The University of Texas at San Antonio
Department of Criminology & Criminal Justice

Traffic Stop and Arrest Analysis

Final Report

February 2021

Rob Tillyer, Ph.D.
Michael R. Smith, J.D., Ph.D.
University of Texas at San Antonio

This research report was supported through an agreement between the City of Livermore, CA. and the University of Texas at San Antonio. The findings and recommendations presented within this report are those of the authors and do not necessarily reflect the official positions or opinions of the City of Livermore, CA. or the Livermore Police Department. The authors wish to thank Chief Young, Lt. Reynolds, and members of the Livermore Police Department who provided data and assisted with the project. **Please direct all correspondence regarding this report to: Dr. Rob Tillyer, Department of Criminology & Criminal Justice, the University of Texas at San Antonio, rob.tillyer@utsa.edu.**

Table of Contents

Tables & Figures

Executive Summary

A research team of criminologists from the University of Texas at San Antonio was contracted to analyze recent stops by the Livermore Police Department (LPD) for patterns of racial and/or ethnic disparity. The analysis addressed two areas of possible disparity: (1) disparities in traffic stops and (2) disparities in arrests.

- Methodology: Examination of 24,944 encounters between LPD officers and civilians between January 1, 2019 and April 30, 2021.
- Traffic Stop Analyses: Use of two benchmarks to operate as proxies for driving and traffic law violating populations in Livermore to compare against the racial/ethnic composition of drivers stopped by the LPD
    - Benchmark #1: A ”veil of darkness” (VOD) analysis examined differences in stop rates of non-White and White drivers during the daytime compared to the nighttime.
    - Benchmark #2: Data containing the racial composition of not-at-fault and at-fault drivers involved in two-vehicle crashes.
- Arrest Analyses: A multivariate model examined whether civilian race/ethnicity predicted the likelihood of an arrest by the LPD after controlling for other relevant factors.
- Key Findings
    - The results from the two benchmark analyses did not reveal a pattern of racial/ethnic disparity in traffic stops experienced by non-White drivers in Livermore.
    - The VOD analysis found no statistically significant differences in the rates at which non-White drivers were stopped in Livermore during the day compared to at night.
    - The traffic crash benchmark analysis found slightly elevated risks for stops of White, Black, and “Other” race drivers (at-fault benchmark only) and a slightly elevated risk for stops of Black drivers relative to White drivers (not-at-fault benchmark only).
    - Taken together, the two benchmark analyses do not show a consistent pattern of disparity in stops based on driver race or ethnicity.
    - The results of the arrest analyses do not reveal a pattern of racial disparity in police outcomes that disadvantages non-White civilians in the City of Livermore.

In sum, no clearly identifiable or concerning pattern of racial/ethnic disparity was found in the 24,944 encounters that took place in Livermore over a 26-month period during 2019-21.

I. Project Background

Researchers at the University of Texas at San Antonio (UTSA) began conversations with the City of Livermore, California and the Livermore Police Department (LPD) in late 2020 to discuss a potential research collaboration. Drs. Rob Tillyer and Michael Smith (UTSA) developed a Scope of Work that was negotiated with Chief Young (LPD) to assist with research questions requiring data analysis and assessment. These discussions culminated in a signed contract to engage with the City on the following matters:

1. Consult with selected City Council and community members to increase awareness of traffic stop data analysis complexities, including benchmarking
2. Review and assess current traffic stop data collection protocols; provide recommendations as needed
3. Analyze recent traffic stop data to identify the racial/ethnic composition of those encounters and compare against appropriate benchmarks
4. Analyze recent LPD arrest data to identify factors associated with arrest

This report addresses Tasks #3 & 4 – assess traffic stops to summarize the racial/ethnic composition of those encounters in relation to appropriate benchmarks and analyze recent LPD data to identify factors associated with arrest. Data required to complete these tasks was provided to the research team during the Summer and Fall of 2021. Data cleaning, variable creation, data analysis, and report writing was undertaken throughout this time period.

This report contains several sections that summarize:

- The scientific knowledge and best practices in traffic stop and arrest data analyses (Section II)
- The methodology applied to the LPD data (Section III)
- The primary findings from the analyses of these data (Section IV)
- The meaning of these findings for the LPD and the City of Livermore, CA.
- Recommendations on next steps (Section V)

II. Officer Decision-Making

Given the focus of this report on traffic stops and arrests, it is important to locate the current analyses within the broader landscape of current best practices and empirical evidence on these two decision points.

Traffic Stops

Law enforcement agencies, communities and their residents, and academics (among others) have expressed continuing interest in the traffic stop practices of the police over the past thirty years. A key concern is that minority groups may receive greater attention from the police and experience elevated or disparate rates of stops and post-stop outcomes such as tickets, searches or arrests. In short, there is a concern that persons of color are targeted for enhanced contact by police and more punitive outcomes.

From a scientific perspective, the initial step in understanding the nature of this concern is to empirically evaluate the prevalence of disproportionate contact with non-White groups. If this is established, then the related question of 'why' this pattern of behavior occurs can be addressed. The primary challenge in analyzing the racial/ethnic composition of traffic stops initiated by the police centers on the identification of an appropriate comparison population or benchmark against which to compare the behavior of the police department.

Several scholars have written extensively on the study and evaluation of traffic stops (e.g., Alpert et al., 2004; Fridell, 2004; Ridgeway, 2007; Smith et al., 2021), and the issues can be distilled as follows. The assessment of law enforcement agency traffic stop behavior begins with the calculation of the racial/ethnic composition of traffic stops as represented by a simple rate of stops for each racial/ethnic group. Critically, the goal is to analyze police-civilian contacts that are officer-initiated and not the result of a call for service from the public. If the concern is that officers may be disproportionately stopping non-White drivers, then it is appropriate to assess their independent decision-making to make a traffic stop rather than stops made as the result of a call for service or other community-based request. As a result, using administrative data collected by the police, the total number of officer-initiated traffic stops for each racial/ethnic group is divided by the total number of officer-initiated stops to produce a rate of traffic stops for each racial/ethnic group. Drivers are frequently categorized into White, Black, Hispanic, Asian, and Other groups with the latter representing a broad category that includes Native Americans, Middle Easterners, or other minority groups.

Once stop percentages are calculated for each group, they must be compared to some independent measure of what is to be expected if no bias in officer decision-making existed. Knowing that 20% of a police department's traffic stops involved Black civilians is meaningless unless we also know what percentage of Black drivers are available or at risk of being stopped on the roadways. Thus, an estimated population of those at <u>risk for being stopped</u> must be identified to compare against the stop rates experienced by non-White drivers. In short, in order to understand whether or not there are racial/ethnic disparities in traffic stops, the analyst must identify and apply an appropriate benchmark against which to compare the rate of stops experienced by non-White drivers (see Alpert et al., 2004; Fridell, 2004; Ridgeway, 2007 for further discussion).

