STOP Program Research Brief

Analytical Approaches to Studying Stops Data

October 2018

Oregon Criminal Justice Commission

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

1. INTRODUCTION

The statistical examination of police traffic and pedestrian stop data was borne out of litigation in the mid-1990s. For decades, advocacy groups cited anecdotal evidence supporting the notion that law enforcement applies different standards to minority drivers and pedestrians concerning various police-citizen interventions. The specific and systematic measurement of these police practices during stops, however, did not occur until it was required in *State of New Jersey v. Soto et al.*, a case where a group of defendants claimed that New Jersey state troopers targeted highway drivers based on the color of their skin. In *Soto*, as well as in *Wilkins v. Maryland State Police*, which occurred at the same time, courts sought empirical data that could properly compare the rates of traffic stops to the driving population at risk for being stopped by the allegedly offending law enforcement agencies. In each of these instances, the courts decided to compare the rates at which minority drivers were stopped by law enforcement with estimates of the driving population based on observational traffic studies prepared during litigation.

Following the conclusions of the cases cited above, the US Department of Justice, along with several professional organizations, began hosting conferences concerning the improvement of police-community relationships. A specific focus of a number of these conferences, including those hosted by the Police Executive Research Forum in 2001, the SOROS Foundation and the Institute on Race and Justice at Northeastern University in 2003, as well as others that followed, was on the collection, analysis, and public reporting of traffic and pedestrian stop data. In response, many states began to mandate the collection of traffic stop data, and in states that did not or have yet to call for data collection legislatively, many local jurisdictions and departments began to collect and analyze stop data on their own.

While numerous individual law enforcement agencies in Oregon have been collecting and analyzing data from traffic and/or pedestrian stops for many years, the State did not mandate statewide collection of this data until the passage of HB 2355 in 2017. Under HB 2355, the Oregon State Police, in cooperation with the Oregon Criminal Justice Commission (CJC), was required to develop and implementing a statewide, standardized method for recording data on officer-initiated traffic and pedestrian stops occurring in the state. Data collection is being phased in over a three-year period, with large agencies to begin data collection in July of 2018, and medium and small agencies following in 2019 and 2020, respectively.[1] Following data collection, the CJC is required to perform statistical analyses examining the prevalence and disposition of officer-initiated traffic and pedestrian stops. This purpose of this brief is to examine the extant research examining stops by law enforcement and represents a foundation for the selection and deployment of appropriate analytical approaches and methodologies to ensure that the CJC meets the requirements outlined in HB 2355.

2. ASSESSING AND ANALYZING TRAFFIC STOP DATA

Since *State of New Jersey v. Soto et al.*, there has been significant development and innovation within the academic and applied research literature concerning the analysis of traffic and pedestrian stop data. As of this time, however, no single approach is considered to be the single, gold standard for analyzing stop outcomes. The lack of a singular approach to the analysis of this type of data stems from several sources. The first can be traced to the nature of the traffic stop itself. Initially, it is tempting to view a traffic stop

[1] For the purposes of the law, large agencies are those with 100 or more officers with stop authority, while medium and small agencies are those with 25 to 99 and less than 25 officers with stop authority, respectively. Lists identifying the tier into which an agency falls can be found at https://www.oregon.gov/cjc/stop/.

as a single instance of police-citizen contact that can be assessed for the presence or absence of discriminatory behavior by a law enforcement official. Within the time it takes to execute and conclude a single stop, however, there are numerous opportunities where racially disparate treatment could or might be present. Beyond the initial decision to stop a driver or pedestrian, race could be a factor in the decision to search, give a citation, or make an arrest. This distinction is critical, because both the data and analytical techniques required to analyze the various decision points found in a single officer-initiated stop differ substantially.

The second challenge encountered by researchers is that it is often difficult to account for three alternative explanations that could possibly explain observed racial disparities found in stop data (Ridgeway 2009). The first explanation when data purports to show different stop rates by race is that *racially biased policing* drives patterns of this type. Indeed, as argued by community groups and many citizens, it very well could be that law enforcement actively seek out individuals of certain groups when engaging in traffic enforcement. The second explanation, however, challenges the first, and suggests that differences in stop rates could be attributable to *differences in driving behavior and/or offending by race*. Under this explanation, if drivers of a certain race are stopped more often than drivers of other races, it could be attributable to the fact that members of the affected groups commit traffic violations at a higher rate. If this were the case, disparities in stop rates would be legitimately tied to differences in violation frequency and thus would not result from racial animus.

The final explanation is that the *exposure to law enforcement may vary by race*, which would indicate that disparities found in stop data could be attributable to exposure or enforcement decisions, not racial discrimination. For example, members of one race could be exposed to law enforcement more often because they drive more frequently or for longer distances than members of other racial groups, which would put them at higher risk of being stopped by law enforcement. Alternatively, if the police patrol more often in areas frequented by certain racial groups, then purported racial disparities could be due to differential exposure and enforcement. This could be particularly relevant in neighborhoods with high calls for service volmen, given that law enforcement resources are more likely to be deployed in high crime areas and because localized crime trends could increase suspicion on the part of officers on patrol. The challenge for researchers, therefore, is to find suitable methods for addressing the various decision points found in a single police-citizen encounter, while also working to address the three possible explanations of racial disparities discussed above.

2.1 The Initial Decision to Stop a Citizen

Conducting analyses examining the decision to initiate a stop presents the most serious challenge of any stage of police-citizen stop data research. When law enforcement agencies collect data on stops, the recorded data only covers the encounters that occur between law enforcement and citizens. This is problematic from a research standpoint, as the means for determining whether a demographic group is treated in a disparate manner is to compare the frequency with which that group is stopped to the *expected* stop rate for the group that would exist if no discrimination was present. This comparison case, often referred to as a *benchmark*, is necessary because knowing that a law enforcement agency stops a certain percentage of people from a single group says nothing about whether the affected group is being unfairly targeted relative to their risk of being stopped. Indeed, this issue is tied directly to the third explanation for racial disparities described above—the differential exposure to law enforcement—as we would expect stop rates to mirror, or at least be very similar to, the degree to which a group is exposed to law enforcement.

Because it does not exist in the stop data itself, researchers must select a benchmark when conducting analyses examining the initial decision to stop an individual. The selection of this benchmark is a critical decision point in the research process, as the researcher must strive to find data that represents, as far as practicable, the population exposed to law enforcement when the stop is made. Failure to select an appropriate benchmark, importantly, can lead researchers to draw incorrect conclusions concerning disparities found in the data. Further complicating this issue is the fact that there are multiple benchmarks that researchers can select from, and the fact that no benchmark can easily represent the population exposed to law enforcement. In the subsections that follow, benchmarks used in past analyses of police stops data will be examined and their strengths and weaknesses will be discussed.

