DISSERTATION

Incorporating Traffic Enforcement Racial Profiling Analyses into Police Department Early Intervention Systems

Brent D. Fulton

This document was submitted as a dissertation in October 2006 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of James N. Dertouzos (Chair), Greg K. Ridgeway, and John M. MacDonald.

RAND PARDEE RAND GRADUATE SCHOOL

The Pardee RAND Graduate School dissertation series reproduces dissertations that have been approved by the student's dissertation committee.

Published 2007 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213
RAND URL: http://www.rand.org/
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

Abstract

In response to legislation, lawsuits, community pressure, and internal concerns, over 4,000 police departments are collecting traffic enforcement data to determine whether their officers racially profile. In this context, racial profiling is defined as when police inappropriately use a motorist's race as a factor in deciding which motorists to stop, cite, search, and arrest, where the appropriateness of the use of race is defined by the law, consent decrees, and department policies. Most studies have estimated the use of race at the department level, but there is a growing interest to estimate its use at the officer level and incorporate the results into Early Intervention (EI) systems, which identify potential problem officers. For either department- or officer-level studies, the dominant concern is to employ methods that accurately estimate the use of race and distinguish between appropriate and inappropriate uses.

This study summarizes the key implications of incorporating racial profiling analyses into an EI system and improves upon existing methods that estimate the use of race in stop, search, and DUI arrest decisions at both the department and officer levels. The methods are applied to the traffic stops of 16 Washington State Patrol troopers who patrolled South Seattle during 2003 to 2005 in order to estimate the troopers' average use of race as well as each trooper's relative use of race as compared to his 15 peers.

For the combined trooper analyses, the study does not find conclusive evidence that the troopers used race as a factor in their stop, search, and DUI arrest decisions. However, the study does find that the use of race significantly varied among the troopers, and a few troopers are identified as *potential* problem troopers who warrant further scrutiny to determine if they inappropriately used race against minority motorists.

The results lead to two key policy recommendations. First, police departments should strongly consider incorporating traffic enforcement racial profiling analyses into their EI systems in order to identify potential problem officers. Second, methods used to identify potential problem officers need to allow for the possibility that an officer's peer group may change over time and that non-racial motorist characteristics differentially affect officers' law enforcement decisions. These types of methods should also be used to identify potential problem officers for other EI system performance indicators.

Table of Contents

List of Figures

List of Tables

Note to the reader: for brevity, male pronouns are used, but they should be considered gender neutral unless stated otherwise.

Acknowledgements

This dissertation would not have been possible without the guidance and support of many individuals. I am especially grateful to my committee—Jim Dertouzos, Greg Ridgeway, and John MacDonald—who spent countless hours providing guidance, reading drafts, and responding to questions. Jim Dertouzos, the chair of my committee, gave strategic direction throughout the process and helped me structure many concepts in economic terms. Greg Ridgeway patiently sat through countless meetings where we discussed methods, propensity score models, and R. Those meetings were invaluable learning experiences and critical in enabling me to complete this dissertation. John MacDonald helped me better understand law enforcement agencies from both a strategic and operational perspective, and also provided keen insight from a criminological point of view.

Professor Geoffrey Alpert, from the University of South Carolina, made valuable comments on an earlier draft, especially pertaining to the intricacies of police department early interventions systems. Professor Clayton Mosher, from Washington State University, discussed his reports on the Washington State Patrol (WSP) and helped me better understand WSP operations and its traffic stop dataset.

John Batiste, chief of the WSP, strongly supported this research and provided me invaluable access to the WSP, including many individuals from his command staff. Assistant Chief Brian Ursino, Captain Mike DePalma, Captain Steve Burns, and Sergeant Rod Gullberg participated in numerous conference calls and responded to many requests for information. They were all very responsive and helpful in answering questions about WSP policies and operations.

I am also grateful to former Dean Bob Klitgaard for his enthusiasm, encouragement, and guidance during my time at PRGS, especially during the dissertation process. The school's faculty equipped me with a solid analytical foundation, and many of my classmates strengthened that foundation, provided helpful comments on this dissertation, and became great friends.

And most importantly, I thank my family, including my parents who valued my education well before I did, and my wife, Lucie, who provided constant encouragement

and support throughout this process, and also made valuable comments on the many drafts that she read.

Chapter 1: Introduction, Policy Questions and Research Objectives

1.1 Introduction

Longstanding racial tension between minorities and the police harms the minority community and reduces the police's effectiveness within those communities. Some of the tension is from real or perceived racial profiling connected to traffic enforcement. In this context, racial profiling is defined as when police inappropriately use a motorist's race as a factor in deciding which motorists to stop for a traffic violation as well as which stopped motorists to cite, search, or arrest as a result of the stop.[1] The appropriateness of using race as a factor in these decisions is defined by the law, consent decrees, and police department policies.[2]

Based on a series of four Gallup polls between 1999 and 2004, most Americans, especially minorities, believe racial profiling is widespread within traffic stops, but relatively few think the practice is justified (Newport, 1999; Ludwig, 2003; Carlson, 2004).[3] According to the June 2004 poll, 53 percent of all adults, 67 percent of black adults, and 63 percent of Hispanic adults believe that racial profiling is widespread, but only 31 percent of all adults, 23 percent of black adults, and 30 percent of Hispanic adults think that the practice is justified (Carlson, 2004).[4]

However, racial profiling does not equally impact all individuals within a minority racial group. For example, the 1999 Gallup poll found that younger blacks were

---

[1] Other areas of concern include whether the average duration of a traffic stop differs among racial groups as well as whether race is used as a factor in deciding whether to check the records of a motorist (e.g., determine whether a vehicle owner, who is presumably the driver, has a suspended driver's license), which would lead to additional stops and arrests for the targeted racial group.

[2] For brevity, I refer to all law enforcement agencies as police departments. Also, note that the law, consent decrees, and police department policies define the circumstances under which the use of race is appropriate; however, a police department may restrict its officers' use of race further than what the law and consent decrees require.

[3] The Gallup polls asked survey respondents "whether police officers stop motorists of certain racial or ethnic groups because the officers believe that these groups are more likely than others to commit certain types of crime." As will be defined below, Gallup's definition of racial profiling is based on statistical discrimination, not racial animosity.

[4] Between the 1999 and 2004 polls, the point estimate of the proportion of adults who believe racial profiling is widespread slightly decreased; however, the decrease is not statistically significant at the 0.05 level.

more likely than older blacks to report that they felt they had been stopped by the police just because of their race. Similarly, male blacks were more likely than female blacks to report the same feeling. To illustrate the largest disparity, 72 percent of black males between the ages of 18 and 34 reported that they felt they had been stopped by the police just because of their race, while only 14 percent of black females over 49 years old reportedly felt the same (Newport, 1999).

During his campaign for president in 2000, President Bush promised to end racial profiling by law enforcement. The End Racial Profiling Act (EPRA) has been introduced several times in both the U.S. Senate and the U.S. House of Representatives, but it has not gained sufficient support for passage. On the other hand, 23 states have passed legislation prohibiting racial profiling (Amnesty International USA, 2004). Moreover, approximately 4,000 police departments are collecting and analyzing data related to traffic enforcement in order to determine if and to what degree racial profiling is occurring (Farrell et al., 2005).

1.2 Policy Questions and Research Objectives

Most traffic enforcement racial profiling studies[5] have estimated the use of race[6] at the department level, but there is a growing interest to estimate each officer's use of race and incorporate the results into an Early Intervention (EI) system designed to identify problem officers and improve officer performance (Walker, 2001; 2003a).[7] A police department's key policy decision is whether to incorporate traffic enforcement data into an EI system in order to estimate each officer's use of race, and ultimately, to

---

[5] The concern about racial profiling on the roadways is less related to enforcing traffic laws *per se*, but is more related to enforcing non-traffic laws, that is, stopping a motorist in order to discover non-traffic law violations (e.g., drug violations). Since the traffic stop is the law enforcement decision that initiates this chain of events, I refer to these law enforcement decisions—both traffic and non-traffic law enforcement—collectively as traffic enforcement.

[6] The term "use of race" includes both appropriate and inappropriate uses of race. Due to data limitations, it is often difficult to separately identify the appropriate and inappropriate uses of race, so most studies estimate the aggregate use of race (while controlling for observed confounding characteristics). However, the law rarely allows an officer to use race as a factor in most law enforcement decisions related to routine traffic stops, and the exceptions are discussed in Chapter 2. In this study, for brevity, when I use the term "use of race," it mostly includes the inappropriate use of race unless stated otherwise.

[7] These systems are also referred to as Early Warning (EW) systems, which are primarily designed to warn police leadership about problem officers. Because these systems are evolving into a general management tool—to identify both good and bad performance—they are now more commonly referred to in the literature as EI systems.

identify potential problem officers. Therefore, if a department decides to incorporate the data, then it needs to develop criteria in order to accurately identify potential problem officers. Whether the analysis is conducted at the department or officer level, the dominant concern among social scientists engaged in this research centers on developing methods to accurately estimate the use of race in both stop and post-stop decisions, where the primary post-stop decisions include the decision to cite, search, or arrest a stopped motorist (e.g., see Farrell et al., 2005; Fridell 2004; Engel and Calnon, 2004).

The growing interest to incorporate traffic enforcement data into EI systems is because EI systems are becoming more prevalent within police departments, and secondly, because an EI system can capture the potential varying uses of race among officers, which a department-level analysis cannot capture. EI systems are part of a broader police reform effort (Walker et al., 2005). The systems collect and analyze officer-level data on performance indicators such as use-of-force reports, resisting arrest charges, vehicular pursuits, vehicular accidents, citizen complaints, and officer involvement in civil litigation. An officer's performance is evaluated based on various criteria such as comparing his performance to a threshold or to a peer group of similarly situated officers. To be similarly situated, the officers must patrol the same geographical area during the same time periods with the same assignment (e.g., traffic patrol). Based on these comparisons, problem officers are identified and referred to interventions such as counseling, coaching, and training, all of which occur outside the formal disciplinary system. The EI system's goal is to address problems before they become serious.

However, the results of a department-level study do not identify whether particular officers are using race as a factor in their traffic enforcement decisions. For example, if a department-level study finds evidence that race is being used on average, the police leadership is not able to determine whether its use varies among officers. The leadership's corrective action would differ if it knew almost all officers were using race versus a situation where only a few officers were identified as using race. Moreover, if a department-level study finds no evidence of the use of race, the actions of a few officers who are using race will likely be hidden.

On the other hand, most officer-level studies estimate each officer's relative use of race as compared to his peers, who may not provide an appropriate benchmark to measure against since they may be using race as a factor in their traffic enforcement decisions. Hence, the results from an officer-level study are more informative when they are coupled with the results from a department-level study, which estimates the average use of race at the department level (i.e., the average use of race for all officers combined).

Whether the use of race is estimated at the department or officer level, a key concern among social scientists is to develop methods that accurately estimate the use of race (e.g., see Farrell et al., 2005; Fridell 2004; Engel and Calnon, 2004). For department-level studies, the key issue in the stop analysis is to estimate each racial group's share of motorists violating a traffic law that would result in a stop. These shares are referred to as the "external benchmark" since they theoretically represent what the racial-group shares of stopped motorists would be if race were not used as a factor in deciding which motorists to stop, assuming each racial group had the same types of violations and exposure to the police. The external benchmark has been estimated using various reference benchmarks such as the U.S. census, traffic observers, DMV data, traffic survey data, accident data, daylight, and aircraft- or radar-measured speeding stops (Fridell, 2004). However, the reference benchmarks vary in their accuracy at estimating the external benchmark and also have varying costs to implement. For department-level post-stop analyses, the primary methodological issue is to control for contextual and motorist characteristics that are associated with a racial group that also influence an officer's decision to cite, search, or arrest a motorist. Without controlling for these characteristics, the estimated use of race will be inaccurate.

This study improves upon the existing methods to estimate the use of race at the department level. To estimate the department's average use of race as a factor for deciding which motorists to stop, I use radar-measured speeding stops to estimate each racial group's share of motorists violating a traffic law since an officer has a degraded ability to detect a motorist's race when using radar (Lovrich et al., 2005; Moose, 2002). The study improves upon the existing radar methodology by allowing for the possibility that radar may be used at different intensities over the evaluation period and that non-racial motorist characteristics may influence stops decisions. For the post-stop analyses, I will improve upon an existing method by distinguishing between non-DUI and DUI incident-to-arrest searches since the events leading to a DUI arrest involve more officer discretion as compared to a typical incident-to-arrest search.[8] To analyze search decisions, I will use existing methods. But for DUI arrest decisions, I begin with an existing methodology used to compare each racial group's share of searches that result in contraband being found (see Knowles, Persico, and Todd, 2001) and adapt it to compare each racial group's share of DUI arrestees who test above the legal alcohol concentration level by taking advantage of the data that approximately identifies marginal DUI arrestees.

Due to the paucity of officer-level studies, the racial profiling literature that analyzes traffic enforcement data at the individual officer level is in the early stages of development, and moreover, the literature that analyzes EI system performance indicators is also in the early stages of development. In the existing officer-level studies (e.g., Smith, 2005; studies cited in Fridell, 2004), the most common empirical methods used to compare similarly situated officers are somewhat limited if officers' work schedules or patrol areas (e.g., neighborhood or roadway) change over time. If a study is limited to when a particular group of officers patrolled together in a particular geographical area, the sample may be restricted to a small subset of their stops, which reduces the ability to detect officer differences. If the schedule and patrol area changes are ignored, then the results could identify the wrong officers as problem officers due to officers working somewhat different schedules or patrolling somewhat different geographical areas than their peers.