<u>Benchmarking</u>

A recently published, peer-reviewed article (Smith et al., 2021) summarizes the current academic efforts to identify an appropriate benchmark and offers the strengths and weaknesses of the most common approaches to addressing this issue. As the authors note, “In order to investigate the existence or magnitude of racial and ethnic disparities in stops of citizens undertaken by the police, it is not sufficient to simply examine the percentage of stops that target minorities. Instead, researchers must determine the extent to which different racial and ethnic groups would be represented in stops if no racial bias was present and then compare the percentage of minority citizens who were stopped to this hypothetical benchmark” (p. 515). While identifying an appropriate benchmark is critical to the assessment of traffic stops, it is without doubt the most challenging and controversial aspect of the effort (Alpert et al., 2004; Ridgeway & MacDonald, 2010; Tillyer et al., 2010). Benchmark strengths and weaknesses are rooted in their ability to accurately offer a proxy for the driving population at risk of being stopped. Importantly, some benchmarks have greater logical and empirical validity than others and represent preferred options for an assessment of possible disparities in the decisions officers make to initiate traffic stops (see Smith et al., 2021 for a more complete discussion).

The most readily accessible benchmark for stops is the Census count of the racial/ethnic composition of the local population. This benchmark is easy to access and presents some initial appeal; however, its utility and appropriateness quickly become problematic once its underlying assumptions are considered. The key challenge to this benchmark, and why it has been rejected by contemporary scholars as a scientifically appropriate benchmark, is rooted in the assumption that each resident or group within a local jurisdiction possesses an equal level of risk for being stopped by the police. This is simply an untenable assumption. Individuals and groups of civilians vary in their likelihood of contact (i.e., risk) based on a number of factors, including their own driving behavior (i.e., if they drive, how they drive, when they drive, what they drive, etc.). Related, the use of this benchmark assumes that only the residential population that lives in a particular area drives in that area. This is an unreasonable assumption as the routine activities of drivers often involve crossing jurisdictional boundaries, such that residents from neighboring cities, counties or even states often cross into the jurisdiction of interest and vice versa. This undermines the accuracy of the residential Census population as a proper representation of the driving population at risk for being stopped by the police in a given city. <u>These two factors (i.e., driving behavior of residents and the cross-jurisdictional travel of non-resident drivers), in addition to others, render this an inappropriate benchmark</u> (for further discussion, please see Alpert et al., 2004; Novak, 2004; Ridgeway & MacDonald, 2010; Tillyer et al., 2010).

Other benchmarking options include the use of red-light cameras to capture the racial/ethnic composition of the driving/violating population or the direct observation of the driving and/or violating populations through the systematic social observation of drivers (see Alpert et al., 2004; Meehan & Ponder, 2002). A separate technique assesses the traffic stop behavior of officers compared to other officers working similar shifts, assignments, and areas. Each of these techniques offer some legitimacy as proxies for risk, but also possess some weaknesses, and their use is often predicated on data availability.

Veil of Darkness Benchmark

Two benchmarks are particularly relevant to this project and offer the most appropriate approach to assess the LPD traffic stop data: (i) the 'veil of darkness' methodology and (ii) the use of traffic crash data. The 'veil of darkness' (VOD) was developed by Grogger & Ridgeway (2006) and is relatively easy and straightforward to apply. This approach makes use of natural changes in lighting based on daylight savings time to allow a comparison of the racial/ethnic composition of vehicle stops made during daylight hours to the racial/ethnic composition of stops made at night during the same hours of the day. Using the sunrise and sunset times published by the Naval Observatory, traffic stops are coded as 'daytime' or 'nighttime' depending on the time of the year. For example, a traffic stop initiated at 7PM in January would be classified as a 'nighttime' stop, while a traffic stop undertaken at 7PM in July would be a 'daytime' stop. Ultimately, this approach focuses on traffic stops that occur in the 'inter-twilight' period, which is defined as the period of time between the earliest sunset (4:48 pm) and the latest sunset (8:33 pm) across the different months of the year (times reflect sunset in Livermore, CA. during the study period).

The use of the 'inter-twilight' period allows a comparison of the racial/ethnic composition of stops during times when daylight could reasonably allow the identification of driver race/ethnicity to the same period of time when darkness would limit the detection of driver race/ethnicity. The underlying logic of the comparison is that the driving patterns of racial/ethnic groups are likely to be similar across the same hours of the day, but make use of daylight savings and seasonal variation in nighttime hours that limit officers' ability to identify driver race/ethnicity prior to a stop. Thus, if officers' decisions to initiate traffic stops were influenced by bias (overt or implicit), then a different racial/ethnic pattern of stops would be evident in the daylight hours compared to nighttime hours (Smith, et al., 2021). Under conditions of bias, daytime stops would reflect a higher proportion of minority drivers when race is more easily identifiable than nighttime stops when skin tone and other features of drivers are more difficult to see.

This approach has been widely replicated in the literature (Chanin, Welsh, Nurge, & Henry, 2016; Pierson et al., 2019; Ritter & Bael, 2009; COPS, 2016; Ross et al., 2016; Smith et al., 2021; Taniguchi et al., 2016; Worden et al., 2012) as it is does not require external data for a benchmark (beyond information supplied from police stop databases themselves). For example, Pierson and colleagues (2019) used the VOD to assess traffic stops initiated by multiple agencies across 21 states and 29 cities. Results indicated that Black drivers were stopped more often during the day and less frequently at night compared to White drivers. Kalinowski and colleagues (2017) also used the VOD methodology to assess stop data from the Massachusetts State Police, Boston, and other Massachusetts municipal agencies with at least 100 speeding stops and an African-American population of 10 percent or higher. They reported that the odds of a Black driver being stopped during daylight hours were 35 to 48 percent higher than for White drivers, depending upon variations in the models. Moreover, they reported Black drivers seemed to adjust their driving behavior (i.e., speeding) downward during the daytime when they would be most visible to the police. As noted by Smith et al (2021), the adjustment in driving behavior of some groups during the daytime identifies a potential weakness in this benchmark that "may overestimate the population of minority traffic violators at night and underestimate minority traffic violators during the day, thus leading to a type II error and a finding of no discrimination in the treatment of minorities in stops by the police" (p. 517).

Crash Data as a Benchmark

The second benchmark used in the analysis of LPD traffic stops uses vehicle crash data. These data offer information on at-fault and not-at-fault drivers involved in vehicle crashes on roadways in Livermore. This approach was pioneered by Alpert and colleagues (2004) and provides an estimate of driving population by using the known race/ethnicity of drivers involved in crashes investigated by the police (also see Lovrich et al., 2007; Withrow & Williams, 2015). Conceptually, these data provide a proxy for the racial/ethnic composition of the driving population, while also accounting for driving frequency and potential exposure to police surveillance (Smith et al., 2021). This is particularly true for not-at-fault drivers who represent a 'random' cross-section of drivers on the roadways that may be at risk for police contact, but also for at-fault drivers who may represent an estimate of traffic violators most likely to draw attention from the police.