Finally, when making comparisons between stops data and benchmarks, researchers have relied on general rules of thumb to determine whether a disparity reaches the threshold at which the difference between reported stop rates and a benchmark can be considered robust. Traditionally, this threshold has been five percentage points, which means that if a group is stopped at a rate that is five percentage points above its share of the population, then a disparity is present that cannot be attributed to natural or random variation (Lovrich et al. 2007). For example, if Hispanics represent 10 percent of the population in a municipality, and 16 percent of all traffic stops made by the local police department involved Hispanics, then the empirical research would classify this hypothetical department as having a racial disparity.

2.1.1 Census Residential Benchmarks

The use of residential Census data for an area as a proxy estimate for that area's population at risk for encountering law enforcement is very common. The types of Census data used to create these benchmarks, however, varies significantly both in scope and complexity. The most commonly used Census benchmark compares stop rates for an aggregate area—a municipality, county, or state—to the residential population reported by the Census for that area (*see* Engel and Calnon 2004). Relatedly, Census data from smaller, disaggregated areas can be used, such as data from neighborhoods, police precincts, census blocks, or census tracts. Finally, some analyses rely on estimates of an area's driving population, which can be constructed using several types of Census data. In general, the primary strength of Census data and its related benchmarks lies in its accessibility and relative simplicity. Census data is free to access, certain programs are updated on a yearly basis, and the nature of the data permits researchers to aggregate or disaggregate it on multiple levels across multiple groups.

In spite of is accessibility, many researchers, law enforcement officials, and various other groups are highly critical of the use of residential Census data as a benchmark in stops analyses. While there are nuances to these critiques, they all center on one central argument: that residential population data does not accurately reflect the driving population or population at risk of encountering law enforcement. For example, many researchers argue that the residential population of a given area is an inappropriate benchmark for the driving population because it cannot account for daily inflows and outflows of individuals from the area under examination. These inflows and outflows can occur for numerous reasons; whether they are caused by commuters entering a city for work on a daily basis, individuals driving to an area for entertainment and shopping excursions, or the fact that the driving population in an area sitting along a major interstate will likely be impacted by flow-through traffic. Relatedly, seasonal population changes can also be problematic, whether driven by tourism during certain months or the significant changes that some municipalities with universities experience when students return to campus for classes.

In an attempt to address the issues raised by the disjuncture between residential demographic data and the driving population, some researchers have developed more complex approaches which attempt to bridge some of the gaps described above. The State of Connecticut, which is often touted as representing the cutting edge in racial profiling analyses, developed a method it argues provides a reasonable estimate of an area's driving population by more accurately accounting for daily work commuters (see Fazzalaro and Barone 2014). This approach, which builds on a similar method first developed by Northeastern University's Institute on Race and Justice for use in Delaware and Massachusetts, uses Census data to construct the demographic features of the population of individuals each feeder town or area contributes to the area under consideration. Those population estimates are then used to augment the residential population of the target area to get a better estimate of the driving population. Thus, for instance, if Census data demonstrates that two suburbs contribute work commuters to a target city's daily population, it is possible to map the demographic characteristics of the contributing suburbs onto the driving population of the target city to get a more accurate representation of the city's driving population. While promising, this approach has its drawbacks, however, as it is still difficult to account for all outside contributions to an area's driving population, such as commuting for recreation, shopping, and/or tourism.

The second major critique of residential census data as a benchmark for the driving population at risk for encountering law enforcement is that research indicates there are differences in driving patterns across age, gender, and racial/ethnic groups. This is problematic because the use of residential Census data relies on the assumption that individuals drive at rates proportionate to their share of the population in a given area. Analyses of the National Household Transportation Survey, for example, found significant racial differences in driver's license possession, the frequency of public transportation usage, and vehicle ownership rates (Engle and Calnon 2004). For example, the 2009 NHTS reported that Hispanics, African Americans, and other non-white, Non-Hispanic groups were all less likely to use private vehicles as their primary mode of transportation on a daily basis (FHWA NHTS Brief 2016). Hispanics were more likely to walk than Non-Hispanic Whites and African Americans, while African Americans were more likely than all other racial/ethnic groups to use public transportation. Thus, even if African Americans make up 5 percent of an area's residential population, research would suggest that it would be inappropriate to assume that they make up 5 percent of that area's driving population.

Third, while Census data is updated on a regular basis, researchers should proceed with caution when using these datasets for developing benchmarks for racial and ethnic demographic groups due to the presence of significant sampling error in many instances. The most accurate and reliable counts of individuals within racial/ethnic groups is found in each decennial Census, as the Census Bureau attempts to count each individual residing in the United States. In an effort to provide population data in between the decennial Census, the Census Bureau releases several types of population data, including yearly estimates at the state and county levels under the Census Population Estimates Program, as well as one and five-year American Community Survey (ACS) estimates.[2] Researchers seeking the most accurate, up-to-date population estimates should use the yearly state/county data released by the Census Bureau pursuant to the Census Population Estimates Program. This data, however, does not permit the researcher to drill down to data on entities within counties, such as cities and towns. For researchers interested in up-to-date data on cities and towns, the one and five-year ACS estimates are the only avenue, but these

[2] For further information concerning the Population Estimates Program, please use this link. For further information on the American Community Survey, please use this link.

estimates are often subject to large amounts of sampling error. In most cases, due to this error, it is advisable to use the five-year ACS data.[3]

2.1.2. Licensed Driver Data

Licensed driver data is another benchmark used in studies of traffic stops. Compared to residential population estimates, even those limited to the population of driving age, it is likely that using data on individuals with a driver's license presents a better picture of the driving population given that licensed individuals have taken at least one active step to identify themselves as a potential user of a motor vehicle. License data, however, also has some serious limitations. First, as described above, differences in driving patterns and transportation usage by race likely impacts the use of licensed driver data. Simply possessing a license is no guarantee that the individual uses it or even has access to a vehicle. Second, information on licensed drivers cannot account for other factors, including miles driven and motor vehicle trip frequency, vehicle ownership, as well as for inflows and outflows from the target area being considered (Alpert Smith, and Dunham 2004). Finally, data on licensed drivers is also more difficult to obtain than Census data. As such, it is rarely used in analyses of stops data.

2.1.3. Roadway Observational Benchmarks

Roadway observational studies offer an intriguing alternative to the creation of benchmarks using Census data as well as other data sources that act as proxies for the driving population. This approach can address many of the problematic issues described above for residential data, as direct observations of the driving population can account for driving patterns, trip frequency, and perhaps even the possible confounding effect of differential offending rates by race. Indeed, Alpert (2004: 47), argues that observational studies of drivers are "the best way to estimate the population of drivers available to be stopped" by police.

Lamberth was the first researcher to collect and examine observational driver data in the early to mid-1990s for the *New Jersey v. Soto et al.* case by observing traffic patterns in 18 sessions over a two week period at randomly selected dates, times, and locations within the State of New Jersey. Relying on the identification of drivers using a White-Black binary coding system, Lamberth's research team was able to identify the race of nearly 42,706 drivers during the data collection process and used their estimates of the driving population as the benchmark for comparison to traffic stop data obtained from law enforcement. Building on this initial study, Lamberth and others created additional observational benchmarks for subsequent litigation as well as for Washington DC Metropolitan Police (Lamberth 2006).