To address these issues, Riley et al. (2005) developed a method that compares a particular officer's stops to other officers' stops that occurred at the same time and location; hence, the method is not restricted to when a particular group of officers patrolled together. An officer's performance is based on his average performance as compared to his peers' stops that occurred during the times and locations he patrolled. I will build on that study and allow for the possibility that an officer and his peers differentially use non-racial motorist characteristics to influence their traffic stop decisions (e.g., one officer focuses on seatbelt violators) as well as their post-stop decisions (e.g., one officer focuses on lane change violators because they are associated with driving under the influence).

[8] DUI stands for driving under the influence of alcohol or drugs.

In summary, my research objectives focus on discussing the implications of incorporating traffic enforcement data into an EI system and improving upon the methods to analyze the data in order to identify problem officers. This requires both estimating the officers' combined average use of race as well as each officer's relative use of race as compared to his similarly situated peers. I focus on estimating the use of race as a factor in deciding which motorists are stopped as well as which motorists are searched or arrested for a DUI as a result of the stop. Specifically, the study's objectives include the following tasks.

- Discuss the key differences between estimating the use of race as a factor in traffic enforcement decisions at the department versus officer level, and summarize the key implications of incorporating traffic enforcement racial profiling analyses into an EI system
- Improve upon existing methods to estimate the use of race in stop, search, and DUI arrest decisions at both the department and officer levels
- Assess how well different empirical methods are able to identify officers who potentially racial profile, and more generally, identify potential problem officers with respect to other EI indicators

The empirical work will be based on analyzing traffic enforcement data from the Washington State Patrol (WSP), which Lovrich et al. (2003, 2005) analyzed from a department-wide perspective. This study will supplement and complement those department-level studies.[9]

1.3 Organization of Dissertation

The remainder of this study includes six chapters. The next chapter defines discrimination, analyzes laws that address racial profiling, discusses EI systems, and reviews the racial profiling literature that covers the key methods used to estimate the use

[9] In Chapter 2, I discuss the use of race as a factor in deciding which motorists to cite; however, an empirical estimate of the use of race in citation decisions is outside the scope of this study. The estimate of the use of race in citation decisions is sometimes difficult to interpret, which will be discussed in the literature review.

of race as a factor in traffic enforcement decisions at the department and officer levels. Chapter 3 is a detachment-level analysis that estimates 16 similarly situated Washington State Patrol troopers' average use of race as a factor in deciding which motorists to stop in the South Seattle area. The chapter includes an introduction followed by sections that describe the data, methods, and results. The chapter concludes with a discussion. Chapter 4 is organized similar to Chapter 3, but instead estimates each trooper's relative use of race as compared to his peers in deciding which motorists to stop. Chapters 5 and 6 are respectively analogous to Chapters 3 and 4, but estimate the use of race as a factor in deciding which stopped motorists are searched or arrested for a DUI. Chapter 7 summarizes the results and concludes with policy recommendations.

Chapter 2: Literature Review of Racial Profiling Law, EI Systems, and Department- and Officer-Level Studies

The key policy questions police departments face are whether to analyze traffic enforcement data in order to estimate the use of race at the officer level (in addition to department level), and if so, how to incorporate the data into an EI system to accurately identify potential problem officers. A key factor in the policy decision is whether adequate methods exist to accurately estimate the use of race at the officer level. To address these questions, this chapter is divided into three sections. The first section discusses two types of discrimination and analyzes the laws that address racial profiling, which will be used to determine whether the use of race as a factor in a law enforcement decision is appropriate or inappropriate. The second section discusses Early Intervention (EI) systems, which are designed to identify problem officers and improve officer performance. The third section reviews the racial profiling literature and discusses the key methods used to estimate the use of race at the department and officer levels as well as discusses the key differences between the estimates. It also summarizes the implications of incorporating traffic enforcement racial profiling analyses into an EI system.

2.1 An Economic and Legal Overview of Racial Profiling

2.1.1 Economic Overview of Racial Profiling

When a racial profiling study within traffic enforcement controls for confounding characteristics (i.e., characteristics that are associated with a racial group that also affect the probability of a motorist being stopped, cited, searched, or arrested) and still finds racial disparities in motorist stop, citation, search, and arrest rates, the disparities are often attributed to racial animosity. However, the disparities may arise from the police's use of race as a proxy for criminal activity, which would also lead to racial disparities in traffic enforcement decisions. This is known as statistical discrimination. Furthermore, disparities may arise due to confounding characteristics that have not been controlled for, typically because they are not observed (by the analyst). The following section discusses statistical discrimination and racial animosity from an economic perspective.

From an efficiency perspective, society's objective is to minimize the social costs of crime, which include offending, enforcement, and penalty costs (Becker, 1968). There are tradeoffs among these costs. For example, as enforcement costs increase, the probability of being caught increases, which decreases the probability of offending, thus reducing the offending costs. The optimal level of enforcement can theoretically be determined.

In order for law enforcement to increase their efficiency at a given enforcement level, they might discriminate against particular groups if the propensity to offend differs among groups. This theory of discrimination is based on incomplete information (Arrow, 1973). In this case, law enforcement does not have perfect information on each person's propensity to offend, and perfect information is costly to obtain. For example, assume there are two characteristics that are associated with each other. The first characteristic is causally related to an outcome of interest, but is relatively costly to measure, while the second characteristic is relatively inexpensive to measure. Even if the second characteristic is not causally related to the outcome of interest, it can be used as a proxy for the more expensive characteristic. This is called statistical discrimination. Statistical discrimination occurs when employers use race (an inexpensive characteristic to measure) to determine whom to hire when they believe race is associated with an expensive characteristic to measure (e.g., quality of education) that causes productivity differences. With respect to policing, race may be associated with an expensive characteristic to measure (e.g., legal employment opportunities) that causes differences in the probability of engaging in criminal activity. Hence, police might use race as one factor in their traffic enforcement decisions if it has the power to predict more serious crimes that may be discovered during a traffic stop.[10]

Alternatively, law enforcement might discriminate against a racial group due to racial animus (hereafter "animosity"). In the labor market, Becker (1957) defines this type of discrimination as when economic actors are willing to pay an economic cost to avoid interacting with particular races. For example, given that a black worker has the same productivity as a white worker, then an employer who discriminates against blacks is willing to pay a relatively higher wage to the white worker. In the case of policing,

[10] Whether it is legal for the police to use race in this manner is discussed below.

racial animosity often surfaces as harassment. For example, an officer who has racial animosity toward blacks is willing to spend time stopping, searching, and harassing relatively unsuspicious black motorists at the cost of not stopping and searching relatively more suspicious white motorists.

Racial disparities found in traffic enforcement studies may be due to racial animosity, statistical discrimination, or confounding characteristics. Because experiments are not feasible in this area of research, it is often difficult to separately identify these sources. Most of the methods estimate the combined use of race due to racial animosity and statistical discrimination by controlling for the confounding characteristics.[11] The total influence of racial animosity and statistical discrimination is often referred to as racial profiling.

2.1.2 Legal Overview of Racial Profiling

In order for a police officer to stop, cite, search, or arrest a motorist, he must establish probable cause[12] that an offense has been, is being, or will be committed to warrant his action. Whether (and the extent) that race may be used as a factor to establish probable cause (or reasonable suspicion)[13] depends on the following four factors (e.g., see Smith, 2005; DOJ, 2003; Alschuler, 2002; Gross and Barnes, 2002; Kennedy, 1997).

---

[11] An outcomes-based method will be introduced below that attempts to estimate the use of race due to racial animosity, but does not identify whether race is being used due to statistical discrimination. This method has only been used to evaluate searches.

[12] The Fourth Amendment protects individuals from being subjected to unreasonable searches and seizures without probable cause. Tomkovicz and White (2001) discuss the probable cause requirement of the Fourth Amendment. *Brinegar v. United States* (1949) states that "'The substance of all the definitions' of probable cause 'is a reasonable ground for the belief of guilt'" (quoting McCarthy v. De Armit, 99 Pa. St. 63, 69). Although probable cause abstractly deals with probabilities, no specific probability of guilt has been specified to establish probable cause. *Brinegar* states that to establish probable cause, the evidence must be more than mere suspicion but can be less than needed to justify condemnation. In order to establish probable cause for an arrest, an officer needs to have reasonably trustworthy information that would lead a prudent man to believe an offense had or was being committed (*Beck v. Ohio*, 1968). Similarly for a search, probable cause is established if an officer has reasonably trustworthy information that would lead a prudent man to believe an "item subject to seizure will be found in the place to be searched" (*United States v. Garza-Hernandez* (Seventh Circuit Court of Appeals, 1980), which references *Brinegar*).

[13] In *Terry v. Ohio* (1968), the Supreme Court created a new legal category, the stop-and-frisk search, in order to balance an individual's rights under the Fourth Amendment with an officer's safety. A stop and frisk is still considered a seizure and search, but is limited to stopping and frisking an individual for weapons that may pose a danger to the officer. The evidentiary standard needed to conduct the stop and frisk has become known as reasonable suspicion, which is a lower standard than probable cause. *Terry* states, "There must be a narrowly drawn authority to permit a reasonable search for weapons for the protection of the police officer, where he has reason to believe that his is dealing with an armed and dangerous individual, regardless of whether he has probable cause to arrest the individual for a crime."

- Whether the Equal Protection Clause's strict scrutiny standard or the Fourth Amendment's reasonableness standard is applied
- How important the state's interest is, that is, the importance of the objective that law enforcement is attempting to meet
- How much influence race has in the probable cause decision, ranging from one of many factors to the sole factor
- How specific the suspect's description is, that is, whether race is part of a specific suspect's physical description, part of a criminal profile, or part of a general statistic concerning offending rates

Many studies show that the vast majority of motorists are violating a traffic law that justifies a stop (e.g., Lamberth, 1994). The factors above also influence whether (and the extent) race may be used as a factor in deciding which of these motorists to stop.

In this section, I discuss the four factors above and develop a figure that plots court cases with respect to the factors (see Figure 2.1). The two primary federal standards used to determine whether it is legal to use race as a factor to establish probable cause or reasonable suspicion originate from the Fourteenth and Fourth Amendments.[14] The Fourteenth Amendment's Equal Protection Clause uses the strict scrutiny standard when a law or policy classifies individuals based on race. In order for the law or policy to be upheld, the standard requires a compelling state interest and that the classification is necessary, or narrowly tailored, to further that interest. The Fourth Amendment, or the reasonableness standard, asks whether using race as a factor to establish probable cause or reasonable suspicion is reasonably related to efficient policing (Smith, 2005; Kennedy, 1997). Note that states may enact laws that further restrict the use of race as compared to the federal standards.

Both the strict scrutiny and the reasonableness standards involve balancing the state's interest with individual rights. When applying either standard, the state's interest is an important consideration. And the state's interest can vary widely depending on the

---

These stop-and-frisk searches are commonly referred to as "Terry" searches. (In addition, a stop-and-frisk search is permissible if an officer establishes reasonable suspicion that a crime has or is about to occur.)

[14] For a discussion of other laws in this area such as Title VI of the Civil Rights Act of 1964, see Alpert (2004).

circumstances, for example, from enforcing minor traffic violations to reducing the illicit drug trade to preventing a catastrophic terrorist attack in a time of war.

The strict scrutiny standard is a much higher standard for the government to meet as compared to the reasonableness standard. The Equal Protection Clause governs actions when a state or local government classifies individuals, which results in different benefits and burdens under the law.[15] The strict scrutiny standard originates from *Korematsu v. United States* (1944). As a response to the Japanese attack on Pearl Harbor, the federal government ordered people of Japanese ancestry to relocate to internment camps. In *Korematsu*, the Supreme Court upheld this policy, but also established the constitutional test, known as the strict scrutiny standard, as to whether a law that classifies individuals based on race will satisfy the Equal Protection Clause. If a law classifies individuals based on a suspect class such as race, ethnicity, national origin, religion, or alienage, then the court will apply a strict scrutiny standard and the burden is on the state to prove that there is a compelling state interest and that the classification is necessary (i.e., the least discriminatory means) to further that interest. When this standard has been applied, most racial classifications have not survived this test (Smith, 2005).[16] The Supreme Court has struck down laws that mandated or permitted segregated education (e.g., *Brown v. Board of Education of Topeka* (1954)), which in turn followed striking down laws pertaining to segregated buses, parks, athletic contests, courtroom seating, and municipal auditoriums (Nowak and Rotunda, 2000). The standard is high because the state must show it has a compelling state interest, how the use of race is related to the state's interest, and whether there are other available means to further the state's interest.

As with the strict scrutiny standard, the reasonableness standard also involves balancing the state's interest and the individual's rights; however, the balance is less burdensome on the state. *Delaware v. Prouse* (1979) states, "The permissibility of a particular law enforcement practice is judged by balancing its intrusion on the individual's Fourth Amendment interests against its promotion of legitimate governmental interests." For this standard, the government neither needs a compelling

---

[15] The due process clause of the Fifth Amendment imposes a similar equal protection requirement on the federal government.
[16] Justice Marshall states in *Fullilove v. Klutznick* (1980) that strict scrutiny analysis is "strict in theory, but fatal in fact" (as cited Smith, 2005).

state interest nor needs to show that the use of race is narrowly tailored to meet that interest.