In California, these data are readily accessible and can be selected for specific time periods and locales. Some researchers have raised a concern that not-at-fault driver benchmarks may not represent an unbiased estimate of the driving population (Ridgeway & MacDonald, 2010), while others (Withrow & Williams, 2015) suggest that at-fault drivers may represent an improved proxy for risky driving and therefore those most at-risk of being stopped. The current analysis uses both types of crash benchmarks to ensure a comprehensive analysis of the LPD traffic stop data. Previous research using this benchmark largely confirms that non-White drivers experience elevated rates of contact relative to White drivers (Alpert, Dunham, & Smith, 2007; Engel, Frank, Tillyer & Klahm, 2006; Farrell et al., 2004; Rojek, Rosenfeld, & Decker, 2004; Smith et al., 2021; Smith & Petrocelli, 2001).

In sum, early benchmarking studies frequently used Census data as a comparison against the rate of traffic stops of non-White populations, but this approach has been soundly rejected as the science of traffic stop benchmarking has improved. Current best practices in traffic stop analyses rely on the VOD and crash-based benchmarks to provide better proxies for the driving and traffic law violating populations and as comparisons for police traffic stop data. Given the availability of the requisite data needed to employ these benchmarks, they were selected as appropriate options for an analysis of the LPD traffic stop data.

Arrests

The other officer decision of interest to the LPD concerned possible racial or ethnic disparities in arrests arising from traffic stops (or other police-civilian encounters initiated by officers). Arrest has a long history of study within the policing literature, and generally such analyses do not suffer from the same data limitations as the stop decision itself. As outlined previously, traffic stop data often require an external data source for comparison (i.e., a benchmark), while in the case of traffic stop outcomes, the universe of encounters in which an arrest could occur is known. While some have argued that selection bias in the decision to make a stop in the first place puts minority drivers at greater risk for arrest (Bronner, 2020), police agencies currently do not collect information on when a traffic stop could have been legally initiated but was not. With this theoretical limitation in mind, examining traffic stop arrest outcomes takes into account all traffic stop encounters, including those in which an arrest could have occurred but did not.[1] As a result, the analytic tools available to identify racial/ethnic disparities in arrest are suitable for revealing patterns of disparity in how drivers of different races and/or ethnicities experience arrest outcomes following traffic stops or other encounters with the police.

Often the goal of an arrest analysis is to identify whether the race/ethnicity of the civilian involved in the police-civilian encounter is a significant factor in whether or not the incident results in an arrest. Because the goal is to understand *officer decision-making* in relation to civilian race/ethnicity, it is important to consider the reason for the arrest and the degree of discretion the officer had in making it. For example, some arrests include a high level of discretion by the officer (e.g., arrests based on probable cause developed on the scene) while others (arrests based on a pre-existing warrant or the discovery of contraband following a search) involve little or no discretion. High discretion arrests should be analyzed separately from low discretion arrests as the goal is to understand whether officers are engaging in bias-based behavior of their own volition. Thus, arrests first should be categorized as high or low discretion and then each group should be analyzed separately. Disparities in high discretion arrests may be indicative of biased decision-making, while disparities in low discretionary arrests address a different question: Do such arrests disproportionately impact non-White civilians?

Critical to any arrest analysis is the ability to measure all relevant other factors that may be

[1] In the case of Livermore, the UTSA research team found no consistent pattern of racial or ethnic disparity in who was stopped by the LPD. Thus, any theoretical impact of stop selection bias on post-stop arrest outcomes was moot since no significant disparities were found in the traffic stops themselves, which made up the great majority of LPD encounters resulting in an arrest. See Section IV below for the results of the traffic stop benchmarking and arrest analyses.

associated with or influence the likelihood of an arrest. These factors can be grouped into several broad categories, including characteristics of the encounter, civilian characteristics, officer characteristics, and contextual factors. The relationship between civilian race/ethnicity and arrest is often of central concern, but it is important to assess this relationship while also considering the impact of other variables. For example, situational characteristics such as time of day or number of bystanders may be important. Likewise, civilian gender or age also may be related to the likelihood of an arrest. In addition, previous studies have demonstrated the importance of civilian demeanor as a predictor of whether or not an arrest is likely to occur (Kochel et al., 2011). The characteristics of the officer or the environmental context of the encounter (i.e., crime rate) may also be influential in understanding the nature of arrests. In sum, the goal is to evaluate the likelihood of an arrest by considering as many potentially relevant factors as possible in order to most accurately identify the contribution, if any, that civilian race/ethnicity has on the likelihood of arrest. The most common method to accomplish this goal is to estimate multivariate models that allow the relationship between each variable, including race/ethnicity, and arrest to be independently assessed while holding the others constant (see the next section for a specific description of this analytic approach).

Previous research on the relationship between civilian demographics and arrest (within traffic stops, in particular) has produced a mixed set of results. Several studies document an elevated likelihood of arrest for non-White groups (e.g., Alpert et al., 2006; Smith & Petrocelli, 2001; Withrow, 2004), while others demonstrate no effect (Alpert Group, 2004; Engel, Frank, Tillyer, & Klahm, 2006; Tillyer & Engel, 2013). In a recent assessment of traffic stop outcomes in San Jose, CA., no relationship was reported between civilian race/ethnicity and warrantless arrests or those conducted due to a warrant (Smith et al., 2016).

Similarly, recent literature on the relationship between civilian sex and age and arrest also reveals some mixed findings. Male drivers consistently experience arrest at a higher rate than female drivers (Alpert et al., 2006; Alpert Group, 2004; Engel et al., 2005, 2006; Engel, Tillyer, Cherkauskas, et al., 2007; Gumbhir, 2004; M. Smith & Petrocelli, 2001; Tillyer & Engel, 2013), whereas, civilian age has demonstrated an inconsistent relationship with arrest with some studies finding that older drivers were more likely to be arrested (Engel et al., 2005) and other studies reporting that older drivers were less likely to be arrested (Alpert et al., 2006; Engel et al., 2006; Gumbhir, 2004; M. Smith & Petrocelli, 2001; Tillyer & Engel, 2013).

III. Methodology

Consistent with the research goals agreed upon with the LPD, this project analyzed LPD traffic stop data to identify possible racial or ethnic disparities among those stopped by the police. Second, these data were examined to identify any racial/ethnic disparities among those arrested by the LPD following a traffic stop. To accomplish these two goals, the following methodologies were used.

Traffic Stops

An assessment of the racial/ethnic composition of traffic stops initiated by LPD officers involved several steps. Initially, all available information on the traffic stops was summarized in a series of descriptive tables. For example, the percentage of stops involving a male civilian or White civilian was calculated to provide a summary of the traffic stop characteristics. Next, we conducted benchmarking analyses that compared the percentage of traffic stops involving each of the racial/ethnic groups to those groups' expected risk of being stopped. As outlined in Section II above, two benchmarking techniques were utilized - a veil of darkness analysis and a comparison of stops to crash data.