In spite of the attractiveness of this approach, however, the use of observational studies to construct driving population estimates presents some serious hurdles. Chief among them is cost, as studies of this type are difficult to implement, costly due to labor requirements, and time consuming (Engle and Calnon 2004). Second, related to concerns over cost, observational studies quickly become outdated as towns and cities grow, and populations shift (Fazzalaro and Barone 2014). Proponents of roadway observational studies have attempted to limit costs by observing patterns in a selected number of sample areas within a jurisdiction. While this approach does have a positive impact on cost, it presents generalizability issues. Third, it is difficult to assess the accuracy of the observations of race recorded by researchers, particularly as coding schemes become more complex and move beyond a white-nonwhite dichotomy. Indeed,

[3] For further detail on the differences between the ACS data collection programs, please use this link.

observations of individuals' racial identities by third parties, even when the target and assessor are in close proximity to one another, are often inaccurate (see Campbell and Troyer 2007, for a review).

2.1.4. Traffic Accident Data

Another approach to the establishment of driving population benchmarks is the use of traffic accident data. First pioneered in Florida in the early 2000s, researchers utilized demographic data from counts of not-at-fault drivers involved in two-car accidents to create a racial distribution of an area's driving population (Alpert, Smith, and Dunham 2004). This approach is based on the assumption that not-at-fault accidents are randomly distributed by race and should approximate the racial demographic profile of an area. More recently, other studies have utilized injury accident data, including analyses conducted by the Portland Police Bureau.[4] The assumption that accident data accurately reflects the driving population has been examined in several recent studies examining police traffic stops (Alpert at al. 2004; Lovrich et al. 2007). The research finds that traffic accident data provides a reasonable estimate of the driving population based on comparisons with observational traffic studies.

The chief drawback to the use of traffic accident data, however, is the fact that it can be difficult to obtain. While law enforcement agencies maintain records of traffic accidents, access to those records is often beyond the reach of organizations outside of law enforcement. Alternatively, other sources of data, such as accident data from state departments of motor vehicles, often does not contain the data fields required to perform analyses, such as data on race.

2.1.5. Traffic Violation Data

A limited number of studies have used data regarding the offending patterns of individuals by race as a benchmark for the driving population. This benchmark, unlike the others outlined previously, offers a different comparison group, as it moves beyond merely benchmarking the pool of at risk drivers to purportedly providing a measure of the population engaging in criminal behavior and thus at risk for being stopped by law enforcement. Studies relying on this type of benchmark generally fall into two categories: studies using data on offending rates for a general cross section of crimes, and data on traffic violations broken down by race.

As to the first type of data, while numerous state level studies have utilized criminal involvement benchmarks for both traffic and pedestrian stops, research generally considers the comparison between general criminal offending and traffic stops to be an incompatible and inappropriate (Alpert et al. 2004). The second type of data, which focuses on driving violation patterns by race, is more promising than the first. As described in the introduction, one possible explanation for disparate stop rates could be that the prevalence of traffic infractions could vary by race. If members of one racial demographic group speed or violate other traffic laws more often than members of other groups, then it would be unsurprising if they stopped at higher rates by law enforcement. Traffic violation data is collected both through direct observation and by using technologies such as red-light cameras and photo enforcement. Direct observations are collected using several approaches, including the use of stationary observation points as well as highway driving at set speeds to determine the share of drivers exceeding the posted speed limit (Lange, Blackman, and Johnson 2001). At times, direct observations are combined with camera

[4] Annual as well as quarterly reports compiled by the Portland Police Bureau are available at https://www.portlandoregon.gov/police/65520.

technologies, as in Lamberth (2006) where stationary observers were used to record driver race when speed cameras detected a traffic violation. Still other studies, including Lamberth (2006), utilized cameras alone, using pictures of drivers to identify their race, and citations have been linked with other sources of driver data to determine race when no picture was available.

While promising conceptually, to date, research examining traffic violation patterns by race are equivocal at best. First, research reports that nearly all motorists commit traffic violations. For example, studies have reported that between 93 and 98 percent of motorists engage in speeding. Furthermore, law enforcement officers can identify an infraction of some type for 94 percent of drivers they observe (Lamberth 2006). Second, while some research conducted by law enforcement reports that African Americans engage in speeding at rates higher than whites (Lamberth 2006), other studies have reported no racial difference (Alpert et al. 2004; Epp, Maynard-Moody, Haider-Markel 2014), racial differences at certain speeds but not at others (Lange et al. 2001), or that purported racial differences are likely attributable to age (Lamberth 2006).

2.1.6. Internal Law Enforcement Agency Benchmarks

A more recent, if somewhat controversial, approach to studying traffic stop patterns is to examine officer behavior against law enforcement agency level internal benchmarks. This approach compares individual offers within a single agency with colleagues patrolling in similar areas at similar times to determine whether variation in stop rates at the officer level may account for agency-level racial disparities in stop rates. To accomplish this, researchers match officers within an agency based on patrol area, shift, and other metrics such that, theoretically, expected patterns of stops by race across the matched group should be identical. Where significant differences exist between the group distribution and an officer, there is the possibility that the officer in question is stopping members of a racial group disproportionately.

The primary advantage to the use of an internal benchmark is that it is the only approach that disaggregates agency-level stop data, thus avoiding an analysis focusing solely on the agency as a whole. This can be beneficial, because most current methodologies risk overgeneralizing the extent to which racial disparities exist within an entire agency. For example, all of the other approaches described in this brief lead to the same general result—either the analysis indicates that an *agency* may be operating in a manner than produces racial disparities or it does not. Research conducted by Gonclaves and Melo (2017), however, found that approximately 20 percent of officers in the State of Florida accounted for nearly 100 percent of racially disparate traffic stops. Thus, while most benchmark approaches can merely identify a law enforcement agency as an organization possibly engaging in racially disparate behavior, internal benchmarks can specify which *officer(s)* may be engaging in such behavior. This avoids painting an entire organization in a possibly negative light and can result in more targeted interventions.

In spite of this benefit, there are also concerns about using internal benchmarks. First, similar to other approaches, internal benchmarks cannot prove that racial profiling has occurred. Thus, while differences may be found between an officer and his or her comparison group, it is impossible to prove whether these differences are attributable to racial profiling or racial animus. This is concerning because it is possible that the behaviors of an officer could be misattributed to racial profiling. Second, this approach assumes that average stop rates across similar officers are made in a race neutral way. If officers across a matched group stop members of a single race at a particularly high or low rate, however, this assumption would likely not hold. Third, internal benchmarking likely brings discomfort to members of law enforcement, as it could be used to single out officers with stop patterns that fall outside of their comparison groups. To combat this possibility, researchers conducting internal benchmark analyses are normally blind to officers' identities in the data, only working with numeric identifiers for officers that do not provide enough information for researchers to determine the actual identities of officers in the data. At the conclusion of the analysis, any individual deemed to be different from his or her comparison group could only be identified by that individual's department, where, if the department chooses to proceed with any internal investigation, the officer can be provided with full due process.