In racial profiling cases, whether the strict scrutiny or the reasonableness is applied depends on the available evidence regarding the use of race as well as what court hears the case. Kennedy (1997) states that the Supreme Court and other lower courts have failed to apply strict scrutiny to police using race as a factor of suspicion, including its use in search and seizure decisions. In order for the strict scrutiny standard to be applied, there is a basic principle that the burden is on the defendant to prove "the existence of purposeful discrimination," which also implies proving a discriminatory effect (*McCleskey v. Kemp*, 1987).[17] Although evidence from a study showed Georgia's death penalty sentences were associated with black defendants, the aggregate statistical evidence was not sufficient to prove purposeful discrimination in McCleskey's particular death penalty sentence. *United States v. Armstrong* (1996) reaffirmed *McCleskey's* two-prong test of discriminatory purpose and effect. On a practical level, proving a discriminatory purpose is difficult to do since a person must show that a similarly situated person of another race was treated differently. In a racial profiling case involving traffic enforcement, the individual typically cannot obtain this data (Smith, 2005).[18]

Racial profiling cases have also sought relief under the Fourth Amendment. If the use of race is considered unreasonable, then remedies for a criminal defendant are better defined based on the Fourth Amendment's exclusionary principle where evidence gained in violation of the Fourth Amendment is suppressed (*Weeks v. United States*, 1914; *Mapp v. Ohio*, 1961). However, the Fourth Amendment cases have not had consistent rulings (Gross and Barnes, 2002).

In addition, the particular standard that has been applied somewhat depends on the court that hears the case. This may be due to different laws that exist at the federal and state level as well as differing opinions at the federal appellate court level due to the law being unsettled. However, after *Whren v. U.S.* (1996) ruled that an officer's

[17] *McCleskey* quotes *Whitus v. Georgia* (1967). Although *McCleskey* finds that statistical evidence will typically not be sufficient to prove intent, it states there may be situations where a stark statistical discrepancy proves intent (e.g., it references *Yick Wo v. Hopkins* (1886)).
[18] In some cases, an officer may admit to using race as a factor (e.g., when race is one characteristic within a criminal profile).

subjective intentions for making the stop are irrelevant,[19] most racial profiling cases now need to be contested on Equal Protection grounds (Smith, 2005).

To illustrate these legal standards as well as the other dimensions that influence the legality of the use of race as a factor to establish probable cause or reasonable suspicion in a traffic enforcement setting, the Figure 2.1 illustrates the legality of the use of race across the range of the critical dimensions: the influence of race, the suspect's specificity, and the state's interest (which includes the legal standard that is applied). The figure is meant to serve as a general conceptual model. To aid the model, I plotted specific court cases to illustrate how the four dimensions affected the outcomes of these cases. Although each case represents a point on the figure, this gives a false impression of precision since each case includes unique characteristics that influence the court's decision. Hence, each case should actually represent a small area on the figure.

The horizontal axis represents the influence of race as a factor to establish probable cause or reasonable suspicion. The influence ranges from low, where race is used as one factor among many, to high, where race is used as the sole factor. The vertical axis represents the specificity of the suspect that law enforcement is attempting to apprehend. The suspect's specificity ranges from particular to general. A suspect's specificity is particular when police are seeking a particular suspect with a physical description who is wanted for a particular crime. A suspect's specificity is more general in a criminal profile where police are seeking suspects who fit a criminal profile that has been developed based on observing a pattern of criminal activity. A suspect's specificity is most general when police are seeking suspects based on one or more characteristics that are predictive of general criminal activity. In this case, neither a specific crime nor a specific pattern of criminal activity has occurred, but instead police actions are being

---

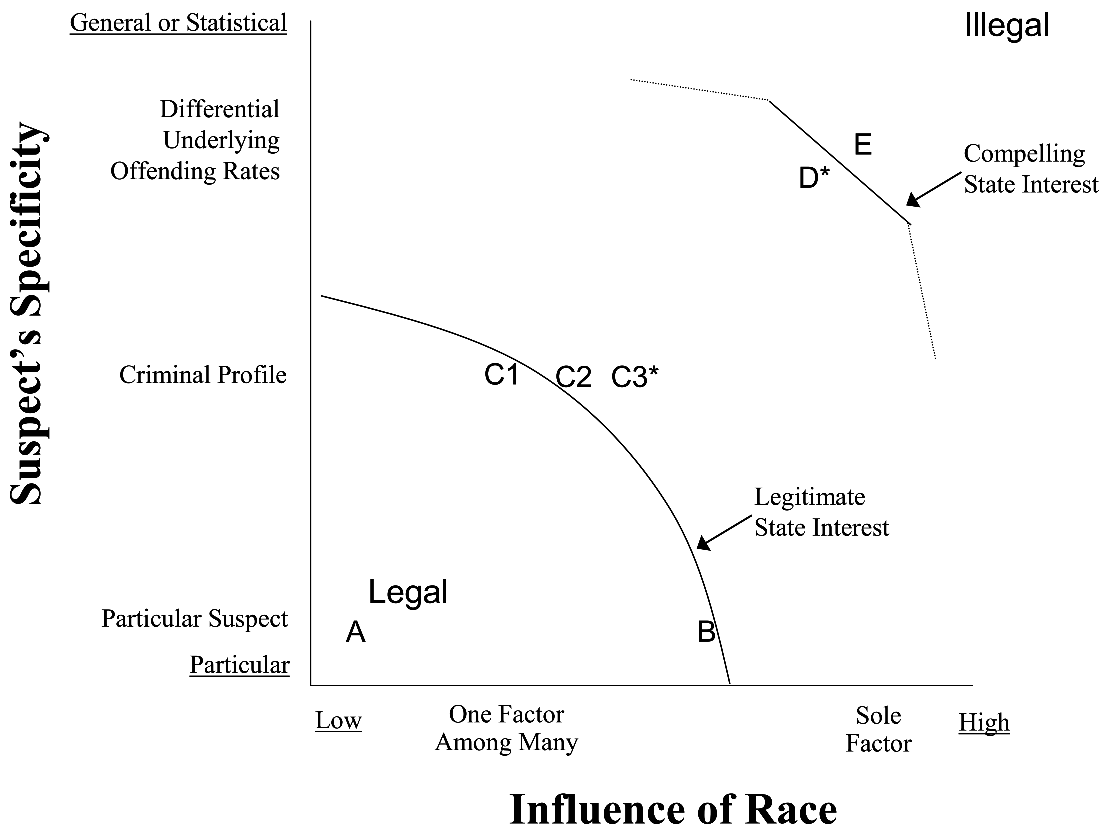
[19] Pre-text stops are known as traffic stops that occur when the officer has probable cause to make a traffic stop due to a traffic violation; however, the officer's real purpose for making the stop is due to suspecting the motorist of a more serious crime (e.g., drug trafficking) for which he lacks probable cause. During the stop, the officer will be better able to ascertain whether he can establish probable to search or arrest the motorist. *Whren v. U.S.* (1996) addressed pre-text stops and found that an officer's subjective intentions for making a stop are irrelevant in deciding on whether a stop violated the Fourth Amendment's reasonableness standard. If an officer establishes probable cause for a traffic violation that calls for a stop, then the stop is considered reasonable. Although, the Fourth Amendment serves as a minimum standard of rights that the states must afford individuals, the states may provide additional rights. For example, based on *State v. Ladson* (1999), the state of Washington requires that an officer to have "clean thoughts" when making a stop, thus, does not permit pre-text stops (Loginsky, 2005).

driven by different criminal incidence rates among individuals' characteristics such as race, sex, or age.

The two curved lines represent two hypothetical state interest levels; however, the state interest level should be thought of as continuous.[20] The state interest increases from the lower left curve to the upper right curve. Remember, the state's interest includes a wide range, for example, from enforcing minor traffic violations to reducing the illicit drug trade to preventing a catastrophic terrorist attack in a time of war. Due to the wide range of the state's potential interests, the legality of the use of race primarily depends on the state's interest and secondarily depends on the influence of race and the suspect's specificity. The use of race is considered to violate the Constitution if the point defined by the intersection of the influence of race and suspect's specificity lies to the lower left of the state interest that is being applied. And if the point lies to the upper right of the state interest, then its use is considered legal. The shape of the curves emphasizes that as the influence of race increases, then the suspect's specificity must increase at a higher rate. And similarly, as the suspect's specificity becomes more general, then the influence of race must decrease at a higher rate. For illustrative purposes, the cases are plotted near one of the two state-interest curves; however, in reality the cases involve different levels of state interest.

[20] The legal standard (i.e., strict scrutiny or reasonableness standard) is embedded in the state interest curve. For example, assume that a particular law is deemed to have a legitimate state interest. If the strict scrutiny standard is applied, the law will be struck down because the state does not have a compelling interest. For a compelling state interest, the state-interest curved line lies to the far upper right of the figure. However, if the reasonableness standard is applied, the law will likely be upheld.

Figure 2.1: The Legality of Using Race as a Factor to Establish Probable Cause or Reasonable Suspicion



*Indicates that the strict scrutiny standard was applied; otherwise the reasonableness standard was applied.

The letters on the figure represent federal and state cases, and for illustration purposes, also represent hypothetical cases. Point "A" represents a hypothetical case. For example, assume a person calls the police to report a purse-snatching incident and states that the suspect is a white male, mid-20s, six-feet tall, wearing a green jacket who just pulled away in a white Ford pickup truck from a particular address. In this case, race may be used as a factor to establish probable cause in order to stop and search motorists who fit the above description near the described location. In this situation, race is one factor among many and a specific suspect is being sought for a specific crime. Using race as a factor to establish probable cause is reasonably related to efficient policing.

Point "E" represents the opposite extreme. For example, assume a person calls the police stating he read that blacks commit more minor crimes per capita as compared to other racial groups. In this case, race may not be used as a factor, for example, to selectively enforce traffic laws against black motorists in order to scrutinize them during the stop for evidence of serious criminal activity. In this case, there is no specific crime and race is being used as the sole factor in deciding which motorists to stop. This use of race would not survive the reasonableness standard of the Fourth Amendment because the use of race is not reasonably related to efficient policing. It would also not survive the strict scrutiny standard because reducing minor crimes is not likely to be considered a compelling state interest and the use of race in this fashion is not the least discriminatory means to further the state's interest of reducing minor crimes.

Point "D*", which lies near point "E" represents *Korematsu v. United States* (1944), where the use of Japanese ancestry was upheld (see case described above). There is debate whether a similar case would be upheld today. While there was a compelling state interest, the question is whether the policy of using ancestry was as narrowly tailored as possible to further that interest.

The next example (see point "B") involves *Brown v. City of Oneonta* (Second Circuit Court of Appeals, 2000), where the court upheld the police's use of race as the predominant factor in deciding whom to stop and question in order to apprehend a burglar. A burglar broke into a 77-year-old woman's home, and although the woman did not see the burglar's face, she identified him as being black based on seeing his hands and forearms. She also said that from his movements, she thought he was young. During their struggle, the burglar apparently cut his hand with his knife. The woman lived in Oneonta, New York, a town of 10,000 permanent residents, including approximately 300 black residents. Additionally, the local state university had a student population of 7,500, including 150 black students. The Oneonta police questioned black, male students based on the state university's list of black, male students. After they discovered no suspects, they began stopping, questioning, and examining the hands of non-white persons on the street, totaling approximately 200 persons. Black, male students and others questioned by the police sued for alleged civil rights violations. However, the court stated that because the police were not alleged to have investigated solely on the basis of race, there was "no actionable claim under the Equal Protection Clause." Alschuler (2002) disagrees and thought the above classification, which was based on race, sex, and age (to some extent) was not narrowly tailored. And Smith (2005) argues that it would not have survived the strict scrutiny test had it been applied.

The next examples (see point "C1") involve a pattern of criminal activity that continues to occur where there is not a specific suspect being sought, but instead the suspects include individuals who are linked to the particular criminal scheme or organization.[21] If the vast majority of individuals linked to the criminal activity share particular physical characteristics, then those characteristics may be sufficient to produce a useful physical description, or what is called a criminal profile. The use of a profile that included the ancestry of illegal aliens was upheld in *United States v. Martinez Fuerte* (1976), and the use of a profile that included the race of drug traffickers was upheld in *United States v. Weaver* (Eighth Circuit Court of Appeals, 1992) and *United States v. Condolee* (Eighth Circuit Court of Appeals, 1990).[22]

On the other hand, there are cases (see point "C2") that have suppressed evidence gained where race is one factor among many in the criminal profile (e.g., *United States v. Brignoni-Ponce* (1975), *United States v. Montero-Camargo* (Ninth Circuit Court of Appeals, 2000), *Lowery v. Virginia* (Virginia Court of Appeals, 1990), and *United States v. Laymon* (730 F. Supp. 332 (D. Colo. 1990)).[23] However, Kennedy (1997) and Smith

[21] Note that points "B," "C1," "C2," and "C3" are compared to the same state interest. This was done for convenience, but in reality, the state's interest in each case is somewhat different.

[22] In *United States v. Martinez Fuerte* (1976), U.S. Border Patrol agents had established a highway checkpoint located 30 miles north of the California-Mexico border to interdict motorists transporting illegal aliens. The agents' suspicion of illegal aliens increased if the driver was of Mexican ancestry, increasing their likelihood of searching the vehicle. The U.S. Supreme Court upheld a conviction stating that the reliance on apparent Mexican ancestry is relevant to the border patrol's objective (Kennedy, 1997). Similarly, in the *United States v. Weaver* (Eighth Circuit Court of Appeals, 1992), a drug enforcement agent at the Kansas City, Kansas airport stopped a person because he fit the profile of a drug courier (e.g., flew in from Los Angeles where cocaine had been originating from, looked "roughly dressed," appeared nervous, had two carry-on bags with no checked baggage, and was black). The U.S. Court of Appeals upheld the conviction because race was one relevant factor among many in the intelligence-based profile. *United States v. Condolee* (Eighth Circuit Court of Appeals, 1990) is a very similar case to *Weaver*. Furthermore, Kennedy (1997) notes that in *State v. Dean* (1975), the Arizona Supreme Court ruled that a person's race could be used as a factor to establish probable cause to stop a motorist, but stated that race could not be the sole factor. In this case, a Mexican male sat in a parked car outside an apartment complex in a predominantly white neighborhood, appeared nervous, and moved his car when a marked police car approached his vehicle.