The veil of darkness analysis calculates the percentage of traffic stops made of each racial/ethnic group during the daytime and nighttime and compares them to identify any difference. A higher rate of daytime stops involving non-White drivers compared to their percentages of nighttime stops suggests a potential difference in the decision-making process to initiate a traffic stop involving these groups. Differences in rates of daytime and nighttime stops for each group were subjected to statistical testing at the group and individual level. An ANOVA test was conducted to assess whether there were differences in the rates between daytime and nighttime stops across all groups that represented a statistical pattern. Individual t-tests were also conducted within each group to compare, for example, whether the daytime rate of stops involving Black drivers differed from the nighttime rate of Black stops. For both analyses, a statistically significant result from these tests would provide empirical evidence of a pattern of disparity whereas a non-significant result would support a conclusion that no substantive difference exists between the groups.

The second benchmark analysis used uniform traffic crash report (CHP 555) data from the State of California for Livermore to provide an estimate of risk for being stopped. These crash data include the race/ethnicity of the drivers involved in traffic collisions and were downloaded from Statewide Integrated Traffic Records System (SWITRS) for two-vehicle crashes that occurred within the City of Livermore between January 1, 2019 and April 30, 2021. In the analyses reported below, traffic crash data were compared to police stop data with not-at-fault drivers serving as an estimate of the driving population in the city and at-fault drivers serving as an estimate of those who violate the traffic laws. A statistically significant higher percentage of stops involving specific racial/ethnic groups in the LPD stop data compared to the crash data benchmarks would indicate disproportionate stops by LPD officers based on the expected risk for stops among those groups of drivers (Alpert et al., 2004; Tillyer et al., 2010; Withrow & Williams, 2015; COPS Smith et al., 2021).

For all benchmarks (i.e., daytime vs. nighttime; traffic stops vs. at-fault crashes; traffic stops vs. not-at-fault crashes), disproportionality indices (DI) were calculated. The DI is a within-group assessment that compares the stop rates for each racial/ethnic group in the traffic stop data to the 'expected' rates of stop for each group based on the selected benchmark. A value of 1.0 indicates alignment between the actual stop rate and the benchmark, while a value above 1.0 indicates that the racial/ethnic group experienced a higher than anticipated stop rate compared to the benchmark. The DI is used to compare the actual rate of stops to the expected rate of stops (based on the benchmarks) within racial/ethnic groups. To further compare stops of non-White drivers to White drivers, a disproportionality ratio (DR) was calculated by dividing the DI rate for the racial/group of interest (e.g., Black) by the White DI rate. The resulting DR value is interpreted as the likelihood of a Black (or any other racial/ethnic group) driver being stopped in comparison to chances of a White driver being stopped. For example, if the disproportionality ratio is 3.0, this indicates that the group of interest is three times as likely to be stopped in comparison to the White group (Smith et al., 2021).

Arrests

The analysis of arrests involved two primary statistics: descriptive and inferential. Descriptive statistics provide a summary of the variables across all cases to allow an assessment of how frequently each variable presents itself within the data. This is most frequently accomplished by calculating a percentage of cases in which this characteristic appears within all records. For example, all records are assessed to identify the number which conclude with an arrest and this information can be used to produce a percentage of encounters involving an arrest.

The second analytic tool used for examining arrests involved the estimation of multivariate regression models. Multivariate models offer the ability to identify the specific effects of each independent variable on the dependent variable by controlling for all other independent variables (Hanushek & Jackson, 1977; Weisburd & Britt, 2004). This approach considers all variables simultaneously to assess which of the encounter characteristics are related to the likelihood of an arrest. This type of modeling is particularly useful in identifying whether a relationship between civilian race/ethnicity and an arrest exists while considering all other potential factors. Importantly, this technique is only as robust as the information that is available, and any variables that may influence an arrest but are not available for inclusion in the model weaken its explanatory power. For example, citizen behavior or demeanor may influence the arrest decision but was not available in the data. With these limitations in mind, the LPD data contained sufficient variables to allow for a meaningful analysis of the arrest decision.

Data

This study analyzed LPD traffic stops generated between January 1, 2019 and April 30, 2021. A total of 24,944 records (i.e., cases) were received from the LPD and considered for analyses.

The initial analytic step involved an assessment of available fields and cases to determine the completeness of each record. Based on conversations with the LPD, the 'Incident Number' represented a unique number that signified a police-civilian encounter that may be eligible for analysis. De-identified officer data also were supplied to the research team and were merged into the stop data so that each record contained information on the primary officer involved in the traffic stop. Information on 120 officers was supplied to the research team.[2] A third data source provided by the LPD included records for violent and selected property crimes that occurred within the various LPD beats during the study period. Information on size of the residential population was also added to allow for creation of a violent and property crime rate. Of note, the internal organizational structure of the LPD was modified during the study period from three beats (i.e., A, B, and C) to four beats (i.e., 1-4). This had implications for the multivariate analyses of arrests and is discussed in more detail below.

Table 1 describes the initial data available for analysis. Of the 24,944 original records, 24,846 cases contained a unique 'Incident Number', and 98 records contained a duplicate 'Incident Number'. An additional 35 records were non-officer-initiated contacts. After removal of these records, 24,811 cases remained and were assessed for missing information in preparation for the analyses of arrests.

With respect to the benchmarking analyses, 1,922 non-traffic stops and 76 cases missing information on 'type' were removed. This resulted in 22,813 records that were assessed for missing data prior to conducting the benchmark analyses.

[2] The merging process (officers to stops) produced a 98.4% matching rate with 24,548 traffic stop records supplemented with officer characteristics.

Table 1: Data

	Records/Cases
Original Data	24,944
Duplicate Records	98
Non-Officer-Initiated Contacts	35
Sub-Total for Arrest Analysis	***24,811***
Non-Traffic Stops (i.e., Ped. Stops, Unlicensed, License Suspended, Other)	1,922
Missing 'Type'	76
Sub-Total for Traffic Stops	***22,813***

The next step was to analyze the variables for potential missing information that would preclude that record from further analysis. Table 2 outlines the available variables and groups them into situational, civilian, officer, and contextual categories. Each variable is described in terms of its measurement and then information on missing records (overall number of records and percentage) is provided along with the total number of records available for analysis. This assessment was conducted independently for the data used for the traffic stop analysis and then replicated for the data involved in the arrest analysis.

Overall, there was a very small percentage of missing data across all fields. The missing data rates for civilian characteristics (i.e., race/ethnicity, gender, and age) were below 1%, which is an unusually impressive level of completeness compared to many police administrative data sets. Missing data rates below 10% are acceptable with 5% or less missing preferred. The low missing rate of less than 1% of civilian characteristics demonstrates the commitment of LPD officers to collecting all required information as part of the RIPA process. In addition, less than 2% of all cases were missing an organizational unit identifier (i.e., beat) and less than 1% were missing officer characteristics (i.e., race/ethnicity, gender, age, years of experience, or assignment). Of note, a violent crime rate was calculated for each beat by counting the number of homicides, aggravated assaults, robberies, and sexual assaults reported within each organizational unit and then standardizing this by the residential population in those beats. The result is a violent crime rate that reflects the number of violent crimes per 1,000 population. A similar process was used to create a property crime rate at the beat level based on burglaries, thefts, and vehicle thefts. Importantly, the LPD changed its beat borders in February 2020, and population estimates were only available for the police-civilian encounters occurring since that date. As a result, there is a high missing rate for this variable. This does not reflect any incomplete information recorded by the LPD; rather, it is a product of not being able to access the population figures for the beats prior to the change in boundaries.