2.1.7. The Veil of Darkness Model

As discussed throughout the previous section, benchmarks abound, yet all benchmarks have drawbacks that make their use impractical and/or inadvisable. In an attempt to avoid the issues related to benchmarks, researchers have increasingly sought out methods that avoid their use altogether. The most promising approach was developed by Grogger and Ridgeway (2006), who proposed a natural experiment that they called the "Veil of Darkness" (VOD) method. The VOD method takes advantage of natural variation in daylight and changes associated with daylight savings time (DST) to estimate the racial distribution of drivers at risk of being stopped.

In its simplest form, a VOD approach can compare the racial distribution of traffic stops made during the day to stops made at night, premised on the fact that race is more easily discerned during the daylight than at night. Indeed, Lamberth (2005) reported that while driver race could be determined in 95 percent of cases during the daylight hours (when race was assessed as a binary white-black determination), the determination of race at night required auxiliary lighting. Further, Greenwald (2001) reported that race could only be determined for 6 percent of drivers around dusk, which indicates that determining race at night is nearly impossible in many cases. According to the VOD framework, therefore, if police officers are relying on racially motivated reasons for stopping motorists, it should only be possible to do so in the daylight when a drivers' race is visible. Thus, in this simple case, if officers stop a higher proportion of non-white drivers in the day time compared to after dark, such a result can be taken as an indication that there is a racial disparity.

As described by Ridgeway (2009), however, simply comparing stop rates at night to rates during the day cannot account for possible confounders that may lead to differences in the racial distribution of drivers at different points over a 24-hour period. Hypothetically, if more African Americans drive at night compared to whites, then a direct comparison between these two times would not be appropriate because racial distributions of the driving population would be different. To account for this possibility, Ridgeway proposed using a natural experiment relying on the shift to and from daylight savings time (DST) because it permits researchers to compare stop rates at the same time of day but in different lighting, which allows the researcher to assume that the distribution of drivers will be comparable. Thus, the share of African American drivers at 6:00pm during the week before DST should be consistent with the share of drivers at the same time one week after DST—all that should change is the lighting and, consequently, the ability of a law enforcement officer to detect race when pulling a motorist over.

The chief benefit of the VOD approach is the fact that researchers do not need a benchmark. This neutralizes the need to rely on flawed proxy measures of the driving population as well as the need to spend time collecting benchmark data and constructing comparisons. Second, the VOD model can also simultaneously address several other possible confounding variables, such as clock time, day of the week, and stop location. The inclusion of these other factors is possible because the VOD approach utilizes a multivariate regression framework, which is a type of statistical model that allows a researcher to include multiple predictor variables in a single model. This is unlike any of the benchmark approaches outlined above.

The VOD approach is considered to be the method that has come the closest to achieving a gold standard in traffic stop research. It is a widely accepted technique that does not suffer from benchmarking issues, and when it is deployed via a multivariate analysis, it provides a very strong test for racial disparities (Fazzalaro and Barone 2014). In spite of this status, however, recent research has called one of its most important assumptions into question—that driver behavior does not change due to changes in lighting. In fact, Kalinowski, Ross, and Ross (2017) do precisely that, and argue that driving behavior differs by race in the daylight versus the night. Thus, according to their study, African American motorists increase their speeds after dark due to the relative anonymity that darkness provides, and it is this change in behavior that leads to the findings reported in many VOD analyses that race is not a contributing factor in stop rates.

2.2 Post-Stop Outcomes

Unlike examinations of the decision to stop a driver, researchers studying post-stop outcomes, such as the prevalence of searches, citations, and arrests by racial group, do not face as many methodological challenges. The chief factor leading to more straightforward analyses is the fact that the benchmark issue is no longer a concern, because researcher have access to data regarding those individuals who both were and were not searched, cited, or arrested. Thus, it is relatively easy to construct the at risk group and compare it to the group who encountered one of these interventions. For example, if 10 percent of all stops involve Asian drivers, then it would be reasonable to assume that 10 percent of searches should involve Asian drivers, if searches occur in a race neutral manner.

While the presence of better benchmark data can lead to a more straightforward analysis of post stop outcomes, other concerns still present themselves in studies examining these outcomes. Even when researchers find an acceptable benchmark to compare individuals searched or arrested by the police to the pool of at risk stopped individuals, the use of simple descriptive or bivariate statistics is insufficient to establish a link between intervention patterns and their likely causes, because these basic approaches cannot account for the litany of possible confounding factors that could lead to disparities in the data. To address this shortcoming, researchers utilize several approaches, the most basic of which is to estimate models of post-stop outcomes using traditional logistic regression techniques or models designed to examine count data, including Poisson and negative binomial regression. The advantage of these models is that they are relatively easy to estimate and interpret, and they provide a means for isolating the effects of covariates (e.g., race, gender, age) while controlling for other factors. This means that the researcher can specify, net of the other factors found in the model, the statistical effect attributable to each variable of interest, which in studies of officer-initiated stops is nearly always race. In spite of the benefits of logistic, Poisson, or negative binomial regression, some researchers argue that additional efforts must be made to ensure that disparities identified in the data are robust.

2.2.1. Propensity Score Matching

In spite of the ability to control for other possible confounding factors, traditional multivariate models, including logistic, Poisson, or negative binomial regression, suffer from some serious drawbacks. As described by Ridgeway (2006), it is impossible to determine whether the results of a traditional multivariate model sufficiently account for potential confounders. This is driven by that fact that the results of multivariate approaches are often sensitive to model form and may be significantly influenced by the inclusion (or not) of interaction terms and non-linear data transformations. According to Ridgeway, propensity score matching (PSM) models present an attractive alternative to the use of traditional multivariate modeling techniques for analyzing post-traffic stop outcomes. PSM models allow researchers to compare stops in a target group to a similarly situated comparison group. This quasi-experimental technique thus allows for an analysis of two groups of drivers who exhibit the same distribution of observed stop features (e.g., time of day, location, reason for being stopped) and, by design, will only differ by race. The primary benefit of this approach is that it allows the researcher to control for several important confounding factors, such as the fact that minorities may be more likely to live in areas with a greater police presence.

While PSM presents a very attractive alternative to traditional multivariate analysis, it too has some drawbacks and limitations. One potential drawback is data related, as the creation of the matched groups for comparison is limited to the observed data recorded during a stop. This means that if other unobserved factors or differences exist between the group of interest and the control group, it is possible that the race estimates could be affected (Ridgeway 2006). Second, as described by King and Nielson (2016), the PSM approach, if conducted haphazardly, can do more harm than good. Finally, the PSM is more difficult to model than traditional multivariate techniques and certainly more difficult to estimate and interpret than descriptive statistics based on benchmarks.

