.1%	1,771
Coburg PD	786	98.7%	10	1.3%	796
Columbia City PD	114	100.0%	0	0.0%	114
Columbia CO SO	1,849	99.4%	11	0.6%	1,860
Coos Bay PD	866	33.8%	1,698	66.2%	2,564
Coos CO SO	924	97.9%	20	2.1%	944
Coquille PD	305	25.2%	907	74.8%	1,212
Cottage Grove PD	206	100.0%	0	0.0%	206
Crook CO SO	2,708	98.7%	35	1.3%	2,743
Dallas PD	1,032	93.2%	75	6.8%	1,107
Eagle Point PD	1,256	98.6%	18	1.4%	1,274
Enterprise PD	118	98.3%	2	1.7%	120
Florence PD	753	88.2%	101	11.8%	854
Gervais PD	795	99.6%	3	0.4%	798
Gilliam CO SO	1,760	100.0%	0	0.0%	1,760
Gladstone PD	2,908	98.1%	55	1.9%	2,963
Grant CO SO	63	100.0%	0	0.0%	63
Hood River PD	1,905	99.3%	14	0.7%	1,919
Hubbard PD	1,792	99.4%	11	0.6%	1,803
Independence PD	1,193	99.3%	9	0.7%	1,202
Jacksonville PD	16	100.0%	0	0.0%	16
Jefferson CO SO	888	99.4%	5	0.6%	893
John Day PD	184	98.9%	2	1.1%	186
Josephine CO SO	2,851	94.2%	175	5.8%	3,026
Junction City PD	617	99.7%	2	0.3%	619
La Grande PD	887	100.0%	0	0.0%	887
Madras PD	582	98.6%	8	1.4%	590
Malheur CO SO	442	99.8%	1	0.2%	443
Malin PD	157	100.0%	0	0.0%	157
Manzanita DPS	340	100.0%	0	0.0%	340
Merrill PD	154	100.0%	0	0.0%	154
Milton-Freewater PD	1,092	99.5%	6	0.5%	1,098
Molalla PD	1,643	91.1%	161	8.9%	1,804
Monmouth PD	933	99.4%	6	0.6%	939
Morrow CO SO	1,850	98.5%	28	1.5%	1,878
Mt. Angel PD	556	98.2%	10	1.8%	566
Myrtle Creek PD	1,141	95.7%	51	4.3%	1,192
Myrtle Point PD	156	99.4%	1	0.6%	157
Newport PD	1,907	96.5%	70	3.5%	1,977
North Bend PD	853	85.6%	144	14.4%	997
Nyssa PD	192	100.0%	0	0.0%	192
Oakridge PD	164	100.0%	0	0.0%	164
Ontario PD	513	99.8%	1	0.2%	514

(Table C.2. continued on next page)

Pendleton PD	2,585	83.4%	516	16.6%	3,101
Philomath PD	1,728	99.4%	10	0.6%	1,738
Phoenix PD	633	90.6%	66	9.4%	699
Pilot Rock PD	90	100.0%	0	0.0%	90
PSU CPS	0	0.0%	13	100.0%	13
Prineville PD	3,708	98.9%	42	1.12%	3,750
Rainier PD	237	99.2%	2	0.8%	239
Reedsport PD	305	96.2%	12	3.8%	317
Rockaway PD	344	94.5%	20	5.5%	364
Rogue River PD	347	95.6%	16	4.4%	363
Sandy PD	1,388	99.2%	11	0.8%	1,399
Seaside PD	2,524	95.4%	121	4.6%	2,645
Sherman CO SO	1,047	100.0%	0	0.0%	1,047
Sherwood PD	2,969	99.4%	19	0.6%	2,988
Silverton PD	852	100.0%	0	0.0%	852
Stanfield PD	1,365	99.6%	6	0.4%	1,371
Stayton PD	1,096	95.8%	48	4.2%	1,144
Sunriver PD	1,218	98.3%	21	1.7%	1,239
Sutherlin PD	1,133	96.1%	46	3.9%	1,179
Sweet Home PD	235	99.6%	1	0.4%	236
Talent PD	756	98.8%	9	1.2%	765
The Dalles PD	1,536	91.8%	138	8.2%	1,674
Tillamook CO SO	1,231	98.9%	14	1.1%	1,245
Tillamook PD	579	98.8%	7	1.2%	586
Toledo PD	1,979	99.6%	7	0.4%	1,986
Turner PD	121	98.4%	2	1.6%	123
Umatilla CO SO	1,223	98.5%	19	1.5%	1,242
Umatilla PD	3,789	99.6%	17	0.4%	3,806
Union CO SO	783	99.9%	1	0.1%	784
Union Pacific Railroad PD	0	0.0%	63	100.0%	63
Vernonia PD	129	100.0%	0	0.0%	129
Wallowa CO SO	45	97.8%	1	2.2%	46
Warrenton PD	1,381	99.9%	1	0.1%	1,382
Wasco CO SO	768	99.4%	5	0.7%	773
Wheeler CO SO	698	99.1%	6	0.9%	704
Winston PD	1,203	98.8%	14	1.2%	1,217
Yamhill PD	905	100.0%	0	0.0%	905
OSU PD	90	79.6%	23	20.4%	113
Total Tier 3	**92,486**	**94.6%**	**5,289**	**5.4%**	**97,775**