Table 2: Variables & Missing Data

Variables	Measurement	Traffic Stops (N=22,813)			Arrests (N=24,811)		
		Missing		Final	Missing		Final
		N	%	N	N	%	N
Situational Variables							
Date & Time	Year, Month, Day of Week	0	0.0%	22,813	0	0.0%	24,811
Type of Contact	Type, Source	0	0.0%	22,813	76	0.3%	24,735
Organizational Unit	Beat	407	1.8%	22,406	415	1.7%	24,396
Civilian Variables							
Race/Ethnicity	White, Black, Hispanic, Asian, Other	76	0.3%	22,737	88	0.4%	24,723
Gender	Male, Female	0	0.0%	22,813	0	0.0%	24,811
Age	15-24, 25-32, 33-39, 40-48, 49-99	39	0.2%	22,774	48	0.2%	24,763
Officer Variables							
Race/Ethnicity	White, Black, Hispanic, Asian, Other	175	0.8%	22,638	216	0.9%	24,595
Gender	Male, Female	175	0.8%	22,638	216	0.9%	24,595
Age	21-55	175	0.8%	22,638	216	0.9%	24,595
Years of Experience	0-28	175	0.8%	22,638	216	0.9%	24,595
Assignment	Various Categories	175	0.8%	22,638	216	0.9%	24,595
Contextual Variables							
Violent Crime Rate	Violent Crime per 1,000 population	N/A	N/A	N/A	15,268	61.5%	9,543
Property Crime Rate	Property Crime per 1,000 population	N/A	N/A	N/A	15,268	61.5%	9,543

IV. Findings

Traffic Stops

Traffic stops initiated by LPD officers during the study period were examined to identify their racial/ethnic composition. White drivers were the most common group contacted by LPD officers with 44.2% of all stops involving that group. Hispanic drivers were the next most common and comprised slightly more than one quarter of all stops (28.1%). The remainder of the stops involved drivers of Other races/ethnicities (12.1%), Black drivers (10.3%), and Asian drivers (5.2%). The distribution of stops by race/ethnicity is summarized in Table 3.

Table 3: Civilian Race/Ethnicity in Traffic Stops

Total Cases: 22,737	**Percentage**
White	44.2%
Black	10.3%
Hispanic	28.1%
Asian	5.2%
Other	12.1%

The first assessment of the traffic stops was a veil of darkness analyses. As described previously, the veil of darkness requires the identification of traffic stops that occurred during the inter-twilight period or the period in Livermore between when the sun set the earliest (16:48) and latest (20:33) during the year. All stops occurring during the inter-twilight period were identified as either a daytime or nighttime stop depending on when during the year the stop was initiated and whether the stop took place before or after sunset on the day the stop was made. These stops were then summarized by the drivers' racial/ethnic composition.

Again, the veil of darkness analysis measures variance in the daytime stop rates of non-White drivers compared to the nighttime stop rates for these groups. Any difference in the within group rates between daytime and nighttime stop rates suggests evidence of a disparity. Two important points are critical when considering disparity. First, a simple difference in the stop rates must be assessed to determine statistical significance or whether the difference is large enough that it is unlikely due to chance. If an observed disparity is statistically significant, this does not necessarily prove bias or discrimination, which typically requires additional evidence that stops were motivated by a discriminatory purpose *(United States v. Armstrong*, 1996; *Ballou v. McElvain*, 2021). Such a determination is beyond the scope of this report and the data available to the research team. Instead, the veil of darkness analysis allows an assessment of patterns of disparity and areas of department action that may need further review or attention from LPD leadership.

Table 4 summarizes the rates of stops for each group during the daytime and nighttime and also reports on the two analytic tests estimated to identify any statistical differences between the experience of these groups depending on the time of day. Overall, slightly more stops occurred during daylight hours (N=1,552) compared to nighttime hours (N=1,148). During both daytime and nighttime, White drivers were the majority group stopped by LPD officers (i.e., 44.5% and 43.8%) with Hispanic drivers being the second most common group involved in traffic stops (i.e., 30.0% and 29.4%).

An overall assessment of the across-group rates using an ANOVA resulted in a non-statistically significant result. Additional within-group analyses using t-tests also demonstrated no statistically significant results. The daytime stop rates of two minority groups – Hispanic (30.0% of daytime stops) and "Other" (11.5% of daytime stops) drivers – slightly exceeded their nighttime stop rates (29.4% and 10.0% of nighttime stops, respectively), but these differences were not statistically significant. The daytime stop rate of Whites also slightly exceeded this group's nighttime stop rate, but this difference also was non-significant. Finally, the daytime stop rate of Black drivers (9.7%) was actually lower than this group's nighttime stop rate (11.4%), which is not consistent with a pattern of disparate enforcement, and a similar pattern was observed for Asian drivers (4.3% of daytime stops vs. 5.4% of nighttime stops).

In sum, the veil of darkness analyses demonstrated no statistically significance difference in rates of traffic stops for the various racial/ethnic groups in stops during the daytime compared to the nighttime. While there were some minor differences in the rates of stops during the study period, these variations do not reflect a statistically significant pattern of racial/ethnic disparities in LPD stop practices.

Table 4: Civilian Race/Ethnicity in Daytime vs. Nighttime

Total Cases: 2,715	**Daytime (N=1,552)**	**Nighttime (N=1,148)**
White	44.5%	43.8%
Black	9.7%	11.4%
Hispanic	30.0%	29.4%
Asian	4.3%	5.4%
Other	11.5%	10.0%

Results were non-significant based on an ANOVA analysis; individual t-tests were also non-significant.
Civilian race/ethnicity was missing on 15 cases (0.6%).

The second assessment of traffic stops involved a statistical analysis of all traffic stops compared to the vehicle crash data. Using the SWITRS database on crashes, data were extracted for the City of Livermore between January 1, 2019 and April 30, 2021. During this period, 4,736 individual drivers were involved in crashes. Of those, 4,436 records were identified as either at-fault or not-at-fault; however, an additional 441 records were missing race/ethnicity. After removal of these data due to missing information, 3,995 records provided required information for analysis (1,775 at-fault drivers and 2,220 not-at-fault drivers). Table 5 reports on the percentage of at-fault and not-at-fault drivers based on their racial/ethnic group.