2.2.2. Outcome Tests

Outcome tests represent another recent development in officer-initiated post-stop research. This class of methods encompasses several related approaches, including the innovative *threshold test* created by the Stanford Open Policing Project in 2017 and the related, slightly older *Knowles, Persico and Todd* (*KPT) hit rate model*. While there are important differences to these approaches, all outcome tests purport to account for differences in offending behavior across racial groups. Specifically, outcome tests seek to determine whether there is purposeful discrimination on the part of officers causing police search (or citation, or arrest) behavior rather than mere statistical discrimination tied to offending patterns (Engle 2008). Thus, outcome tests seek to determine whether, when faced with a decision to search a minority versus non-minority driver, an officer is more likely to search the minority driver even if the likelihood of a successful search is lower than it would be for a non-minority individual.

Outcome tests have a long history, as they were first used over fifty years ago to study racial differences in the granting of loans by banks (Becker 1957). Becker argued that if banks granted loans to minority applicants in a race neutral fashion, then it would be logical to assume that successful outcomes—measured by repayment rates—would be identical across different racial groups. If, however, minorities repaid their loans at a higher rate than whites did, it would suggest that lenders were not funding loans in a race neutral way, as minority applicants would then be held to a higher standard than non-minorities. When applied to traffic stops, if police officers search minorities and non-minorities in a race neutral way, then the discovery of contraband (i.e., a successful search outcome) should be nearly identical between these two groups. If, however, the discovery of contraband is substantially lower in one group compared to another, then it is possible to infer that the group is being searched more often than warranted because officers are applying a different, lower threshold for searching members of one group compared to another (Tiller, Engle, and Cherkauskas 2010).

Outcome tests examining post-stop outcomes have become very popular among researchers, in the judiciary during racial profiling lawsuits, and among law enforcement professionals. These approaches are touted as clear-cut statistical solutions that provide easily interpretable results. Further, proponents argue that they do not suffer from many of the drawbacks that plague traditional multivariate models (discussed above) and they purport to separate out intentional versus statistical discrimination.

While outcome tests are widely accepted by many groups and stakeholders concerned about racial discrimination in officer-initiated post-stop analyses, there are also some serious critiques. First, while the results are easily interpreted, the statistical model itself is complex and difficult to explain to laypersons. Second, the assumptions underlying the various types of outcome tests are not without criticism from researchers and scholars. As discussed by Engel (2008), one of the implicit assumptions underlying outcome tests is that law enforcement officers have near perfect knowledge or foresight of criminality when making search decisions. Put another way, it is assumed that officers know that members of a certain group possess contraband at a specific rate, which justifies searching members of that group. Relatedly, these tests also assume that all officers will respond similarly to the same contextual information. Both of these assumptions, according to Engle (2008) are highly questionable. The first assumption forms the foundation of much of economic theorizing and has often been criticized by scholars from various disciplines as being unrealistic. The second assumption has been called into question by research that reports that differences in demographic characteristics, experiences and training, work assignments, management and oversight, and personal attitudes can affect officers' behaviors and decision-making (see Engle 2008: 24-25 for more than a dozen studies reporting these findings).

Finally, studies utilizing the myriad variants of the outcome test also often assume that when it comes to search decisions, officers have complete discretion and that the decision to search a stopped individual only focuses on discovering contraband. In reality, however, discretion varies significantly. Some searches are mandatory under agency policies, whether they are searches conducted during an arrest or administrative searches when vehicles are impounded.[5] Further, even among discretionary searches, there are different levels of discretion applied. If an officer views contraband in plain sight upon stopping a car, or smells the odor of drugs, the amount of discretion to search is different compared to searches based on reasonable suspicion. Finally, consent searches also present problems, as at least part of the decision to search is in the hands of the motorist and thus outside of the officer's discretion. In fact, the inclusion of consent searches has been deemed so problematic by some researchers it has been suggested that only non-consent discretionary searches can be included in outcome tests.

3. Developing a Research Approach for Analyzing Oregon STOP Data

Due to the complexities involved in a single officer-citizen interaction, the great degree of variation in techniques and approaches used to analyze traffic stops, and the fact that each approach to analyzing this type of data has its own strengths and weaknesses, the selection of an appropriate analytical technique is challenging. Current best practice is to utilize a suite of approaches to address these issues. By utilizing multiple approaches, researchers are able to consider all aspects of a stop, from the initiation of the encounter through any searches, citations, or arrests that occur. Second, in the event the data indicates an agency may have a racial stop discrepancy, the examination of multiple data points helps to ensure that

[5] The issue of whether to include mandatory searches in analyses utilizing outcome tests is further complicated by the fact that departments vary in both their policies regarding mandatory searches as well as the frequency with which they conduct mandatory searches (Engel et al. 2005).

any analytical findings are more robust than could be achieved using a single test. Third, this has been the approach taken by the state of Connecticut, which is relevant given that the Connecticut model informed the development of the STOP program and HB 2355 in Oregon. Finally, recent studies of traffic stops in Oregon have utilized multiple analyses when examining encounters between citizens and law enforcement, including benchmarking research conducted for the Portland Police Bureau by the Criminal Justice Policy Research Institute at Portland State University (Renauer et al. 2009) as well as internal studies of police stops conducted by the Portland Police Bureau between 2013 and 2015[6] and a study examining nine years of stop data for the Corvallis Police Department (Renauer 2016).

As such, the CJC has elected to follow in the spirit of the approach taken by the State of Connecticut and the IMRP Project Team housed at Central Connecticut State University, which has employed an analytical matrix in its analysis of traffic stop data since 2014. Following the Connecticut research group, the CJC identified several guiding principles to guide the selection of the best suite of approaches for inclusion in analyses of Oregon STOP data. Techniques utilized by the CJC:

1. Must be ***evidence-based*** and ***conform to best practices*** approaches for the analysis of stops data.

2. Must provide findings that are widely ***accessible***, insofar as the selected approaches should occupy a middle ground between robust but complex statistical models with difficult to interpret and communicate findings and simpler models that, while easy to communicate, possess numerous theoretical and methodical shortcomings.

3. Must use data that are reasonably ***easy to obtain and update*** as measured by:

    a. Ease of ***obtaining***, downloading and working with data obtained from other sources (e.g., Census, other Agency or State-level data.);

    b. The ***costs*** associated with obtaining data from third parties or collecting data by the CJC; and

    c. The frequency with which data is ***updated*** by third parties or needs to be updated by the CJC.

4. Must be selected by considering each technique's ***methodological strengths and weaknesses*** identified by researchers and practice.