[23] *United States v. Brignoni-Ponce* (1975) involved the U.S. Border Patrol stopping a car near the Mexican border. In this case, the stop was ruled unconstitutional because race was used as the sole factor for the stop (Smith, 2005). In *United States v. Montero-Camargo* (Ninth Circuit Court of Appeals, 2000), the court announced in dicta that the use of race as a factor in deciding which motorists to stop violates the Fourth Amendment (Smith, 2005). In *Lowery v. Virginia*, the Virginia Court of Appeals ruled that the law requires a compelling (not just a reasonable) justification to use race as a factor in establishing probable cause (Kennedy, 1997). In *United States v. Laymon* (1990), the court suppressed incriminating evidence where an officer allegedly used race as a factor to establish probable cause to search the motorist's vehicle. The court ruled that the officer had not established probable cause to search the motorist's vehicle (Kennedy, 1997).

(2005) state these cases are either in the minority or do not represent the dominant opinion.

Point C3* represents *State v. Soto* (New Jersey Supreme Court, 1996), which is the only racial profiling case involving stops where the strict scrutiny standard has been applied (Gross and Barnes, 2002). The case involved 17 defendants who had been arrested for drug trafficking after being stopped and searched by the New Jersey State Police. The court suppressed the evidence partly due to an observational study of New Jersey Turnpike motorists, where blacks were found to represent 14.5 percent of the violators, but represented 35 percent of the stops and 73 percent of the searches (Lamberth, 1994). Although the use of race may have met the reasonableness standard, this was irrelevant since the strict scrutiny standard was applied, which it did not meet.

As seen above, although there have been some exceptional court decisions, the dominant legal view is that a person's race can be used as a factor to establish probable cause or reasonable suspicion if its use is reasonably related to efficient policing, is one factor among many, and is not used as a pretext for harassment (Kennedy, 1997). Although this view prohibits race from being used as a factor as a pretext for harassment, inequitable outcomes may still result. Assuming that the legal use of race leads to efficient policing, a net benefit results. However, those who benefit from the use of race may not be the same individuals as those who are burdened by its use. Individuals benefit if they would have been harmed by a crime (or paid for additional security) had the use of race not occurred. Individuals are burdened if their civil liberties are curtailed due to the use of race. In high-crime neighborhoods, those who are burdened by the use of race are sometimes the same individuals who benefit from its use. However, it is also the case that the individuals who benefit and the individuals who are burdened do not always include the same individuals.

In recent years, many states have passed legislation banning racial profiling; however, what actually constitutes racial profiling varies by state. Amnesty International USA (2004) surveyed the 23 states that had laws banning racial profiling. They found that 11 states do not allow race, ethnicity, or national origin to be used as a factor in deciding which motorists to stop, while the other 12 banned using these characteristics as the sole factor. For example, California prohibits casting suspicion on an entire class of people without any individualized suspicion of the particular person being stopped, and Connecticut prohibits race from being used as the sole factor. Pending federal legislation would prohibit race from being used to any degree.

Because "racial profiling" is a politically charged and ill-defined term, this study uses the term the "use of race." The use of race includes the use of race due to both racial animosity and statistical discrimination, except where racial animosity can be separately identified. While the use of race due to racial animosity is clearly inappropriate, the appropriateness of the use of race due to statistical discrimination depends on the factors discussed above: the state's interest, the influence of race, and the suspect's specificity.[24]

2.2 Overview of Police Reform and EI Systems

Although the police enforce the law for the betterment of society, they are also bound by the law, including the use-of-race laws defined above. Due to the substantial power given to the police to enforce the law, many internal and external oversight systems have been implemented to regulate the use of that power. Perez (1994) compared internal, civilian, and civil monitor review systems and states that an effective police review system needs to have integrity, legitimacy, and a learning component. Integrity means that the system will evaluate performance and investigate complaints fairly, thoroughly, and objectively. Legitimacy is the perception of the system's integrity, from various stakeholders such as police officers, police leadership, and the community. The learning component includes whether officers learn from the system, whether it deters inappropriate behavior, and whether it identifies problem officers.

Although Perez's study involved comparing internal, civilian, and civil monitor review systems, these principles are applicable to any organization's reform system. Corporations have invested billions of dollars in enterprise resource systems, which collect and analyze financial, operational, and human resource data in order for management to make better-informed decisions. Within a police department, these systems are known as Early Intervention (EI) or Early Warning (EW) systems. The systems have historically been known as EW systems because they were primarily

[24] As stated in Chapter 1, in this study, for brevity, when I use the term "use of race," it mostly includes the inappropriate use of race unless stated otherwise. This is because the law rarely allows an officer to use race as a factor in most law enforcement decisions related to routine traffic stops.

designed to *warn* police leadership about problem officers, such as officers who inappropriately use force or generate a large number of citizen complaints. Walker et al. (2005) state that an early *intervention* system better describes the ideal system, which should be part of a police department's overall effort to improve officer performance, where the interventions are not considered punitive. Moreover, Walker and Alpert (2000) state that the systems should have an impact at all levels of the police department, not just at the line officer level. At the department level, trends in areas such as the use of force or citizen complaints can be tracked. At the sub-department level, supervisors can be encouraged to monitor their officers' performance and intervene when necessary.

The use of EI systems is increasing across law enforcement agencies (Walker et al., 2005). In *Principles for Promoting Police Integrity*, the U.S. Department of Justice recommended EW systems to promote police accountability and effective management (DOJ, 2001). Additionally, when agencies have settled civil rights lawsuits with the Civil Rights Division of the U.S. Department of Justice, the consent decrees and memoranda of understanding have mandated agencies to implement EI systems. These agencies include the Pittsburgh Police Bureau, the New Jersey State Police, the Metropolitan Police Department of Washington, D.C., the Cincinnati Police Department, and the Los Angeles Police Department (Walker, 2003). In fact, the Los Angeles Police Department's consent decree was recently extended for three years, in part because its EI system, TEAMS II, was not fully operational (Orlov, 2006).

The components of an EI system include collecting officer-level data on various performance measures; analyzing the data and developing criteria to determine when an intervention is warranted; and developing and implementing appropriate interventions (e.g., counseling, training, reassignment). I will discuss each of these in turn.

While most EI systems collect data on potentially problematic behavior such as use-of-force reports, resisting arrest charges, vehicular pursuits, vehicular accidents, citizen complaints, and officer involvement in civil litigation, some of the systems also track positive behavior. For example, the Pittsburgh Police Bureau's EW system is called the Performance Assessment and Review System (PARS). The system tracks a large range of performance measures such as use-of-force reports and citizen complaints, but also tracks traffic stop data, searches, and arrests as well as awards, commendations, and recognition (Davis et al., 2002).

The data collected within an EI system are analyzed to identify officers who warrant an intervention. The objective is to develop criteria that will identify all officers who warrant an intervention, while at the same time, not falsely identify officers who do not warrant an intervention. Hence, the criteria should minimize misidentifying officers, which includes falsely identifying non-problem officers and failing to identify problem officers. In a statistical hypothesis test, these are referred to Type I and Type II errors, respectively. Because an EI system does not have complete information and its data are subject to measurement error, no set of criteria will perfectly identify the appropriate officers. For any given set of criteria, there is typically a tradeoff between Type I and Type II errors. For example, criteria that result in identifying all problem officers will also likely identify some non-problem officers, and conversely, criteria that result in never falsely identifying any non-problem officers will also likely fail to identify some problem officers.

Due to this tradeoff and because many EI systems have recently been developed, there is no consensus regarding the criteria that should be used to identify problem officers. However, the most common methods include thresholds, peer officer comparisons, and ratio-based formulas (Walker, 2003b). For example, the Miami-Dade Police Department EI system uses thresholds to identify potential problem officers. If an officer has two or more personnel (e.g., citizen) complaints or three or more use-of-force reports during a quarter, then the officer must be reviewed by his supervisor.

However, when department thresholds are used, they may falsely identify officers who have a different type of assignment that may lead to a higher number of citizen complaints or uses of force; hence, peer officer comparisons are used by police departments. The peer officer comparison method compares officers who are similarly situated, that is, officers who patrol the same geographical area during the same time periods with the same assignment (e.g., traffic patrol). If an officer's performance differs significantly from his peers, then that officer would be referred to an intervention. While this comparison will likely perform better at identifying the appropriate officers (as compared to using than department-wide thresholds), there may be differences within the peer group that explain an officer's performance.

A performance indicator ratio is used to account for different activity levels among officers. For example, if an officer is involved in more arrests than his peers, he is more likely to have more citizen complaints and uses of force due to the higher number of contacts with the public. The Miami Dade Police Department's EI system also compares the number of use-of-force reports per arrest among officers.

When an EI system identifies a potential problem officer, the next step normally involves an informal interview with a supervisor. The purpose of this interview is to determine whether there were extenuating circumstances (e.g., being on a special assignment that differed from his peers) that explain the officer's unusual performance. If there is still a concern, then the officer is referred to various interventions, which may include one or more of the following: supervisor counseling, peer mentoring, training, crisis intervention teams, reassignment/relief from duty, and professional counseling on family or personal issues (Walker et al., 2005). Interventions are outside of the formal disciplinary system because the EI system's focus is on improving performance. Walker (2001) emphasizes that the data used in an EI system are just a starting point to begin an inquiry, because alone, they are not conclusive. This is because there may be extenuating circumstances that may explain an officer's unusual performance.

The use of EI systems is increasing and they are expanding from focusing on identifying problem officers to becoming more of a management tool. The tool will still identify problem officers, but also enable police leadership and other decision-makers to make more informed decisions. The success of EI systems depends on whether they meet Perez's (1994) principles of an effective reform system, which include integrity, legitimacy, and learning.

2.3 Department- Versus Officer-Level Racial Profiling Analysis

Although racial profiling within traffic enforcement is a serious concern, traffic enforcement data have not typically included officer-level identifiers, precluding it from being incorporated into an EI system at the officer level. Therefore, most racial profiling studies have been conducted at the department level (Engel and Calnon, 2004). A department-level study that estimates the use of race as a factor in traffic enforcement decisions differs from an officer-level study in several respects. First, a department-level study estimates the department's average use of race while an officer-level study typically estimates each officer's relative use of race as compared to his peers. Second, if an officer-level study (along with a department-level study) is incorporated into an EI system, then corrective action can take place at the individual level. Third, analyzing traffic enforcement data at the officer level may reduce officer support for traffic enforcement analyses if officers think the data will be misinterpreted. Fourth, the analysis of traffic enforcement data at the officer level impacts data collection and analysis costs. The racial profiling literature is discussed below, focusing on how it applies to each of these issues.

2.3.1 Nature and Accuracy of Use-Of-Race Estimate

The dominant concern among social scientists engaged in this research centers on developing methods to accurately estimate the use of race in both stop and post-stop decisions (e.g., see Farrell et al., 2005; Fridell 2004; Engel and Calnon, 2004). As discussed above, the three potential sources of racial disparities include racial animosity, statistical discrimination, and confounding characteristics. Because experiments are not feasible in this area of research, it is often difficult to separately identify these sources. Most of the methods estimate the combined use of race due to racial animosity and statistical discrimination by attempting to control for confounding characteristics. The analyst's ability to control for confounding characteristics partly depends on whether the study is done at the department versus the officer level.

This section reviews the literature that estimates the use of race as a factor in stop decisions as well as post-stop decisions such as the decision to cite, search, or arrest a motorist. For the decisions to search or arrest a motorist for a DUI, this section also reviews the literature that estimates the use of race by examining whether the search resulted in contraband being found and whether the DUI arrest resulted in the arrestee testing over the legal alcohol concentration level. For each of the above decisions, the literature review includes department- and officer-level studies.

2.3.1.1 Stop Decisions

Department-level and officer-level studies use different approaches to estimate the use of race as a factor in deciding which motorists to stop.[25] For a department-level study, in order to accurately estimate the use of race as a factor in determining which motorists are stopped for a traffic violation, it is important to accurately estimate each racial group's share of motorists violating a traffic law that would result in a stop,[26] that is, the racial-group shares of motorists at risk of being stopped. These shares are referred to as the "external benchmark" since they theoretically represent what the racial-group shares of stopped motorists would be if race was not used as a factor in deciding which motorists to stop, assuming each racial group had the same types of violations and exposure to the police. The external benchmark has been estimated using various methods: U.S. census, traffic observers, DMV data, traffic survey data, accident data, daylight, and aircraft- or radar-measured speeding stops (Fridell, 2004). The use of traffic observers and the U.S. census are the two most common ways to estimate the benchmark (Farrell et al., 2005).

When a study estimates the benchmark using traffic observers, observers are stationed at various locations (or ride along with traffic) to record each motorist's race and whether the motorist was violating a traffic law (see Lamberth, 1994; Solop, 2002; Rickabaugh, 2003; Alpert, 2004). Whether this approach provides an accurate estimate of the external benchmark depends on whether observers sampled motorists at various locations and times to account for the possibility that the racial-group shares on the road are not constant throughout the jurisdiction over time. As the motorist sample size increases, the cost of the study will increase as well. As a less expensive alternative, the U.S. census is used to estimate the external benchmark and it is often used in conjunction with some of the other benchmarks mentioned above (see e.g., Steward, 2004; Lovrich et al., 2005). However, the census data may not accurately reflect each racial group's share of motorists on roadways where commuters travel, especially on highways.