Table 5: Civilian Race/Ethnicity in Crashes

Total Cases: 3,995	**At-Fault Crashes (N=1,775)**	**Not-At-Fault Crashes (N=2,220)**
White	40.7%	41.1%
Black	9.4%	8.4%
Hispanic	33.7%	28.6%
Asian	5.4%	8.1%
Other	10.9%	13.9%

Driver's race/ethnicity was missing in 441 (.9%) of the 4,436 crashes

These crash data rates were then used as a benchmark to compare against LPD traffic stops by racial/ethnic group. Table 6 provides a summary of the rates of traffic stops, at-fault crashes, and not-at-fault crashes for each racial group. Thereafter, disproportionality indices (DI) and ratios (DR) are reported for each benchmark. A DI above 1.0 indicates that the group of interest experienced a higher rate of stops compared to the rate of stops for that group using the benchmark. For example, White drivers possessed a DI of 1.1 when using the at-fault and not-at-fault benchmarks suggesting that their rate of stops was slightly above what was expected based on their representation in each benchmark. Black drivers also showed a slightly elevated DI rate when using the at-fault (1.1) and not-at-fault (1.2) benchmarks. The comparisons for the remaining racial/ethnic groups were either at or below 1.0 (with the exception of Other drivers when using the at-fault benchmark) suggesting no pattern of differential stops of these groups.

The DR statistic extends the analyses by comparing the DI rate for the group of interest (i.e., Black, Hispanic, Asian, and Other drivers) to the DI rate for White drivers. In short, this assessment compares the experience of groups of primary interest to that of White drivers. Similar to the DI, a rate above 1.0 indicates that the group of interest experienced rate of stops at an elevated rate compared to White drivers. The only comparison that demonstrated a slightly elevated rate of stops was for Black drivers compared to White drivers when using the not-at-fault benchmark. In that case, Black drivers were 1.1 times more likely to be stopped compared to White drivers when using the not-at-fault benchmark as a proxy for risk of stop, which indicates a slightly elevated disparity in stops relative to Whites. However, the DI for Black at-fault drivers was not elevated compared to Whites.

Table 6: Disproportionality Indices & Ratios

	Traffic Stops (N=22,737)	At-Fault Crashes (N=1,775)	Not-At-Fault Crashes (N=2,220)	At-Fault DI	Not-At-Fault DI	At-Fault DR	Not-At-Fault-DR
White	44.2%	40.7%	41.1%	1.1	1.1	--	--
Black	10.3%	9.4%	8.4%	1.1	1.2	1.0	1.1
Hispanic	28.0%	33.7%	28.6%	0.8	1.0	0.8	0.9
Asian	5.2%	5.4%	8.1%	1.0	0.6	0.9	0.6
Other	12.1%	10.9%	13.9%	1.1	0.9	1.0	0.8

Arrests

Arrests arising from activities initiated by officers (as opposed to calls for service) were analyzed for patterns of racial disparity using multivariate modeling. Most arrests arose from traffic stops (91.9%), but pedestrian stops (7.0%) and other miscellaneous types of encounters (0.8%) also contributed to the 445 arrests that took place during the 24,065 LPD officer-initiated encounters with civilians analyzed from January 1, 2019 through April 30, 2021 (Table 7).[3]

Arrests occurred in 1.8% of all encounters during the study period. Importantly, no information was recorded about the reason for the arrest or the type of arrest that was undertaken. As described in Section II, high discretion arrests ideally would be analyzed separately from low discretion arrests. However, due to the data collection protocols used by the LPD during the study period, this was not possible.

In addition to the arrest outcomes, additional variables were available to help inform the multivariate disparity analysis. These variables also are summarized below in Table 7. The majority of contacts occurred in 2019 (54.0%), during a weekday (82.4%), and during daylight hours (62.1%). The racial/composition of civilians involved in these encounters was predominately White (45.6%), with slightly more than a quarter of all civilians identified as Hispanic (28.1%). Black civilians comprised 10.0% of all contacts, Asian civilians were involved in 4.9% of all incidents, and persons of Other races/ethnicities comprised the remaining 11.5% of all encounters. Male civilians were involved in 71.2% of all encounters, and 20.3% of these incidents involved a civilian under the age of 24.

Officers initiating these contacts were predominately White (89.0%), with a small representation of Black (0.7%), Hispanic (6.6%), Asian (0.5%), or Other (3.2%) officers. Male officers initiated the encounters in 92.7% of the cases, and officers were, on average, 38 years of age with 11 years

[3] 70 records did not indicate the 'type' of stop (0.3%).

of experience. Slightly more 60% of the contacts were initiated by an officer assigned to Patrol. Finally, these contacts occurred in beats with an average violent crime rate of 2.7 per 1,000 population and an average property crime rate of 23.9 per 1,000 population.[4]

Table 7: Descriptives

N=24,065	**Percent**		**Percent/ Average**
Arrest	1.8%		
Encounter Variables		*Officer Variables*	
Year 2019	54.0%	Race/Ethnicity	
Year 2020	34.0%	White	89.0%
Year 2021	12.1%	Black	0.7%
Weekend	17.6%	Hispanic	6.6%
Daytime	62.1%	Asian	0.5%
Civilian Variables		Other	3.2%
Race/Ethnicity		Male	92.7%
White	45.6%	Age	37.65
Black	10.0%	Years of Experience	11.26
Hispanic	28.1%	Patrol Officer	61.0%
Asian	4.9%	*Contextual Variables*	
Other	11.5%	Violent Crime Rate	2.72
Male	71.2%	Property Crime Rate	23.93
Under 24 Years of Age	20.3%		

Violent and property crime rate are based organization beats in effect since Feb 2020.

Two models were estimated using the data provided by the LPD (see Table 8). Model 1 uses all available records to analyze the impact of encounter, civilian, and officer variables on the likelihood of arrest. Model 2 includes all these variables but also considers beat-level crime rates as predictors. In these models, three key pieces of information are provided. First, statistical coefficients are provided that indicate the direction of the relationship between the variable shown and the arrest outcome. A positive value indicates an increased likelihood of arrest associated with this variable; conversely, a negative value means that the chances of an arrest are reduced when this variable is present. Statistical significance is denoted with asterisks, which indicate that the variable influenced the arrest outcome to a degree unlikely due to chance. The number of asterisks indicates the level of confidence in that relationship. For example, a single asterisk represents a 95% degree of confidence that the relationship was not due to chance. Two asterisks represent a

[4] The violent crime and property crime rates were only able to be calculated for encounters occurring since Feb 2020 due to a lack of data on population size in the beats pre-Feb 2020.

confidence interval of 99% and so on. The magnitude or impact of statistically significant coefficients is shown with an odds ratio, which provides an interpretable number to indicate how much more likely an arrest is to occur when that variable is present in the encounter. An odds ratio of 2.0, for example, would indicate that the odds of arrest were two times higher when that variable was present during the police-civilian encounter.

The results from Model 1 reveal several statistically significant variables. Of primary interest, the race/ethnicity of the civilian was related to the likelihood of an arrest across several racial and ethnic groups. In this analysis, White civilians serve as the referent group to which minority groups should be compared. Black civilians were statistically indistinguishable from White civilians in terms of arrest likelihood, while Hispanic, Asian, and Other civilians all had lower odds of an arrest compared to Whites. Similarly, males were 1.91 times more likely to be arrested than females who served as the referent gender, while civilians under the age of 24 experienced a lower likelihood of arrest.