3.1 The Initial Decision to Stop a Citizen

As described above, there are two general approaches to studying the decision made by law enforcement to initiate a stop: those that utilize a benchmark and the veil of darkness model. As discussed at length, all benchmark approaches have significant strengths and weaknesses. Table 1 presents these strengths and weaknesses using the goals listed above as a guide. The primary strength of nearly all benchmark approaches is the accessibility of their results. For stakeholders, the comparison between the share of stops for a single demographic group and the percentage of that demographic group found in the

[6] https://www.portlandoregon.gov/police/72040

population is quite intuitive and easy to understand. Comparing across benchmarks, there is some variability regarding the degree to which the benchmark makes accessible comparisons, as the construction of some benchmarks are more complex than others. For example, while the residential population is straightforward and easy to explain, the Census driving population approach discussed above presents a much more complex approach which rests on a number of assumptions. Similarly, traffic violation data, given that it is the only benchmark that does not rely on a more traditional view of the driving population, would also be less accessible to stakeholders. The accessibility of benchmarks is not only in the simplicity of their construction (for most approaches), but also in the fact that they can be used to quickly and succinctly make comparisons across a wide variety of units. For example, a statewide comparison of stopped drivers to a selected benchmark is just as easy to construct, display, and digest as a table of comparisons for individual departments.

Table 1. Assessment of Reviewed Benchmarking Approaches

Benchmark Technique	Accessibility of Findings	Obtaining Data	Data Costs	Data Updating	Other Weaknesses
Census Residential	●	●	●	◐	○
Census Driving Population	◐	◐	●	◐	◐
Licensed Driver Data	●	○	◐	◐	◐
Roadway Observations	●	○	○	○	◐
Traffic Accident Data	●	○	○	○	●
Traffic Violation Data	◐	○	○	○	○
Internal Benchmark Data[7]	◐	○	○	○	○

Note: Rankings on a three-point scale from low (○), to (◐) medium, to high (●).

While most benchmarks score relatively high on accessibility for stakeholders, other areas of assessment are much more variable. This is certainly true concerning the other goals outlined above concerning data accessibility, cost, and updating. Census residential data are the easiest to obtain. Data is available online at https://factfinder.census.gov, which is a free, relatively easy to use web platform for selecting and downloading Census data. Indeed, any researcher who spends more than an hour on the site is likely to become competent in its use. Further, the data on the Census website are updated frequently and periodically.

Beyond the ease with which researchers can obtain and update Census data, however, there are significant weaknesses associated with the use of Census residential data. As described above, the residential population of an area may not be an accurate reflection of the driving population. In Portland, for example, only 42.6 percent of the daily working population consists of Portland residents, while the remaining 57.4 percent, which accounts for over 235,000 individuals, comes from surrounding cities and communities that would not be measured if simple population demographics were used to construct a benchmark (US Census Bureau 2018). Similarly, only 50.5 percent of workers in Salem live in Salem (or the connected areas of Keizer, Hayesville CDP, and Four Corners CDP). Beyond work commuters, tourism likely also has a significant impact on driving patterns in the state. Indeed, estimates of the non-resident tourist population in several regions of Oregon demonstrate that non-residents can make up between 4 and 27 percent of an area's driving population at different times throughout the year (Dean

[7] The collection of data identifying officers is explicitly prohibited by HB 2355. Thus, the use of any internal benchmarking approach in Oregon using data collected pursuant to the STOP program would be impossible.

Runyan Associates 2017). Thus, reliance on residential population numbers would be highly suspect simply due to work commuting patterns and tourism within many of the larger cities of Oregon.

Finally, reliance on Census data of any kind can present several challenges. The foremost concern is accuracy, as decennial Census data will not be available until after 2020, which will be several years after data collection and analysis of officer-initiated stops in Oregon has begun. In the interim, the only data available will be the yearly population estimates and ACS data. While the yearly population estimates are the most accurate non-decennial Census data (Rynerson 2018), they are limited to state-level and county-level estimates, which only accounts for 42 percent of Tier 1 reporting agencies and 30 percent of Tier 2 reporting agencies, all of which will be reporting before the next decennial census occurs.

The remaining 37 non-county Tier 1 and 2 reporting agencies represent cities and towns, which can only be analyzed using ACS data. The primary drawback to this data is that both the one-year and five-year calculations rely on sampling, and thus have sampling error. According to the Population Research Center at Portland State University, for example, when estimating the Hispanic population for Medford, the one-year ACS provides a point estimate of 13.1 percent, with a 90 percent confidence interval of +/- 3.9 percent (a range of 9.2 to 17 percent), while the five-year ACS data is more precise, as it provides a point estimate of 14.6 percent, with a 90 percent confidence interval of +/- 1.1 percent (a range of 13.5 to 15.7 percent) (Rynerson 2018). Further, internal CJC analyses comparing five-year ACS data to yearly population estimates shows significant disparities throughout Oregon, particularly in small counties, where estimates of individual racial groups can differ by as much as 50 percent. Thus, when using data of this type to determine whether racial disparities exist, the lack of precision in the ACS estimates may make it difficult to draw adequate conclusions regarding the presence of a disparity, particularly in Oregon where the size of some minority groups is rather small and fluctuates significantly depending on the source consulted.

Estimates of the Census driving population present similar benefits and challenges. While much of the component data for estimating driving populations are also accessible via the Census website and other related sites, such as https://onthemap.ces.census.gov, the construction of this benchmark requires the researcher to have some experience working with spatial data, other GIS related files, and related approaches. Costs, similar to working with residential data, are low other than accounting for the staff time that must be dedicated to constructing the benchmark. Updating falls in the middle, as the reconstruction of the driving population benchmarks can be an involved process, although if established data scripts and programs are used, updating can be streamlined. Finally, the issues associated with the use of Census data are compounded when using it as component data for the construction of the residential driving population. For example, 205 communities contribute at least ten workers to the working population in Salem (US Census Bureau 2018). If there is error traceable to demographic estimates from each of these areas, then the compounded error resulting as the researcher aggregates each area's contribution to the Salem working population could rise to an unacceptable level.

Licensed driver data, traffic accident data, and traffic violation data all require the researcher to obtain data from other agencies, such as the state Department of Motor Vehicles (DMV), Department of Transportation (DOT), or individual law enforcement agencies. This can present challenges when agencies are not willing to share data or when complex interagency data sharing agreements must be negotiated. Costs for these data can be higher than other sources as well, as some agencies charge flat fees for data requests or an hourly rate when compiling data for other organizations. These challenges are multiplied if regular updating is required. Finally, at least in Oregon, while it may be possible to obtain licensed driver data from the DMV or accident data from ODOT, this data does not contain the most important variable necessary for these analyses: race. Oregon DMV data does not contain a value for driver race. Similarly, while ODOT publishes yearly crash numbers by county and for municipalities with populations over ten thousand people, it does not report the racial demographics for accidents.