Appendix D – Veil of Darkness Technical Appendix and Detailed Results

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:

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

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 or 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, 2019 to June 30, 2021. 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, and the latest end of civil twilight is 9:48pm in Clatsop

[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, and 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.

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 2019 and March 2020. 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.

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].

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

Appendix E – Predicted Disposition Technical Appendix and Detailed Results

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 $= \frac{(propensity\ score)}{1-(propensity\ score)}$
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 estimand 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).

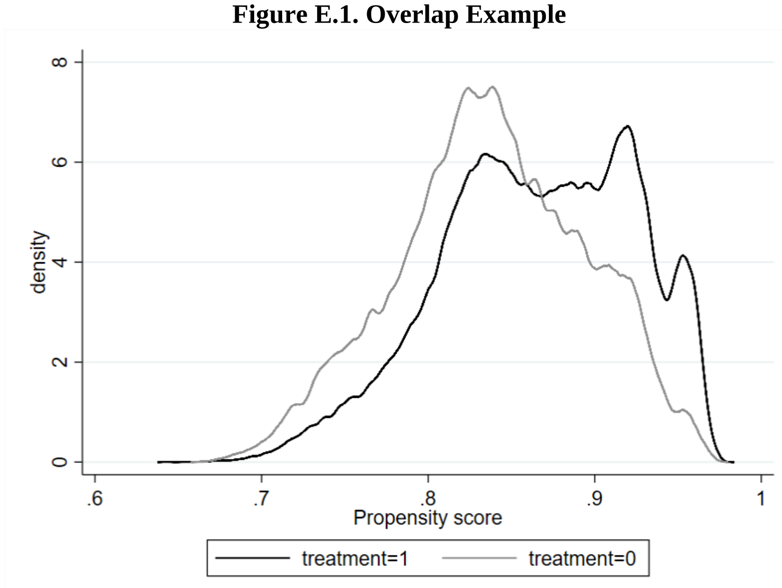
[35] For a thorough discussion of IPWRA methods, see Wooldridge 2010, Chapter 21.3.4.

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 E.1. 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 E.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]

[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.

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 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 Latinx individual causes another Latinx 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 Asian/PI, = 0 if white
2. Officer perception of race/ethnicity: = 1 if Black, = 0 if white
3. Officer perception of race/ethnicity: = 1 if Latinx, = 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 is 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 or Very Serious Moving/Speeding; Equipment, Cell, or Seatbelt; Registration/License; Other
7. Day of week indicators
8. Agency stop volume $= \frac{Total\ \#\ of\ stops\ by\ agency\ on\ day\ of\ stop}{Maximum\ \#\ of\ daily\ stops\ by\ \boldsymbol{agency}\ over\ year\ of\ analysis}$
9. County stop volume $= \frac{Total\ \#\ of\ stops\ by\ agency\ on\ day\ of\ stop}{Maximum\ \#\ of\ daily\ stops\ in\ the\ \boldsymbol{county}\ over\ year\ of\ analysis}$