Other important predictors of an arrest included time of the day. Arrests were 1.78 times more likely to occur during daylight hours compared to nighttime hours. Also, two officer characteristics were associated with the likelihood of an arrest. Officers with less experience and those assigned to patrol were more likely to conclude an encounter with an arrest. Of note, the race/ethnicity and sex of the officer were not related to the likelihood of an arrest.

Model 2 included the same variables as Model 1 but also included beat-level crime rates for encounters that took place after February 2020. Results were largely consistent with Model 1 with two exceptions. First, the time of day become statistically non-significant and was no longer related to the likelihood of an arrest, and second, encounters occurring within beats with higher violent crime rates were more likely to result in an arrest.

Table 8: Arrest Multivariate Models

	Arrest Model 1 N=24,065		Arrest Model 2 N=9,425	
	Coeff.	**Odds Ratio**	**Coeff.**	**Odds Ratio**
Intercept	-4.114***	--	-5.451***	--
Encounter Variables				
Year 2019	.248	--	--	--
Year 2020	.014	--	.066	--
Weekend	.111	--	.040	--
Daytime	.576***	1.78	-.122	--
Civilian Variables				
Black	-.040	--	-.422	--
Hispanic	-.301**	0.74	-.046	--
Asian	-2.293***	0.10	-2.171*	0.11
Other	-1.443***	0.24	-1.120*	0.33
Male	.648***	1.91	.611*	1.84
Under 24 Years of Age	-.715***	0.49	-.794**	0.45
Officer Variables				
Black	-.261	--	--	--
Hispanic	.167	--	-.158	--
Other	-.267	--	-.009	--
Male	-.147	--	-.111	--
Years of Experience	-.082***	0.92	-.050*	0.95
Patrol Officer	.436**	1.55	1.06**	2.86
Contextual Variables				
Violent Crime Rate	--	--	.440**	1.55
Property Crime Rate	--	--	-.017	--
Model R^2 (Nagelkerke)	.086		.113	

*p ≤ .05, **p ≤ .01, ***p ≤ .001
Reference Groups: Year 2021, White Drivers, White Officers
Black officers were not included in Model 2 as they only accounted for 19 cases and generated unstable standard errors which make statistical modeling inappropriate.

V. Summary & Conclusions

A research team of criminologists from the University of Texas at San Antonio analyzed 24,944 encounters between LPD officers and civilians that took place between January 1, 2019 and April 30, 2021 for patterns of racial and ethnic disparity. The analysis addressed two areas of possible disparity: (1) disparities in traffic stops and (2) disparities in arrests. The traffic stop analysis made use of two benchmarking techniques that have been well-accepted in the peer reviewed literature. A "veil of darkness" (VOD) analysis examined differences in stop rates of non-White and White drivers during the daytime compared to the nighttime. A higher rate of non-White stops during daylight hours when race and ethnicity are more visible to officers prior to the stop is suggestive of possible racial bias (Grogger & Ridgeway, 2006). In addition, data obtained from a State of California-maintained database (SWITRS) containing the racial composition of not-at-fault and at-fault drivers involved in two-vehicle crashes was used as a proxy for the driving and traffic law violating populations in Livermore and was compared against the racial/ethnic composition of drivers stopped by the LPD during the period of study (see Alpert, Smith, & Dunham, 2004). These two benchmarking analyses allowed for an assessment of whether LPD officers stopped non-White drivers at rates that exceeded the risk for a stop expected for these groups and if so, how that increased risk compared to stops experienced by White drivers. In the same vein, a multivariate analysis of arrests examined whether civilian race/ethnicity predicted the likelihood of an arrest by the LPD after controlling for other relevant factors available in the data or from external sources (e.g., beat-level crime rates).

The VOD analysis found no statistically significant differences in the rates at which non-White drivers were stopped in Livermore during the day compared to at night. This finding suggests that the race/ethnicity of the driver did not influence the decision by LPD officers to initiate traffic stops. The traffic crash benchmark analysis found slightly elevated risks for stops of White, Black, and "Other" race drives (at-fault benchmark only) and a slightly elevated risk for stops of Black drivers relative to White drivers (not-at-fault benchmark only). Together, **the results from these two benchmark analyses do not suggest a pattern of racial/ethnic disparity in traffic stops experienced by non-White drivers in Livermore.**

The multivariate arrest analysis found a *decreased* risk for arrest among non-White civilians in Livermore compared to White civilians when other relevant factors (day of week, time of day, officer race/ethnicity/gender, area crime rates) were held constant. **The results of the arrest analyses also do not reveal a pattern of racial disparity in police outcomes that disadvantages non-White civilians in the City of Livermore.**

Compared to most other traffic and arrest disparity studies reported in the literature, **no clearly identifiable or concerning pattern of racial/ethnic disparity was found in the 24,944 police-**

civilian encounters that took place in Livermore over a 26-month period during 2019-21. This is an unusual and encouraging result and suggests the LPD and city leadership are committed to providing fair and constitutional policing to the community of Livermore. They should be commended for these findings.

With these encouraging results in mind, the UTSA research team recommends regular audits of the LPD's RIPA data to assess its completeness and validity and to ensure that officers remain in compliance with the letter and spirit of the law. In addition, the team recommends an annual analysis of the RIPA data to identify any racial/ethnic disparities of concern embedded within the detailed information that RIPA now mandates be collected. For example, a fulsome analysis of citations and searches will be possible once sufficient cases are accumulated in the data, which may reveal areas that require additional training or monitoring. Subsequent arrest and search analyses can make use of RIPA's improved level of detail to separate out high and low discretion searches and arrests and examine potential disparities in outcomes that fall on the higher end of the discretion continuum. Racial and ethnic disparities in the use of force are of national concern and the LPD may consider working with an experienced research team to ensure that it is collecting the appropriate information on use of force cases and analyzing the resulting data to its full potential.

VI. References

Alpert Group. (2004). *Miami-Dade Police Department Racial Profiling Study*. Available at: http://www.policeforum.org/library/racially-biasedpolicing/supplemental-resources/Alpert_MDPDRacialProfilingStudy%5B1%5D.pdf.

Alpert, G. P., Smith, M. R., & Dunham, R. (2004). Toward a better benchmark: Assessing the Utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy*, 6(1), 43–69.

Alpert, G. P., Becker, E., Gustafson, M. A., Meister, A. P., Smith, M. R., & Strombom, B. A. (2006). Pedestrian and Motor Vehicle Post-stop Data Analysis Report. Available at: http://www.analysisgroup.com/AnalysisGroup/uploadedFiles/Publishing/Articles/LAPD_Data_Analysis_Report_07-5-06.pdf.

Alpert, G. P., Dunham, R., & Smith, M. R. (2007). Investigating racial profiling by the Miami-Dade police department: A multimethod approach. *Criminology & Public Policy*, 6(1), 25–56.