Roadway observations have the steepest data acquisition costs, as field studies must be deployed to collect the data directly. As described previously, this has proven to be costly, and other states have determined that this approach is simply not feasible for statewide programs aimed at analyzing officer-initiated stops due to the logistical issues involved and monetary costs. This issue also raises concerns for keeping data up to date, as roadway observation studies can quickly become outdated as areas grow and traffic patterns shift. These issues are particularly concerning and relevant to Oregon. First, the feasibility of collecting data has been called into question in small states, such as Connecticut. In Oregon, the resources needed to conduct observational studies would be staggering in comparison, as Oregon is approximately eighteen times larger than Connecticut geographically. Second, the need to update roadway data would be particularly acute in Oregon, as it is one of the fastest growing states in the U.S. (Leins 2017), having grown by nearly one million people (approximately 25 percent) in the past two decades. Relatedly, Oregon is also experiencing significant demographic shifts, as the population growth of Hispanics has outpaced this group's growth in much of the U.S. Since 2000, for example, while Hispanic population growth has been approximately 50 percent nationwide, Oregon has seen a 72 percent increase (Parks 2016). Taken together, these patterns suggest that roadway observational studies would be expensive to implement and require frequent updates, which would further compound cost issues.

Roadway observation data would also likely suffer from inaccuracies in Oregon. As described above, the determination of race for this type of data relies on observations of individuals' racial identities by third parties. Research demonstrates that even when the target and assessor are in close proximity to one another, even just a few feet during an interview, racial identifications are often inaccurate (Campbell and Troyer 2007). If an interviewer in the same room with an individual cannot reliable identify a subjects' race, identifications by stationary observers of drivers in fast moving vehicles are likely be very inaccurate.[8] This limitation is particularly notable for Hispanic individuals, as research demonstrates that the identification of Hispanics is difficult for both other Hispanics and non-Hispanics alike (Alpert et al. 2004). Furthermore, prior studies utilizing roadway observation technique found that in addition to the distance between the target and the observer, identification becomes increasingly difficult as driver speed increases (Lambreth 2006). Based on the foregoing, Table 1 provides a comparison of the six different benchmarking approaches discussed in this analysis. As shown in Table 1, while census residential data earns high marks for accessibility and on the three measures of data availability, it is the weakest when it comes to other methodological concerns. The other five approaches present various concerns, with traffic violation data presenting the most serious issues. Indeed, following the investigation into these approaches, it would appear that perhaps the most prudent course of action would be avoiding the use of benchmarks all together. Taking this approach, however, ignores the most important strength of benchmarks, which is the accessibility of their findings for stakeholders. Thus, while these approaches have numerous drawbacks, as an initial descriptive analysis of officer-initiated stops in Oregon, the CJC believe that the use of benchmarks would be beneficial.

[8] Lambreth, in his 2006 study, describes how attempts at observing race from a stationary point were quite difficult on a roadway where automobiles traveling 56 mph triggered a citation and thus an observation by researchers.

Therefore, if a single benchmark approach is to be used (or a suite of benchmarks), it is necessary to identify the strongest one(s) for inclusion in the analysis of Oregon officer-initiated stop data. As discussed, Census residential data is the easiest to interpret, obtain, and update, but it comes with numerous methodological concerns. CJC believes, however, that some of these concerns can be overcome. First, given that one data element included in the data collection effort in Oregon is information on the residency of the individual who is stopped by the police,[9] it would be possible to include benchmark comparisons between Census residential information and data for stops limited to citizens of the area being analyzed. Second, concerning data accuracy, for the state and for counties the more accurate one-year estimates can be used so that only cities and towns will utilize the less accurate ACS data. Further, once data collection has been fully implemented across all law enforcement tiers in 2021, each Oregon law enforcement agency will have reported a full year of data, prior demographic comparisons using yearly estimates and ACS data can be compared with the results of the 2020 Census, which will provide the most accurate benchmark data possible.

Relatedly, following the best practices of Connecticut (Fazzalaro and Barone 2014), CJC will also build and pilot residential driving samples following the method prescribed by the Connecticut research group. At this time, however, CJC remains uncommitted as to the inclusion of these results in the final analyses, as it requires time to investigate the degree to which the approach utilized by Fazzalaro and Barone (2014) approximates the driving population of different Oregon communities given the sampling error found in its component population estimates.

Beyond the ability of the CJC to address some of the shortcomings of Census residential data or driving population data, the use of these approaches are justified for other reasons as well. Chief among them is legitimacy, as layperson stakeholders will likely expect analyses of possible racial disparities utilizing population benchmarks. Given that this has been the traditional approach and the fact that it is easy to understand and its conclusions are accessible, to leave the benchmark approach out of the analyses would be risky and could be viewed as a means for criticizing the conclusions. If, on the other hand, benchmarks are used descriptively as a means for providing an initial reference point, these data points could be used to set the stage for the more complex analyses that follow.

Due to the drawbacks identified for all benchmarks, the primary analytical approach that will be utilized to study the initial decision to stop an individual will be the Veil of Darkness model. As discussed above, the VOD model takes advantage of natural variations in lighting to test whether stop rates differ during the day versus after dark or during different periods surrounding shifts to and from daylight savings time. The beauty of this approach is that there is no need for benchmarks. Further, it is easy to address different lighting conditions that may be influenced across different latitudes and longitudes across the state, as well as across different times of year when days are longer or shorter, as eligible stops can be coded to account for these differences before estimating the models. In addition, other predictors can be included in these models, which provides the opportunity to further isolate race as a predictor of stop rates. Finally, while the concerns raised by Kalinowski, Ross, and Ross (2017) may be valid, the VOD still remains the most highly respected, tested approach to analyzing racial disparity patterns during the initial stop window (Fazzalaro and Barone 2014). Thus, while a benchmark approach will be used to start the

[9] HB 2355 explicitly mandates that law enforcement agencies collect data on a number of factors concerning the stop as well as specific demographic characteristics of the stopped individual. Residency status, however, was not included in HB 2355. In discussions leading up to the collection of data in Oregon, however, the stakeholder groups tasked with implementing the HB 2355 data collection elected to include residency status as one of the data elements.

conversation about possible disparities in the stop data for initial decisions to stop drivers and/or pedestrians, the VOD approach will provide a more statistically sound, robust test of each submitting law enforcement agency.

3.2 Post-Stop Outcomes

Fortunately, post-stop outcomes do not rely on benchmarks, which tends to make the selection of analyses of the latter stages of an officer-initiated stop easier. Relying on the goals and principles outlined in Section 3, the CJC plans to utilize both propensity score matching techniques as well as at least one variant of the outcome test to examine post-stop outcomes. The CJC selected these approaches because both are viewed by researchers and practitioners as robust, sound tests for examining searches, arrests, and citations. As such, they are evidence based and considered examples of current best practices in the literature. Second, because these approaches do not rely on benchmarks, there is no requirement that CJC obtain outside data from third parties, which means that concerns about obtaining data, data costs, and updating are relatively small and limited to ensuring that data is correctly and consistently submitted to the CJC by law enforcement agencies pursuant to the requirements of HB 2355. Finally, the primary weaknesses of both PSM models and the suite of outcome tests is that they are complex and thus require more time and care to build, test, and refine. This complexity encompasses the concerns regarding the construction of the analytical samples discussed above for outcome tests, and CJC will take care to ensure that different samples will be analyzed and sensitivity tests will be used to examine the results to confirm that the appropriate samples are being utilized.