---

[25] Prior to the stop, race might also be used as a factor in determining whether to check the records of a motorist (e.g., determine whether a vehicle owner, who is presumably the driver, has a suspended driver's license), which would lead to additional stops and arrests for the targeted racial group.

[26] The seriousness of the violation is an important predictor for whether a motorist is stopped. Not all violations observed by the police result in a stop.

Even with an accurate estimate of each racial group's share of motorists on the roadways, if racial groups' propensities to commit violations that trigger a stop or their relative exposure to police differ, then these characteristics need to be accounted for. Although many studies have shown that nearly all motorists routinely violate minor traffic laws, there is less agreement about which motorists violate more serious traffic laws that trigger a stop. Lamberth (1994) found that 98 percent of motorists were speeding on the New Jersey Turnpike, and in their experiment, Farrell et al. (2005) found that 94 percent of motorists were violating a traffic law. MacDonald (2001) notes that the National Highway Traffic Safety Administration (NHTSA) reported that blacks represent 10 percent of drivers nationally, but represent 13 percent of drivers involved in fatal accidents and 16 percent of drivers involved in injury accidents.[27] However, using observers in Miami, Alpert (2004) found that white males had a higher number of violations per driver as compared blacks and females. Additionally, for particular violations, there may be different offending rates among racial groups. For example, Lovrich et al. (2003) cited a number of studies that found that blacks had higher seat belt violation rates than non-blacks. If the police are focusing on seat belt violations, then the black share of stopped motorists will increase.

Moreover, especially in a city setting, the police are not uniformly distributed throughout their jurisdiction, but instead are concentrated in high-crime areas. If those areas are associated with a particular race, then the expected share of stops for that racial group may increase.

To estimate the external benchmark and account for potential different violation types and police exposure among racial groups, Grogger and Ridgeway (in press) compare the racial-group shares of stops that occur when it is dark to the shares when it is light based on the premise that officers have a degraded ability to detect the race of the motorist when it is dark; hence, the racial-group shares of stops during darkness represent the external benchmark. Because the racial-group shares of motorists may change between daylight and darkness, they limit their evaluation to stops that occur near the same clock time during a several week period, which is feasible near the Daylight Saving

[27] However, she notes that seatbelt use differences among racial groups may explain some of the disparity.

Time change.[28] This method also accounts for a motorist's violation type and relative exposure to the police, which should not change between daylight and darkness, given a particular clock time.[29] One limitation of this approach is that it is limited to evaluating stops at clock times near sunrise or sunset; hence, it will not detect the use of race outside of these hours.

In general, the benchmarking estimates have been criticized because they may not accurately represent the racial-group shares violating traffic laws that trigger a stop and because racial groups may be differentially exposed to the police. But as an equal or even greater concern, the department-level studies only estimate the average use of race and do not capture potential variation in the use of race among officers. Therefore, even if a department-level study finds no evidence of the use of race, the actions of a few officers who are using race will likely be hidden. Moreover, if a department-level study finds evidence that race is being used, the police leadership is not able to determine whether its use varies among officers. The leadership's corrective action would differ if it knew almost all officers were using race versus a situation where only a few officers were identified as using race.

To address these concerns, Walker (2001, 2003a) proposes comparing traffic enforcement data among officers who are similarly situated. To be similarly situated, officers must patrol the same geographical area during the same time periods with the same assignment (e.g., traffic patrol). An officer's similarly situated peer officers serve as his benchmark, where the peers' stops represent the racial group shares that were at risk of being stopped by the officer. The benchmark is referred to as an internal benchmark since it is internal to the police department. However, the officer's estimated use of race is not an absolute estimate (i.e., as compared to a race-neutral officer who does not inappropriately use race as a factor in traffic enforcement decisions), but instead, is a relative estimate of his use of race as compared to his peers.

In order to better interpret the results of an officer-level analysis, it is important to know the similarly situated officers' average use of race as compared to the hypothetical

---

[28] In most contexts, the racial group shares of motorists should be similar during the same clocktimes. However, if the context involves a large number of motorists whose schedules depend on daylight (e.g., motorists in the construction or farming industry or elderly drivers), then this assumption may not hold.

race-neutral officer. Without that estimate, one is not even able to conclude whether an officer who most uses race relative to his peers to stop a higher share of minority motorists actually uses race to stop a higher share of minority motorists as compared to a race-neutral officer. This is because if peer officers use race to stop a higher share of non-minority motorists, then as compared to a race-neutral officer, this officer of interest may either use race as a factor (albeit the least intensive use among his peers) to stop a higher share of non-minority motorists, or may not use race as a factor in his stop decisions, or may use race as a factor to stop a higher share of minority motorists.[30] Hence, the results from an officer-level analysis are more informative when they are used in conjunction with results from a department-level analysis, which estimates the average use of race across the department (or similarly situated officers).

However, if an average use of race estimate is not available, the officer-level analysis could be useful by itself. For example, if the analysis identifies an officer that uses race significantly more than his peers, then that officer could be flagged by an EI system. This information could be used with other data being collected by the EI system to see if there is a pattern that this officer may potentially be using race as a factor in other decisions (e.g., for searches, citations, use of force, etc.). Walker (2001) emphasizes that the data used in an EI system are just a starting point to begin an inquiry, because alone, they are not conclusive because officers are often being compared to their peers, not an absolute standard.

Because the traffic enforcement data rarely include officer identifiers, there have been relatively few officer-level racial profiling studies (Engel and Calnon, 2004). The three studies below show the different empirical methods that have been used to identify problem officers. Fridell (2004) states that the Saint Louis Metropolitan Police Department (see Decker and Rojek, 2002) and Ohio State Patrol (see Friday, 2002) used internal benchmarking to estimate their officers use of race.[31] Decker and Rojek (2002) compute a z-score for each officer, which normalizes the distribution of the share of

---

[29] Because some violation types are more prevalent during daylight or darkness (e.g., headlight violations) and may be associated with a racial group, these stops were excluded from the analysis.

[30] If the peer officers are race neutral or use race to stop a higher share of minority motorists, then one can conclude that this officer uses race to stop a higher share of minority motorists as compared to a race-neutral officer.

minority stops, resulting in a mean of zero and a standard deviation of one. This calculation is shown in Eq. (2.1) where M_j is a binary variable that indicates whether a motorist stopped by Officer *j* is minority:

$$z_j = \frac{\overline{M}_j - \overline{M}}{s.d.(\overline{M}_j)} \tag{2.1}$$

where $\overline{M}_j$ and $\overline{M}$ are the minority shares of Officer *j*'s and all stops, respectively, and $s.d.(\overline{M}_j)$ is the standard deviation of the $\overline{M}_j$ observations.

They do not include officers who had less than a certain number of stops. Friday (2002) uses a similar method to compare similarly situated officers' arrest rates of minorities.

As shown in Eq. (2.2), Smith (2005) calculates a t-statistic for each officer that is based on comparing the subject officer's proportion of minority stops with his peer's proportion. As compared to the z-score, the t-statistic is more informative since it is sensitive to the number of observations.[32]

$$t_j = \frac{\overline{M}_j - \overline{M}_{\sim j}}{\sqrt{\frac{Var(M_j)}{N_j} + \frac{Var(M_{\sim j})}{N_{\sim j}}}} \tag{2.2}$$

where $\sim j$ indicates a motorist stopped by an officer other than the Officer *j*.

Officers are then ranked by their t-statistics.[33]

However, the above approaches are somewhat limited if an officer's work schedule or patrol area (e.g., neighborhood or roadway) changes over time. When these changes occur, his peer officers will change; hence, his peer group will not always be the

---

[31] Decker and Rojek (2002) is an internal police report not available to the public, but it was discussed in Fridell (2004).

[32] The effective number of observations within the z-score is the number of officers that are evaluated, while the number of observations within the t-statistic is the number of stops.

[33] Additionally, Alpert et al. (2005) developed an internal benchmarking methodology for the Los Angeles Police Department to employ in the future. The methodology is similar to Smith (2005) and also included a discussion on how to analyze a particular officer's stops when the officer is similarly situated to more than one peer group of officers.

same officers. Moreover, when an officer's peers change their work schedules or patrol areas, these changes will also cause the officer's peer group to change. If an analysis is limited to when a particular group of officers patrolled together in a particular geographical area, the sample may be restricted to a small subset of their stops, which reduces the ability to detect officer differences. If the schedule and patrol area changes are ignored, then the results could identify the wrong officers as problem officers due to officers working somewhat different schedules or patrolling somewhat different geographical areas than their peers.

To address these issues, Riley et al. (2005) developed a method that compares a particular officer's stops to other officers' stops that occurred at the same time and location; hence, the method is not restricted to when a particular group of officers patrolled together. An officer's performance is based on his average performance as compared to his peers' stops that occurred during the times he patrolled. During 2004, 91 officers from the Cincinnati Police Department made over 100 stops. For each officer, the black share of his stops was compared to the weighted black share of his peers' stops. The weights were generated from propensity scores so that the peers' weighted distribution of stop characteristics closely matched the subject officer's distribution of stop characteristics based on the neighborhood, the month, the day of the week, and the time the stops occurred. Hence, the subject officer's stops and his weighted peers' stops are considered to be similarly situated. To test whether black share of a subject officer's stops statistically differed from the weighted black share of his peers' stops, a beta-binomial distribution was used to generate a confidence interval of the number of black stops made by the subject officer, assuming the black share of his stops was the same as the weighed black share of his peers' stops. If the actual number of black stops exceeded the upper end of the confidence interval, then this provided evidence that the subject officer used race as a factor to stop a higher share of black motorists as compared to his peers. Of the 91 officers, four had minority stop counts that exceeded the upper end of the confidence interval. Based on this analysis, these officers may have used race as a factor to stop a higher share of minority motorists relative to their peers. Moreover, these officers may have also used race as a factor to stop a higher share of minority motorists on an absolute basis (i.e., as compared to a race-neutral officer) since the department-level study showed no evidence that race was used on average to stop a higher share of non-minority motorists.

2.3.1.2 Post-Stop Decisions: Citation, Search, and Arrest Rates

Once a motorist has been stopped for a traffic violation, an officer might use race as a factor in deciding which stopped motorists are issued a citation (instead of a warning) and which motorists and vehicles are searched during the stop.[34] Additionally, he might use race as a factor in deciding which stopped motorists are arrested (e.g., for driving under the influence of alcohol or drugs (DUI)). Unlike stops, the racial-group shares of the at-risk population that may be cited, searched, or arrested are known because the population consists of motorists who are already stopped for traffic violations. However, like stops, there are many contextual and motorist characteristics that influence the decision to cite, search, or arrest a motorist, some which are not included in the data. If these characteristics are associated with a racial group, then the use-of-race estimate for citing, searching, and arresting motorists will be biased if these characteristics are not accounted for. Additionally, these confounding characteristics may also influence whether a motorist is stopped in the first place; hence, stopped motorists may not have the same risk of being cited, searched, or arrested.

During a traffic stop, an officer has discretion over the number and type of violations that he records for a motorist; hence, the number of violations and the number of serious violations are not necessarily only based on the motorist's actions. An officer who uses race as a factor in the citation decision might also use race as a factor in deciding how thoroughly he scrutinizes a motorist for violations. Moreover, he might use race as a factor in deciding whether to record each violation, whether to record the relatively more serious versus less serious violations, and whether to reduce a serious violation to a non-serious violation (e.g., speeding from 25 miles per hour over the speed limit to 10 miles per hour over the speed limit).

Once a violation is recorded, an officer also has a great deal of discretion whether to issue a citation or give a warning (Farrell et al., 2005). Citations typically include a

[34] Additionally there is a concern that the duration of stops involving minority motorists lasts longer than those involving non-minority motorists either because officers more thoroughly scrutinize minority motorists or because officers search a higher share of minority motorists.

fine; hence, if an officer has a racial animosity toward a particular racial group, he may cite that group at a higher rate than other groups. The factors that influence his decision to cite include the type and number of violations, contextual characteristics (e.g., time of day, traffic conditions, weather), police department priorities (e.g., emphasizing seatbelt violations), motorist's demographic characteristics (e.g., gender, age, residence, social status), and other motorist's characteristics (e.g., behavior, actions, demeanor). Lovrich et al. (2003, 2005) found that as the number and seriousness of the violations increase, the likelihood of a motorist being cited significantly increases. For example, if an officer stops a motorist for multiple serious violations (e.g., negligent driving with a headlight out), he is more likely to cite him as compared to a motorist who is stopped for one minor violation (e.g., failure to signal). As data collection efforts have evolved, more of the above factors are being captured in the data; however, some factors such as a motorist's behavior are difficult to capture, making it more difficult to estimate an officer's use of race in a citation decision. Moreover, controlling for the number and seriousness of the violations in a model may mask the use of race if race was used as a factor in deciding the number and seriousness of the violations.[35]

Although an officer may issue a citation to any motorist who has violated a traffic law, other conditions must be met before he can search a motorist and his vehicle or arrest him (e.g., for a DUI). In general, an officer may search a motorist and his vehicle if he possesses a search warrant, can establish probable cause or reasonable suspicion, or is given consent to search by the motorist. Moreover, if an officer establishes probable cause to arrest a motorist, then he may search the motorist for weapons and evidence as part of that arrest. These are known as incident-to-arrest searches.