Ballou v. McElevain, No. 3:19-cv-05002-RBL (9th Cir. Sept. 28, 2021)

Bronner, L. (2020, June 25). Why statistics don't capture the full extent of the systematic bias in policing. *FiveThirtyEight.* Available at: *https://fivethirtyeight.com/features/why-statistics-dont-capture-the-full-extent-of-the-systemic-bias-in-policing/*.

Channin, J., Welsh, M., Nurge, D., & Henry, S. (2016). *Traffic Enforcement in San Diego, California: An Analysis of SDPD Vehicle Stops in 2014 and 2015*. San Diego, CA: San Diego State University. Available at: https://www.sandiego.gov/sites/default/files/sdpdvehiclestopsfinal.pdf

COPS. (2016). *Collaborative Reform Initiative: An Assessment of the San Francisco Police Department*. Washington, DC: Office of Community Oriented Policing Services. Available at: https://riczai-inc.com/Publications/cops-w0817-pub.pdf

Engel, R. S., Frank, J., Tillyer, R., & Klahm, C.F. (2006). *Cleveland Division of Police Traffic Stop Data Study: Final Report*. Cincinnati, OH: University of Cincinnati. Submitted to the Cleveland Division of Police, Cleveland, OH.

Engel, R. S., Calnon, J. M., Tillyer, R., Johnson, R., Liu, L., & Wang, X. (2005). *Project on*

Police-citizen Contacts: Year 2 Final Report (May 2003-April 2004). Cincinnati, OH: University of Cincinnati. Submitted to the Commissioner of the Pennsylvania State Police, Harrisburg, PA.

Engel, R. S., Tillyer, R., Cherkauskas, J. C., & Frank, J. (2007). *Traffic Stop Data Analysis Study: Year 1 Final Report.* Cincinnati, OH: University of Cincinnati. Submitted to the Arizona Department of Public Safety, Phoenix, AZ.

Farrell, A., McDevitt, J., Bailey, L., Andresen, C., & Pierce, I. (2004). *Massachusetts Racial and Gender Profiling Final Report*. Boston, MA: Institute on Race and Justice, Northeastern University.

Fridell, L. (2004). *By the Number: A Guide for Analyzing Race Data from Vehicle Stops.* Washington, D.C.: Police Executive Research Forum.

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.

Gumbhir, V. K. (2004). *Final Report on the Eugene Police Department's Vehicle Stop Data (2002-2003), Condensed Report*. University of Oregon, Submitted to the Eugene Police Department, Eugene, OR.

Hanushek, E. A., & Jackson, J. (1977). *Statistical Methods for Social Scientists*. New York, NY: Academic Press.

Kalinowski, J., Ross, S. L., Ross, M. B. (2017). *Endogenous Driving Behavior in Veil of Darkness Tests for Racial Profiling*. Chicago, IL: The University of Chicago. Available at: https://econresearch.uchicago.edu/sites/econresearch.uchicago.edu/files/Kalinowski_Ross_Ross_2017_driving-veil-darkness_0.pdf

Kochel, T.R., Wilson, D.B., & Mastrofski, (2011) Effect of suspect race on officers' arrest decisions. *Criminology*, 49(2), 473-512.

Lovrich, N. P., Gaffney, M. J., Mosher, C. C., Pratt, T. C., & Pickerill, M. J. (2007). *Results of the Monitoring of WSP Traffic Stops for Biased Policing*. Pullman, WA: Division of Governmental Studies and Services, Washington State University.

Meehan, A. J., & Ponder, M. C. (2002). How roadway composition matters in analyzing police data on racial profiling. *Police Quarterly*, 5, 306-333.

Novak, K. J. (2004). Disparity and racial profiling in traffic enforcement. *Police Quarterly*, 7, 65−96.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., & Goel, S. (2019). A Large-scale Analysis of Racial Disparities in Police Stops across the United States. Stanford Computational Policy Lab. Available at: https://openpolicing.stanford.edu/publications/.

Ridgeway, G. (2007). *Analysis of Racial Disparities in the New York Police Department's Stop, Question, and Frisk Practices*. Santa Monica, CA: RAND.

Ridgeway, G., & MacDonald, J. (2010). Methods for assessing racially biased policing. In S. K. Rice, & M. D. White (Eds.), *Race, Ethnicity, and Policing: New and Essential Readings* (pp. 180–204). New York: NYU Press.

Ridgeway, G., & MacDonald, J. M. (2009). Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association*, 104: 661-668.

Ritter, J. A., & Bael, D. (2009). Detecting Racial profiling in Minneapolis Traffic Stops: A New Approach. Minneapolis, MN: University of Minnesota. Available at: http://www.cura.umn.edu/publications/catalog/reporter-39-1-2-1

Rojek, J., Rosenfeld, R., & Decker, S. (2004). The influence of driver's race on traffic stops in Missouri. *Police Quarterly*, 7: 126−147.

Ross, M. B., Fazzalaro, J., Barone, K., & Kalinowski, J. (2016). *State of Connecticut Traffic Stop Data Analysis and Findings, 2014–15.* Racial Profiling Prohibition Project. Connecticut State Legislature. New Britain, CT: Central Connecticut State University. Available at: https://www.ccsu.edu/imrp/Publicatons/Files/May%202016%20Connecticut%20Racial%20Profiling%20Report.pdf

Smith, M.R., Tillyer, R., Lloyd., C. & Petrocelli, M (2021) Benchmarking disparities in police stops: A comparative application of 2nd and 3rd generation techniques. *Justice Quarterly*, 38(3): 513-536.

Smith, M. R., & Petrocelli, M. (2001). Racial profiling? A multivariate analysis of police traffic stop data. *Police Quarterly*, 4: 4−27.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016). *A Test of Racial Disproportionality in Traffic Stops Conducted by the Raleigh Police Department*. Research Triangle Park, NC: RTI International. Available at: https://www.rti.org/sites/default/files/resources/rti-publication-file-5d8618d6-1723-45ad-96f8-33c08ed5ebd3.pdf

Tillyer, R., Engel, R. S., & Cherkauskas, J. C. (2010). Best practices in vehicle stop data collection and analysis. *Policing*, 33: 69-92.

Tillyer, R. & Engel, R.S. (2013) The impact of drivers' race, gender, and age during traffic stops: Assessing interaction terms and the social conditioning model. *Crime & Delinquency*, 59(3): 369-395.

United States v. Armstrong, 517 U.S. 456 (1996).

Weisburd, D., & Britt, C. (2004). *Statistics in Criminal Justice (2nd ed.)*. Belmont, CA.: Thomson and Wadsworth.

Withrow, B. L. (2004). Race-based policing: A descriptive analysis of the Wichita stop study. *Police Practice and Research*, 5: 223−240.

Withrow, B. L., & Williams, H. (2015). Proposing a benchmark based on vehicle collision data In racial profiling research. *Justice Review*, 40: 449–469.

Withrow, B. L., McLean, S. J., & Wheeler, A. P. (2012). Testing for racial Profiling with the veil-of-darkness method. *Police Quarterly*, 15(1): 92–111.