4. Conclusion

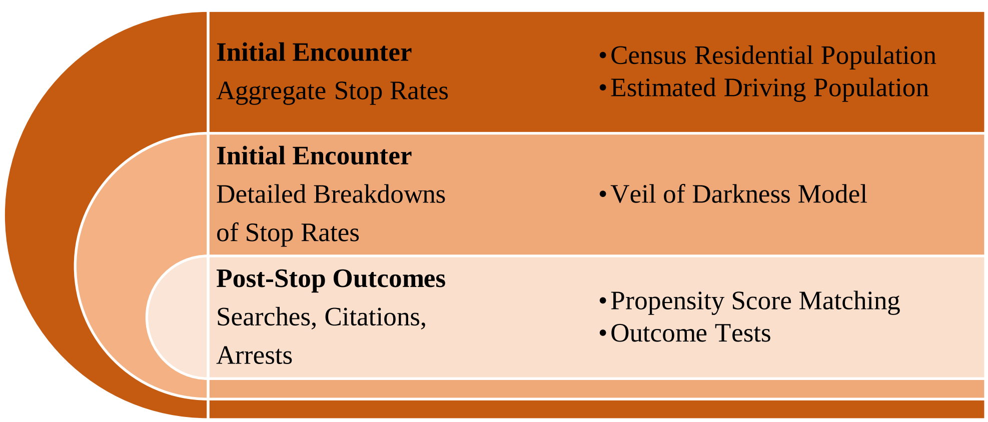

Figure 1 is provided to illustrate the multi-modal approach described in the foregoing sections of this briefing. As discussed, the initial police-citizen encounter will be examined using two classes of statistical approaches, beginning with aggregate comparisons of benchmarks of stop data before moving on to the more complex, but methodologically sound, Veil of Darkness approach. Post-stop outcomes will be examined using propensity score matching models and one or more variants of the outcome test. CJC believes that this combination of approaches best fit the goals and principles identified in Section 3 of this briefing and will result in the most complete analysis of stop data in Oregon.

5. References

Alpert, G.P., Smith, M.R., Dunham, R.G. (2004). "Toward a Better Benchmark: Assessing the Utility of Not-at-Fault Traffic Crash Data Racial Profiling Research." *Justice Research and Policy* 6(1): 43-69.

Becker, G.S. (1957). *The Economics of Discrimination*. University of Chicago Press: Chicago, IL.

Dean Runyan Associates. (2017). Oregon Travel Impacts 1992-2016. Accessed at http://www.deanrunyan.com/doc_library/ORImp.pdf.

Campbell, M.E., Troyer, L. (2007). "The Implications of Racial Misclassification by Observers." *American Sociological Review* 72: 750-65.

Engel, R.S. (2008). "A Critique of the "Outcome Test" in Racial Profiling Research." *Justice Quarterly* 25(1): 1-36.

Engel, R.S., Calnon, J.M. (2004). "Comparing Benchmark Methodologies for Police-Citizen Contacts: Traffic Stop Data Collection for the Pennsylvania State Police." *Police Quarterly* 7(1): 97-125.

Engle, R.S., Calnon, J.M., Tillyer, R., Johnson, R.R., Liu, L., and Wang, X. (2005). *Project on Police-Citizen Contacts, Year 2 Final Report*. Submitted to the Pennsylvania State Police, Harrisburg, PA.

Epp, C.R., Maynard-Moody, S., Haider-Markel, D.P. (2014). *Pulled Over: How Police Stops Define Race and Citizenship*. University of Chicago Press: Chicago, IL.

Fazzalaro, J., Barone, K. (2014). "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.

FHWA NHTS Brief. (2016). "Hispanics and Transportation: 2009 National Household Travel Survey. Accessed at: http://nhts.ornl.gov/briefs/HispanicBrief.pdf.

Greenwald, H.P. (2001). "Final Report: Police Vehicle Stops in Sacramento, California." Sacramento Police Department.

Grogger, J., Ridgeway, G. (2006). "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness." *Journal of the American Statistical Association* 101(475):878-887.

Kalinowski, J., Ross S.L., Ross M.B. (2017). "Endogenous Driving Behavior in Veil of Darkness Tests for Racial Profiling." Working Paper, University of Connecticut Department of Economics Working Paper Series: 2017-03.

King, G., Nielson, R. (2016). "Why Propensity Scores Should Not Be Used for Matching." Working Paper, Institute for Quantitative Social Science, Harvard University. https://gking.harvard.edu/files/gking/files/psnot.pdf

Lamberth, J.C. (2005). “Racial Profiling Data Analysis Study: Final Report for the San Antonio Police Department.” Lamberth Consulting.

Lamberth, J.C. (2006). “Data Collection and Benchmarking of the Bias Policing Project.” Final Report for the Metropolitan Police Department in the District of Columbia.

Lange, J.E., Blackman, K.O., Johnson, M.B. (2001). “Speed Violation Survey of the New Jersey Turnpike: Final Report.” Calverton, MD: Public Services Research Institute.

Leins, C. (2017). “These Are the 10 Fastest-Growing States in the U.S.” *U.S. News & World Report*, Dec. 20, 2017 (https://www.usnews.com/news/best-states/slideshows/these-10-states-saw-the-greatest-population-growth-this-year)

Lovrich, N., Gaffney, M., Mosher, C., Pratt, T., Pickerill, M. (2007). Report to the Washington State Patrol, Results of the Monitoring of WSP Traffic Stops for Biased Policing: Analysis of WSP Stop, Citation, Search and Use of Force Data and Results of the Use of Observational Studies for Denominator Assessment. Washington State University. Division of Governmental Studies and Services.

Parks, C. (2016). “Oregon’s Latino population is growing faster than the nationwide average.” *The Oregonian*, Sept. 15, 2016 (http://www.oregonlive.com/pacific-northwest-news/index.ssf/2016/09/oregons_latino_population_is_g.html)

Renauer, B.C. (2016). “Analysis of Corvallis PD Stops and Searches 2007 to 2015.” Submitted to Corvallis Police Department, September 22, 2016.

Renauer, B.C., Henning, C., Covelli, E. (2009). “Benchmarking Portland Police Bureau Traffic Stop and Search Data. Report submitted to Portland Police Bureau, September 23, 2009.

Ridgeway, R. (2009). *Cincinnati Police Department Traffic Stops: Applying RAND’s Framework to Analyze Racial Disparities*. Rand Corporation: Santa Monica, CA.

Rynerson, C. (2018). “Race and Ethnicity in the Census: HB2355 Stakeholder Engagement Meeting.” Presented on January 8, 2018.

Tiller, R., Engle, R.S., Cherkauskas, J.C. (2010). “Best Practices in Vehicle Stop Data Collection and Analysis.” *Policing: An International Journal of Police Strategies & Management* 33(1):69-92.

U.S. Census Bureau. (2018). LEHD Origin-Destination Employment Statistics (2002-2015). Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program, accessed on 1/3/18 at https://onthemap.ces.census.gov. LODES 7.3