For some situations, the officer has little discretion as to whether to conduct a search, and in other situations, he has a greater degree of discretion. For example, if an officer stops a motorist and discovers that there is an outstanding warrant for his arrest involving a serious felony, the officer has little discretion but to arrest and search the motorist for weapons and evidence. Similarly, if an officer spots contraband in plain site

[35] Additionally, it is not clear whether a higher citation rate for one racial group indicates that racial group was adversely impacted. For example, if minority motorists are stopped on relatively dubious grounds because the officer wants to perform a quick walk-around search, then if nothing illegal is discovered, the

(e.g., drugs), then he has probable cause to arrest the motorist and search him and his vehicle for additional evidence.

On the other hand, some situations are more ambiguous whether probable cause or reasonable suspicion has been established. Because the precise degree of evidence needed to establish probable cause or reasonable suspicion is ambiguous, the officer's judgment and discretion can be influential as to whether he conducts a search in many situations. For example, if an officer suspects a motorist of carrying drugs (e.g., due to the motorist's furtive movements as he was approaching the vehicle or contradictory statements during questioning), it may still be ambiguous whether probable cause or reasonable suspicion is established. In another example, an officer may stop a motorist he suspects of driving under the influence of alcohol or drugs (e.g., due to the motorist weaving back and forth). Once the motorist is stopped, if the officer still has a reasonable basis that the motorist is impaired, he may ask the motorist to submit to field sobriety tests. If the motorist refuses to submit to the tests, then there may be cases where the officer is unsure whether the motorist is sufficiently impaired in order to establish probable cause that would justify an arrest. Terry searches also involve officer judgment and discretion. If an officer has reasonable suspicion that a motorist is carrying a weapon that poses a danger to the officer, the officer may pat search or frisk the motorist. These searches are considered to involve more discretion than a probable cause search since the evidentiary standard needed to search is lower. Additionally, even if the officer does not have reasonable suspicion or probable cause, he is legally permitted to ask the motorist for consent to search. Some police departments have restricted consent searches by requiring the officer to have an articulatable suspicion before asking for consent, and due to a racial profiling lawsuit, the California Highway Patrol instituted a moratorium on consent searches (Richardson, 2003).

In order to estimate a department's use of race as a factor in search decisions, it is important to distinguish between low- and high-discretion searches if they vary among racial groups (e.g., see Ridgeway (in press); Riley et al., (2005); Lovrich et al. (2003, 2005)). Otherwise these confounding characteristics will bias the use-of-race estimate.

---

motorist may be released without a citation. This would lead to a low citation rate for stopped minority motorists, even though they were adversely impacted.

However, sometimes the data do not identify the legal basis for the search, making the results more difficult to interpret. For example, Steward (2004) found that blacks and Latinos were more likely than white motorists to be searched during a traffic stop in 86 percent of 413 Texas law enforcement jurisdictions. However, the data from the jurisdictions often did not include whether the search was low discretion versus high discretion; hence, whether a disparity exists for high-discretion searches is unknown.

But even if the discretion level of the search is known, because an officer has discretion over the number and types of violations that he records (as was discussed above with citations), then, for example, if he uses race as a factor against minorities, he may record a higher number and more serious violations for minority motorists than non-minority motorists who have the same violation(s). If this occurs when the search discretion level (i.e., these variables) are included in the model, then the use-of-race estimate will be understated. However, if minority motorists have more violations and more serious violations than non-minorities, then since these variables are positively associated with searches, then the use-of-race estimate will be overstated if they are not included in the model. Hence, the results of either model should be interpreted knowing that the estimates may be biased in a particular direction.

In high-discretion search situations, the factors that influence an officer's decision to search include the seriousness of the violation(s), contextual characteristics (e.g., location of stop, time of day, day of week), motorist's demographic characteristics (e.g., gender, age, residence), and other motorist's characteristics (e.g., behavior, actions, demeanor, appearance). For example, the neighborhood where the stop occurred can be very influential. Ridgeway (2006) found that the search rate in high-crime Oakland neighborhoods was higher across all races as compared to the search rate in low-crime neighborhoods. In New York City, stop-and-frisk rates significantly varied among police precincts (Office of the Attorney General, New York, 1999). In studies where data distinguishes between low- and high-discretion searches and also includes the factors above, then multinomial logistic regression (Lovrich et al., 2003, 2005) or propensity score weighting (Ridgeway, 2006) models have been used to estimate the use of race in search decisions.

As data collection efforts have evolved, the information in the data has improved; however, characteristics such as a motorist's behavior, actions, demeanor, and appearance are difficult to measure and are typically not included in the data. An officer's suspicion is heightened if the driver appears nervous or if the driver and passengers give different answers when asked about their driving origin or destination. Similarly, when an officer is tailing a vehicle or is walking up to a stopped vehicle, his suspicion is heightened if the driver or passengers make furtive movements as though they are hiding contraband or reaching for a weapon. An officer's suspicion is also heightened when he smells drugs or alcohol, or when the driver and passengers attempt to mask the smell by driving with their windows down when it is uncomfortably cold outside and/or lighting a cigarette when the officer approaches the vehicle.

Due to many contextual and motorist characteristics that affect citation and search decisions that are not observed by the analyst, an internal benchmarking approach can mitigate some concerns. If officers are similarly situated, then they should be exposed to similar stopped motorists within each racial group. The characteristics used to make citation and search decisions that are not included within the data should be similar across the motorists (including motorists within a racial group) that each officer stops since similarly situated officers should be stopping similar motorists. Therefore, an internal benchmarking approach that compares each officer's citation and search rates of different racial groups to his peer officers' respective rates (while controlling for observed characteristics that are associated with racial groups and citation and search decisions) provides a method to estimate each officer's relative use of race as a factor in these decisions. Because the unobserved confounding characteristics should be similar among similarly situated officers, the estimated use of race will be relatively more accurate than an absolute estimate of the use of race. This is similar to the improved accuracy in the stop analysis where the internal benchmark does not require the elusive external benchmark. However, similarly situated officers may have different production levels (e.g., number of stops, citations, and searches per shift), which may result in them being exposed to different types of motorists than their similarly situated peers. In addition, the internal benchmarking estimate of the use of race is only a relative measure that should be complemented with an estimate of the average use of race. However, even if an average use of race estimate is not available, the officer-level analysis could be useful by itself. For example, if the analysis identifies an officer that uses race significantly more than his peers, then that officer could be flagged by an EI system to see if a pattern exists across other traffic enforcement decisions. To my knowledge, there have been no empirical studies published using an internal benchmarking approach to evaluate post-stop decisions.[36]

2.3.1.3 Post-Stop Decisions: Search and DUI Arrest Hit Rates

In order to mitigate the effect of unobserved confounding characteristics as well as to estimate the use of race that is solely due to racial animosity in search decisions, Knowles, Persico, and Todd (2001) use an outcomes-based test that compares the share of each racial group's searches where contraband is found. When contraband is found, these searches are considered successful searches and the percentage of searches that result in contraband being found is known as the hit rate (i.e., the number of searches where contraband is found divided by the total number of searches). Their model assumes that an officer maximizes the number of successful searches minus the cost to search.[37] Hence, an officer will search motorists based on his estimated probability that the motorist is carrying contraband, which can be referred to as the officer's suspicion level. The officer's suspicion level is based on multiple contextual and motorist characteristics $\boldsymbol{x}$ (e.g., location, time of day, motorist's non-racial characteristics) and one of the characteristics may be the motorist's race, R. If the officer's suspicion level for a motorist is above the officer's threshold suspicion level for searching a motorist, then the officer will search the motorist. If the officer has no racial animosity, then the least suspicious motorist that he searches within each racial group should yield the same expected hit rate. The expected hit rate of the least suspicious motorist searched is called the expected

[36] Although the following studies have not been published, Avery Guest's 2005 declarations in *Lacy v. Villeneuve* deserve mention. The case involved a federal civil rights lawsuit against the Washington State Patrol for alleged police misconduct and racial discrimination during a traffic stop in 2002 in the Seattle area. As an expert witness, Guest used internal benchmarking methods in his court declarations to provide evidence that the trooper racially discriminated against black motorists in stop and post-stop decisions. One of the post-stop methods he used is similar to a post-stop method that I use in Chapter 6. In early 2006, the case was settled out of court, but Guest's declarations are publicly available: United States District Court, Western District of Washington at Seattle, Case No.: CV03-2442JLR, see Guest (2005).

[37] Assume the expected benefit of finding contraband does not vary among racial groups, that is, the type and quantity of contraband found is similar among racial groups.

marginal hit rate. If the marginal hit rate is lower for one racial group (e.g., $R = A$) as compared to another racial group (e.g., $R = B$), then the officer is willing to search relatively less suspicious motorists in group A as compared to group B.[38] This reduces the value of the objective function above, which is to maximize the number of successful searches (minus the cost to search). The officer could increase the value of the objective function by not searching the relatively unsuspicious motorists in group A, and instead, search the relatively more suspicious motorists in group B. The officer bears the cost of the decrease in the value of the objective, which Becker (1957) defines as discrimination or racial animosity. Hence, this approach estimates the use of race solely due to racial animosity and does not attempt to determine whether race is being used as a statistical discriminator.

However, since the search data do not identify the marginal motorists, the data only allow an analyst to calculate the average hit rate for each racial group, not the marginal hit rate. And comparing the average hit rate of each racial group may not estimate the use of race as a factor in search decisions due to racial animosity. For example, assume that an officer does not possess racial animosity. If the searched motorists from group A were on average more suspicious than the searched motorists from group B, then group A's expected average hit rate is higher than group B's expected average hit rate. Hence, in this situation, comparing each group's average hit rate is an inaccurate estimate of racial animosity.

However, the Knowles et al. (2001) model addresses this data limitation. Their model shows that in equilibrium, each motorist type ($\boldsymbol{x}$, R) has the same probability of carrying contraband due to the different search rates that each group is exposed to.[39] Their model allows the benefits and costs of carrying contraband to differ among racial groups, which may be due to different employment and education opportunities among racial groups. Because the benefits and costs of carrying contraband are costly for the police to directly measure, the police use race as proxy for these values and adjust their search rates accordingly, which Arrow (1973) calls statistical discrimination.

[38] This assumes the thoroughness of the officer's search is the same among racial groups.
[39] This applies to motorist types ($\boldsymbol{x}$, R) that have a positive probability of being searched. There are motorist types who will not carry contraband even if the probability of being searched is zero.

If the probability of carrying contraband differed among motorist types, then police would adjust their search rates to maximize the objective function. As police adjust their search rate for motorist type ($\boldsymbol{x}$, R), then that motorist type ($\boldsymbol{x}$, R) will adjust their carry rate. This is because motorist type's ($\boldsymbol{x}$, R) carry rate depends on the probability of being searched. As the probability of being searched increases (decreases), the motorist's optimal carry rate will decrease (increase). When both the police and motorists are maximizing their respective objective functions, the equilibrium search and carry rates for each motorist type ($\boldsymbol{x}$, R) is defined by a Nash equilibrium. For example, if one motorist type ($\boldsymbol{x}$, R) had a higher probability of carrying contraband as compared another motorist type ($\boldsymbol{x}$, R) then police would search the first motorist type at a higher rate and search the second motorist type at a lower rate, until the two motorist types' probability of carrying became equal.[40] In equilibrium, the probability of carrying is the same among all motorist types ($\boldsymbol{x}$, R) due to the differential search rates they are exposed to. Hence, the average and marginal hit rates are the same; therefore, if the average hit rate is lower for one racial group as compared to another, then this is evidence of racial animosity toward the former group.

However, several papers have criticized the Knowles model (e.g., see Anwar and Fang, 2006; Bjerk, 2006). They argue that $\boldsymbol{x}$ depends on whether a motorist is carrying contraband, that is, $\boldsymbol{x}$ is endogenous to the probability that the motorist is carrying contraband. They state that a motorist type ($\boldsymbol{x}$, R) who is carrying contraband will emit a higher level of suspicion to the police as compared to motorist type ($\boldsymbol{x}$, R) who is not carrying contraband; hence, the motorist who is carrying contraband will be become motorist type ($\boldsymbol{x}_c$, R) where the subscript c indicates a motorist is carrying contraband. The difference in the level of suspicion could be due to several factors. First, if a motorist is carrying contraband, there is a positive probability that the police will directly detect the contraband either due to seeing or smelling it (in the case of drugs). Second, the motorist carrying contraband may exhibit more signs of nervousness as compared to the non-carrying motorist. Therefore, the probability of finding contraband within the searched members of motorist type ($\boldsymbol{x}$, R) will be higher than the probability of finding

[40] The model assumes that the police's benefit of finding contraband does not differ by racial group; hence, the average nature and quantity of contraband found must be similar among racial groups.

contraband within all members of motorist type ($\boldsymbol{x}$, R). If the change from $\boldsymbol{x}$ to $\boldsymbol{x}_c$ differs among racial groups, then the average hit rate of the searched motorists may not be equal when the officers have no racial animosity.[41]

2.3.2 Ability to Act on the Use-Of-Race Estimate to Change Officer Behavior

Before discussing corrective action alternatives for a police department that inappropriately uses race as a factor in their traffic enforcement decisions, it is important to understand the organizational characteristics of a police department. Klitgaard (2000) states that governance involves establishing processes and systems to reduce monopoly power, manage discretion, and increase transparency (or accountability). If an organization establishes sound governance, it will better meet its objectives and reduce corruption. However, the state gives the police a monopoly on the legitimate use of force; hence, due to this inherent power, establishing sound governance within a police organization is challenging. Therefore, sound governance critically depends on managing discretion and increasing transparency in order to reduce the potential for the police to abuse their power.