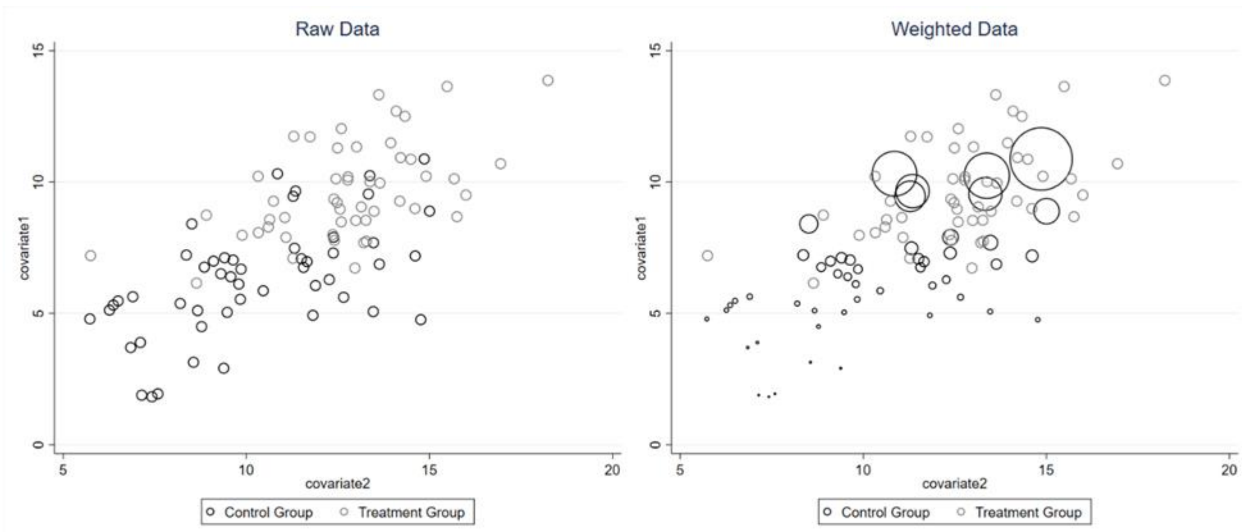
The first step of the analysis uses a probit model to estimate the propensity of being in the treatment group based on the covariates listed above. Overlap of propensity scores is evaluated and any non-overlapping observations are removed from the sample. Inverse Probability Weights (IPWs) are estimated for each observation based on the propensity scores. For the treatment group in an ATET framework

[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).

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 E.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 E.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 E.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

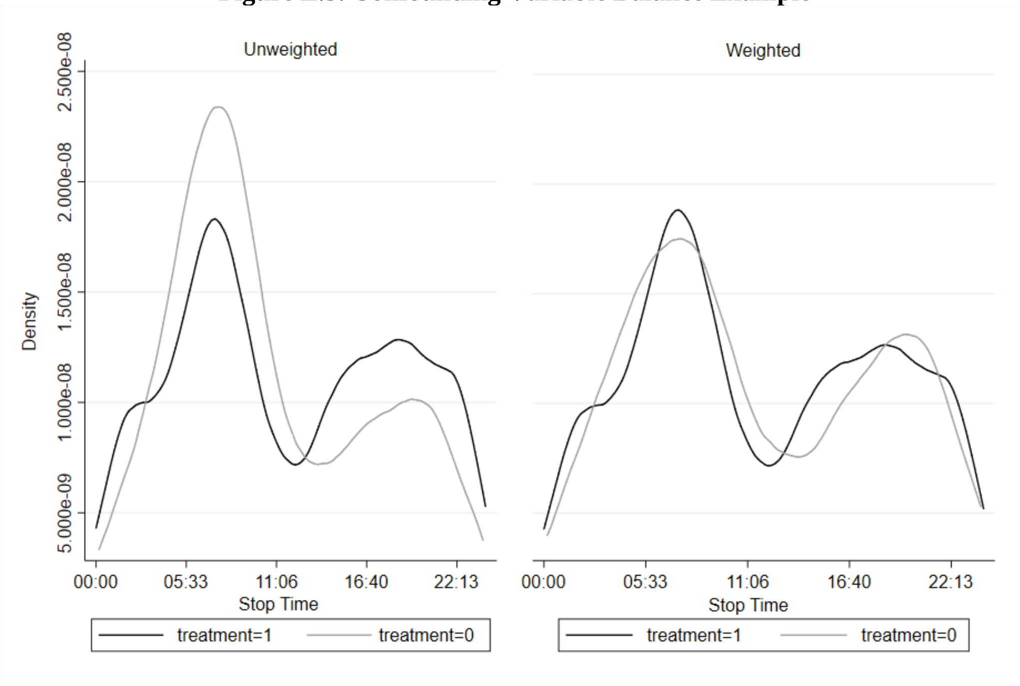
---

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

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

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 E.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 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 (minority) group 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

[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 minority groups and averages across all observations.

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 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-R^2s 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.

Appendix F – 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 over searching 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" (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-square 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% with a Bonferroni correction for multiple testing determined significance. Each agency's white hit-rate was compared to each race group (Black, Latinx, Asian/PI, Middle Eastern, and Native) dependent upon sample size, so a Bonferroni corrected p-value of $0.05/5 = 0.01$, $0.05/4$, $0.05/3$, $0.05/2$, or 0.05 was used, dependent upon the number of groups for which the analysis was able to be performed. Hit-Rate analyses and accompanying statistical tests were performed with the statistical software R.

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.

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% of the time (while all other members of the group carry contraband less than 50% of the time). Suppose also that group B has some portion of members that instead carry contraband 75% of the time (while all other members of the group carry contraband less than 50% of the time). If an officer only searches every person (regardless of group) who has over a 50% 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 infra-marginality 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.