Even in a traffic enforcement setting, a traffic enforcement officer has a great deal of discretion in deciding whom he stops for a traffic violation as well as a great deal of power over the stopped individual. And for any given stop or post-stop decision, there is little transparency with respect to the officer's decision-making process. Although legislation and department policies limit an officer's discretion in using race as a factor in traffic enforcement decisions, data collection and analysis at the department level provide some transparency as to whether the laws and policies are being followed. If a racial profiling analysis is also done at the officer level, there is a greater degree of transparency on the use of race and individual officers can be held accountable. Additionally, incentives and corrective action can be tailored to increase the likelihood that individual officers will comply with use-of-race laws and policies.

[41] Anwar and Fang (2006) develop a less powerful test that involves comparing officers from different racial groups, whereby the hit rate of officers and motorists who share the same race are compared to hit rate of officers and motorists who do share the same race. As compared to the Knowles model (assuming the model's assumptions are correct), this test is less able to detect the use of race when it is being used as a factor in search decisions.

In general, incentives are designed to align a worker's (e.g., a police officer's) objectives with his employer's (e.g., a city's) objectives. While incentives are not always monetary in nature, each incentive could theoretically be denominated as a monetary value. An incentive with a high monetary value is considered high powered, and an incentive with a low monetary value is considered low powered. High- and low-powered incentives are not binary, but instead represent segments on a continuum, where an extremely high-powered incentive is where a worker's wage is solely tied to performance, and where an extremely low-powered incentive is where a worker's wage is mostly guaranteed with a small portion based on performance.

The organization's characteristics dictate whether high- or low-powered incentives are optimal (see e.g., Holmstrom and Milgrom, 1991; Klitgaard and Light, 2004). An employer prefers organizational characteristics that result in high-powered incentives being optimal because if low-powered incentives are optimal, the employer is less able to influence workers to align their objectives with his. High-powered incentives are less optimal than low-powered incentives when the organization's objectives are multi-dimensional and some of the objectives are unobserved or difficult to measure; when monitoring and measuring costs are high; and when performance is partly based on random factors outside a worker's control. In a police organization, many of these characteristics are acutely present; hence, high-powered incentives are used less often. This does not mean that high-powered incentives such as reassignment, suspension, and termination do not exist, but their use is limited due to the above organizational characteristics.[42] I will discuss each of the organizational characteristics as they relate to a police organization, and then discuss how the effectiveness of low-powered incentives vary when a racial profiling study is done at the department level versus the officer level.

In a police organization, eliminating the inappropriate use of race as a factor in law enforcement decisions is one objective among many objectives, namely deterring and preventing crime, and in a traffic law enforcement setting, deterring and preventing

[42] When organizations are mostly limited to low-powered incentives, then it becomes optimal to expend more resources screening and indoctrinating applicants to ensure their objectives mostly align with the organization's objectives, which is part of the reason why police departments have a very thorough applicant screening and indoctrination program. Part the indoctrination includes inculcating organizational values and norms. Perez (1994) states that socialization, which includes social values, individual desires,

traffic accidents and fatalities by enforcing traffic laws. In a traffic law enforcement setting, a police organization will attempt to create incentives so the traffic officer's objectives align with the organization's objectives. Incentives may be linked to an officer's performance across various dimensions such as the number of stops, citations, searches, and DUI arrests that an officer makes per shift as well as dimensions that may indicate problems such as uses of force, officer involvement in lawsuits, and citizen complaints. Other dimensions that are more difficult to measure include the degree of professionalism and respect that an officer uses when stopping motorists (as well as the degree of respect that the stopped motorist shows toward the officer). If an incentive is tied to measurable outcomes, then the incentive's effect on other outcomes (especially unobserved outcomes) must be considered. For example, a policy that specifies racial-group share quotas for stopped and searched motorists could hinder the police's ability to stop or search persons when they are driving dangerously or potentially committing another serious crime.[43] Moreover, the incentives must be closely managed, otherwise a perverse incentive may be created for an officer to, for example, falsely arrest a motorist for aggressive driving.

Within a traffic enforcement setting, some performance measures can be measured at a relatively low cost, while others are more difficult to measure. The common measures include the demographic characteristics of stopped motorists, the date and time of the stop, the reason for the stop, the number and types of violations, whether a citation was issued, and whether a searched occurred, and if so, its legal basis and whether contraband was found. However, due to the large number of motorists an officer observes violating a traffic law but does not stop, it would be very costly and practically impossible for him to record their characteristics as well as the reason for not stopping them. And in a post-stop situation, there are many qualitative factors that affect a decision to cite or search that are difficult to measure, especially in a quantitative fashion. Some police departments are beginning to mount cameras on patrol cars so they can better

---

professional standards, and cultural expectations, can be very effective at influencing behavior, but these norms can be difficult and costly to precisely control.

[43] In these instances, the policy may waive the quota requirement; however, if enough waivers are permitted, the racial-group shares of stopped and searched motorists may return to their pre-quota levels, making the policy irrelevant.

monitor an officer’s performance as well as evaluate the legitimacy a of motorist’s claim of police misconduct.

Measuring an officer’s performance is also difficult because many situations are partially outside his control. For example, the share of stopped motorists from a particular racial group that an officer searches is partially based that racial group’s share of stopped motorists whose actions warrant a search and partially based on his inappropriate use of race (if any) in deciding whom to search. The first component is random and outside of the officer’s control. Over a short evaluation period, this random component can vary widely, and thus, may dominate the effect of an officer’s use of race. When a racial profiling study is done at the department level with a large number of observations, the influence of the random effects is somewhat negligible because officers who are exposed to an above average share of one racial group that warrants a search will likely be counterbalanced by other officers who are exposed to a below average share that warrants a search. When a study is done at the officer level, the influence of these random effects is magnified because there are fewer observations. This is especially a concern for rare events such as the use of force. For example, the Pittsburgh Police Bureau is analyzing whether officers are not using force in situations when their safety is in danger.

Due to the above three characteristics being present in a police organization: multi-dimensional objectives (where some objectives are unobserved or difficult to measure), high monitoring and measuring costs, and random factors outside an officer’s control, then low-powered incentives will generally be more optimal than high-powered incentives. Because low-powered incentives are relatively ineffective at influencing behavior as compared to high-powered incentives, it becomes more important that the low-powered incentives be closely managed. To improve the management of these incentives, police departments have been developing and using EI systems (which were discussed above).

In order for a racial profiling study to be used by an EI system at the officer level, an officer-level analysis is needed. Although the incentives within an EI system are mostly low powered, if a study identifies an officer who inappropriately uses race as a factor in traffic enforcement decisions, the EI system could be effective at changing the officer’s behavior because he could be subjected to the interventions that have a somewhat higher degree of power. When racial profiling may be occurring at the department level and the potential varying use of race is unknown at the officer level, then police leadership is limited to corrective actions that affect all officers somewhat the same, including officers who were not using race. For example, in *Curtis V. Rodriguez, et al. v. California Highway Patrol, et al.*, the California Highway Patrol (CHP) was sued for racial profiling. As part of the settlement, the CHP extended its moratorium on consent searches for three years (Richardson, 2003). This decision was partially based on internal and external analyses (e.g., see McCormick and Zamora (2001)) that showed that stopped black and Hispanic motorists were searched at a higher rate than stopped white motorists. If these analyses had been done at the officer level, then particular officers may have been identified with significantly higher than average search rate disparities among racial groups. Although the search rate disparities were widespread across the CHP's eight divisions, if an officer-level analysis showed that a few officers were most responsible for the disparities, then corrective action might have taken place at the individual officer level. Even if the department-wide corrective action of instituting a consent search moratorium was still deemed appropriate, then CHP leadership could scrutinize the officers with the largest disparities to see if their other activities warranted corrective action.

2.3.3 Officer Support for EI Systems and Traffic Enforcement Data Collection

If a racial profiling analysis is going to identify problem officers, it is critical that the empirical methods identify the appropriate officers. Otherwise, officers will not support the data collection and analysis efforts, making it difficult to do the analysis. In general, officer-level analyses will be favored by some officers and resisted by others (Fridell et al., 2001). If a department has a few officers that bring disrepute to the entire organization due to their inappropriate uses of race, then the majority of officers may be inclined to favor officer-level transparency. However, that transparency will be resisted by officers who fear that their legitimate actions will be misjudged (as well as those officers who want to hide their illegitimate actions). Moreover, this transparency could lead to lawsuits. Hence, most police departments are not collecting officer identifiers within their traffic enforcement data (Engel and Calnon, 2004). In order to better ensure officer-level transparency is not abused, then an officer-level analysis that is incorporated into an EI system should trigger an inquiry, not a conclusion (Walker, 2001).

The use of EI systems is increasing across law enforcement agencies (Walker et al., 2005). In its survey, Walker (2003) found that approximately 40 percent of law enforcement agencies have an EI system.[44] In the survey, 64 percent of managers reported a positive general assessment of the EI system and only 4 percent reported negative general assessment. Sixty-five percent reported that the system had a positive impact on management and supervision, and 2 percent reported a negative impact. Walker states the most surprising finding was that 84 percent of the agencies reported that the police union did not oppose the EI system. However, the effect on morale was more uneven. Twelve percent reported that the EI system had a positive effect, 52 percent reported a mixed impact, 30 percent reported no impact, and 6 percent reported a negative impact. However, rank-and-file officers' initial skepticism and distrust tended to change and become more supportive over time.

Based on the survey, the overall assessment of EI systems is quite positive. However, whether these results can be generalized to other agencies depends on whether those agencies that responded to the survey are representative of other agencies, and more importantly, whether those that responded and had an EI system are representative of other agencies. The agencies that had implemented an EI system likely include agencies where police union resistance was expected to be low, otherwise the implementation would have been more difficult or not possible. Hence, agencies that have not implemented an EI system may face more resistance from their police unions. However, because the overall assessment of the EI systems has been quite positive, police unions that were initially skeptical may be more willing to support an EI system, and moreover, they may be more willing to support incorporating racial profiling data into them.

2.3.4 Cost of Estimating the Use of Race

Although beneficial in many respects, an officer-level analysis will likely cost more than a department-level analysis due to increased data collection and analysis costs. The analysis requires additional steps in order to ensure similarly situated officers are

[44] The survey was emailed to 521 Police Executive Research Forum (PERF) members and 135 responded.

being compared. For a large police organization, this could require doing several internal benchmarking analyses with different groups of similarly situated officers.[45] Additional data needs to be collected in order to identify officers who are similarly situated. However, if this data collection and analysis becomes part of an already existing EI system, then the additional costs may be minimal. On the other hand, from the police department's perspective, an officer-level analysis could save money by identifying problem officers before their actions result in costly lawsuits.

As discussed above, maintaining an EI system requires line officers to do more paperwork and requires supervisors to analyze the data and conduct interventions. Additionally, there are computer hardware and software costs, especially the initial costs to set up the system. Although there have been some studies that evaluate EI systems, future research needs to be done to compare the costs of an EI system with its benefits.

2.3.5 Summary of Department- Versus Officer-Level Racial Profiling Analyses

Table 2.1 summarizes the key differences of estimating the use of race at department versus officer level, and summarizes the implications of incorporating traffic enforcement racial profiling analyses into an EI system. The table includes the five factors discussed above: nature and accuracy of the use-of-race estimate, ability to act on the estimate, methods to identify potential problem officers, officer support, and analysis costs.

[45] For example if there were two groups of officers whose patrol areas did not overlap, a separate analysis would be required for each group.

Table 2.1: Summary of the Issues of Conducting a Racial Profiling Analysis at the Department Versus Officer Level and Incorporating the Results into an EI System

Issues	Department-Level	Officer-Level
Nature and accuracy of use-of-race estimate	• Estimates the average • Large sample size	• Estimates the variance among officers • Not reliant on difficult-to-obtain external benchmarks (for stop analysis), but estimated use of race is relative to peers • Omitted variables may not bias results if troopers are similarly exposed to them
Ability to act on the use-of-race estimate	• Able to justify a department-wide corrective action	• Able to individualize the corrective action
Methods to identify potential problem officers	• Not applicable	• If peer groups are stable, descriptive statistics may be sufficient, but if not, multivariate analysis is needed
Officer support	• Not concerned that findings will falsely implicate particular officers	• Will identify peers who inappropriately use race
Analysis costs	• Less analysis costs, especially for a large police department that would require many separate similarly situated officer-level analyses	• Additional costs may be reduced if can incorporate racial profiling data and analysis into an existing EI system • May reduce the likelihood of costly lawsuits

2.4 Summary

This chapter discussed how racial disparities found in traffic enforcement studies may be due to racial animosity, statistical discrimination, or confounding characteristics. The chapter also discussed that the legality of the use of race as a factor to establish probable cause or reasonable suspicion depends on the influence of race, the suspect's specificity, and the state's interest (which includes the legal standard that is applied). Most studies have estimated the use of race as a factor in traffic enforcement decisions at the department level, but there is a growing interest to estimate each officer's use of race and incorporate the results into an EI system designed to identify problem officers and improve officer performance (Walker, 2001; 2003a). This chapter concludes that department-level and officer-level studies are different in many respects, including the following: nature and accuracy of the use-of-race estimate, ability to act on the estimate, methods to identify potential problem officers, officer support, and analysis costs. A police department's key policy decision is whether to incorporate traffic enforcement data into an EI system in order to estimate each officer's use of race, and ultimately, to identify potential problem officers. That decision should be based on considering the issues summarized in Table 2.1. If an officer-level analysis is going to be conducted and ultimately used within an EI system, then the empirical methods used to identify problem officers must accurately identify the appropriate officers. This will help ensure the EI system has integrity, legitimacy, and a learning component.

Chapter 3: Detachment-Level Stop Analysis

3.1 Introduction

An officer-level analysis within an EI system often involves comparing officers who are similarly situated, which includes officers who patrol the same geographical area during the same time periods with the same assignment (e.g., traffic patrol). The primary purpose of this chapter is to estimate 16 similarly situated WSP troopers' average use of race as a factor in deciding which motorists to stop. If the 16 troopers' average use of race for stop decisions is estimated, then each trooper's relative use of race as compared to his peers can be better interpreted (and each trooper's relative use of race for stop decisions will be estimated in Chapter 4). In Chapter 2, the primary methodologies used to estimate the use of race in stop decisions were discussed and critiqued, so they are not repeated here.

3.2 Washington State Patrol Overview and Data

3.2.1 Washington State Patrol Overview

Within the Washington State Patrol (WSP), the Field Operations Bureau's (FOB) is responsible for traffic law enforcement on Washington's highways, state routes, and certain county roads, totaling 17,500 miles. Its mission also includes criminal interdiction and terrorism prevention; collision investigation; and roadside assistance (WSP, 2005a). The priorities within traffic enforcement are known as the "core four" and include the following types of violations: driving under the influence (DUI), dangerous speeding, seat belt, and aggressive driving (WSP 2005b). During this study's analysis period, which runs from January 1, 2003 through March 31, 2005, there were between 700 and 800 traffic officers assigned to the FOB.

The FOB is divided into eight geographical districts that span the state (see map in Figure 3.1, which lists each district's headquarters). Each district is divided into four to six autonomous patrol areas (APA), for a total of 39 APAs.[46] Each APA includes

[46] The 39 APAs are numbered from 1 to 40 (excluding 17). For a short period, from 2002 to October 2003, a 41st APA existed near APA 33 in District 7.

between one and six detachments, which include approximately six to ten troopers; however, the detachments are not always geographically distinct. For example, APA 6 includes four detachments that patrol the same geographical area, but each detachment is responsible for a different time shift. From a command perspective, a captain oversees a district, a lieutenant oversees an APA, and a sergeant oversees a detachment.

Figure 3.1: Map of Washington State Patrol's Eight Districts

Source: WSP

The FOB collects trooper-level traffic enforcement data within its Time and Activity Reporting System (TARS).[47] However, TARS is not considered to be an EI system, but has some features that are found within an EI system. TARS has been in place for over 20 years and was originally designed to automate troopers' payroll. In the early 1990s, it was expanded to track each trooper's activity on each shift. This activity includes the number of stops, the type of violation (including DUI arrests), whether a citation was issued, and whether the motorist was searched. On a daily basis, a

[47] TARS is also referred to as the Time and Activity System (TAS).

detachment sergeant reviews the activity of each of his troopers, and on semimonthly basis, he reviews each trooper's aggregate activity for the period. As within an EI system, the sergeant will speak with a trooper if his activity deviates from what is considered normal. For example, if a trooper does not make any traffic stops during a shift, or works a Saturday night shift and does not make any DUI arrests, then the sergeant will ask the trooper for an explanation. This is the same as an informal intervention within an EI system. If a trooper's activity level is consistently unusual, then the sergeant may elevate the issue to his superiors. Each trooper is formally reviewed semiannually.[48] However, whether the review is based on a trooper's daily, semimonthly, or semiannual activity, the review does not include evaluating the racial-group shares of a particular activity (e.g., racial-group share of a trooper's stops or searches).

3.2.2 Washington Racial Profiling Legislation and WSP Racial Profiling Policy

On October 1, 1999, the WSP voluntarily began collecting traffic stop data in order to evaluate the racial-group shares of its activities, primarily at either the district or WSP level. The data collection began in response to the growing concern of racial profiling throughout the nation. Shortly thereafter in 2000, the state of Washington enacted a law mandating the WSP to collect and report on the traffic stop data.[49] In 2002, the state enacted a racial profiling law requiring all local law enforcement agencies to establish a written policy that condemned and prevented racial profiling. The law did not create a new legal definition for racial profiling, but instead used existing law by stating that racial profiling is "the illegal use of race or ethnicity as a factor in deciding to stop and question, take enforcement action, arrest, or search a person or vehicle with or without a legal basis under the United States Constitution or Washington state Constitution."[50] The WSP's policy covering racial profiling states that "Biased-based profiling is the selection of individuals based solely upon a common trait of a group, including, but not limited to, race, ethnic background, gender, sexual orientation, religion,

[48] On a monthly and annual basis, the troopers' activity levels are aggregated up to the detachment and district levels. If a particular detachment or district had unusual activity during a particular period, then the WSP leadership will seek to find an explanation, including reviewing an individual trooper's activity to see if corrective action is necessary.
[49] Revised Code of Washington (RCW) 43.43.480, enacted March 24, 2000.
[50] RCW 43.101.410, enacted March 12, 2002.

economic status, age, or cultural group" (WSP Regulation Manual, 2004). As stated in the legal analysis in Chapter 2, there are very few situations where federal law would permit race being used as the sole factor in a traffic enforcement decision. Therefore, although the WSP policy is more restrictive than federal law in these few situations, for most situations, federal and Washington state[51] laws are more restrictive than the WSP policy, and hence, those laws, not the WSP's policy, will restrict the WSP's ability to use race in its traffic enforcement decisions.

3.2.3 Washington State Patrol Data

The traffic stop dataset includes all WSP contacts with the public, including when a citation is not issued. Most of the contacts involve private motor vehicles, but some involve commercial vehicles, and on rare occasions, pedestrians. The WSP categorizes contacts between a WSP trooper and a motorist as either proactive or reactive. The majority of the contacts are proactive where the trooper initiates the contact for reasons such as stopping a motorist for routine traffic violations (e.g., speeding, lane change, seatbelt, or equipment violations) as well as for more serious and criminal violations such as aggressive driving or driving under the influence (DUI). Reactive contacts occur when a trooper is dispatched to respond to an incident such as an accident or a stranded motorist.[52] Troopers also make commercial vehicle contacts that involve inspections and weighing operations.

The traffic stop dataset includes the type of contact (e.g., routine stop, aggressive driving, commercial vehicle, collision investigation); the date, time, and location of the stop (identified by a highway and milepost number); the primary violation that initiated the stop[53]; the number and types of violations, including the enforcement action for each violation (i.e., warning or citation); whether a search occurred, and if so, the legal basis of

---

[51] The following Washington case restricts the use of race in certain situations. In *State v. Barber* (Washington Supreme Court, 1992), "racial incongruity" is defined as when a person of any race is considered "out of place" because the person's race is not the dominant race in the geographic area. *State v. Barber* prohibits using racial incongruity as a factor in establishing reasonable suspicion or probable cause (Loginsky, 2005).

[52] When a trooper initiates a contact with a stranded motorist (i.e., is not dispatched), the WSP considers that to be proactive contact; however, for the purposes of the analyses that follow, this type of contact is considered to be a reactive contact since the motorist is already stranded.

[53] The primary violation that initiated the stop is recorded as "violation one" in the data.

the search and whether contraband was found; the driver's[54] race, gender, and age; and the WSP trooper's name, gender, race, rank, and months of experience. There are eight racial-group categories for motorists: white, black (or African American), American Indian, Asian, Pacific Islander, East Indian, Hispanic, and other.[55]

Lovrich et al. (2003, 2005) and WSP officials identified South Seattle as an area of concern for the potential inappropriate use of race in traffic stop decisions involving minority motorists. Additionally, in early 2006, the WSP settled a federal civil rights lawsuit against a WSP trooper for alleged police misconduct and racial discrimination during a traffic stop just north of South Seattle in 2002.[56] The WSP designates South Seattle as APA 6, which is located within District 2. In order to analyze the traffic stops within APA 6, I selected 16 troopers who made over 500 stops between January 1, 2003 and March 31, 2005 on Interstate 5 between mileposts 140 and 162, inclusive.[57] The 16 troopers include 13 white males, one black male, and two white females. The median trooper has 4.5 years of experience, but the range varies widely from 1 to 26 years. I only include troopers with at least 500 stops because I want to compare troopers who consistently patrolled the same stretch of highway and want to ensure there are a sufficient number of stops to estimate each trooper's use of race, especially for relatively rare events such as searches. This section of Interstate 5 begins approximately 14 miles south of Seattle-Tacoma International Airport (Sea-Tac Airport) and ends approximately eight miles north of the airport, just south of Seattle's central business district (see map in Figure 3.2). With an average daily traffic flow of approximately 200,000 motorists, this is one of the busiest sections of the state highway system (Washington State Department of Transportation, 2005). The heavy traffic is mostly due to commuters driving to and from Seattle and air travelers using Sea-Tac Airport, which handles approximately 75,000 passengers per day.

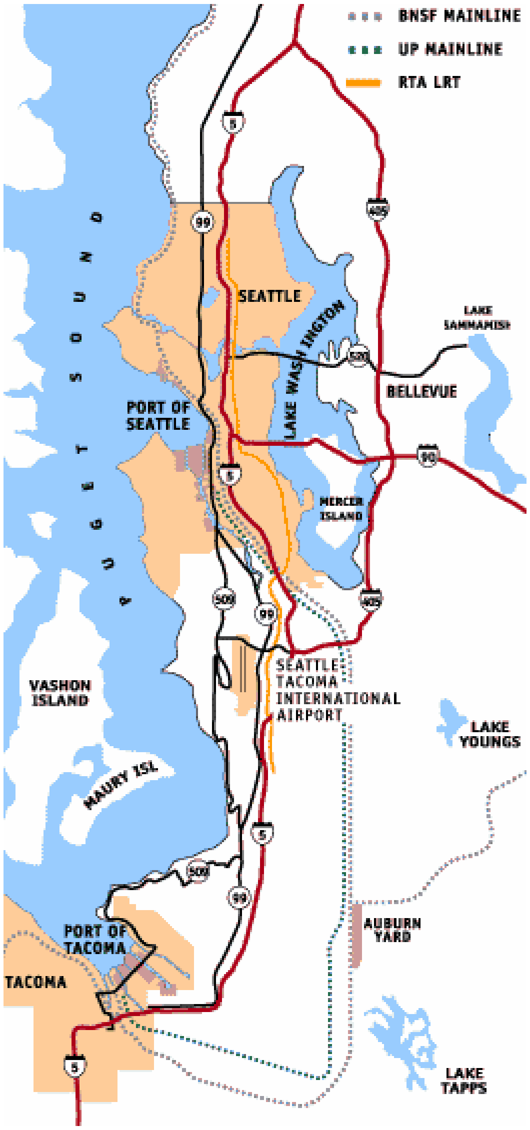
[54] In this study, I use the terms "driver" and "motorist" interchangeably.

[55] These racial group categories are based on the Washington State Department of Personnel and the U.S. Department of Commerce.

[56] *Lacy v. Villeneuve*, United States District Court, Western District of Washington at Seattle, Case No.: CV03-2442JLR (see Guest, 2005).

[57] Although these 16 troopers are from one of four detachments (including a "ghost" detachment described below), when I refer to the group of troopers, I use the term "detachment" for convenience.

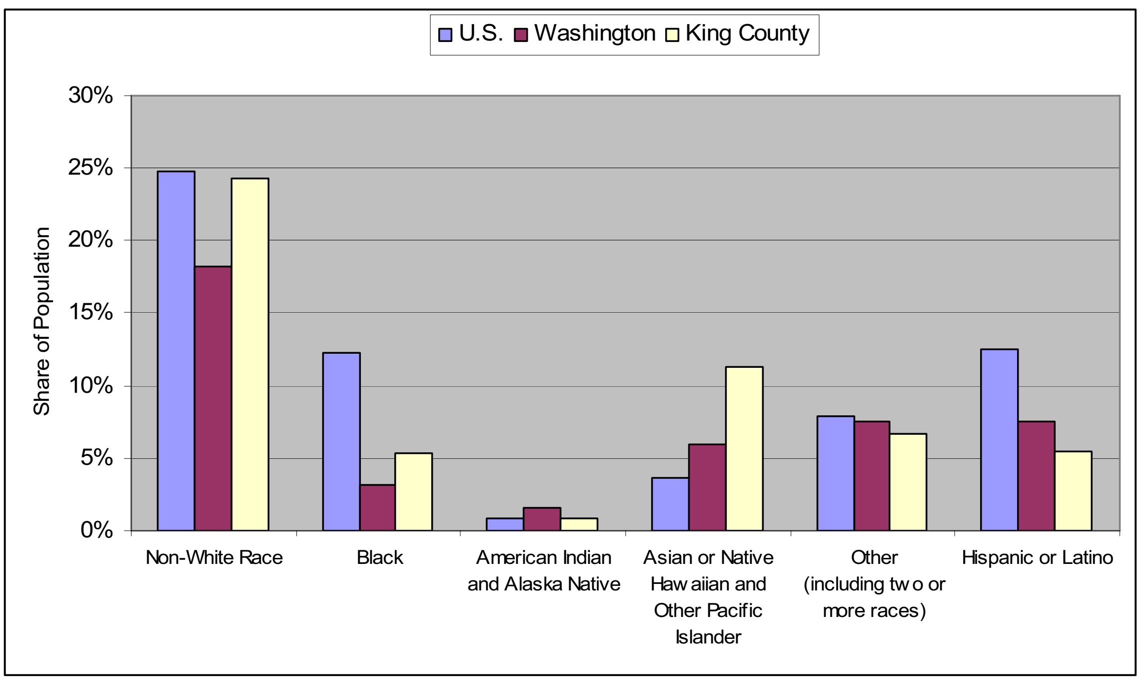
Figure 3.2: Map of Seattle-Tacoma Metropolitan Area



Source: Port of Seattle

Because Seattle commuters and Sea-Tac Airport travelers heavily influence the racial-group shares of motorists on this stretch of highway, the racial-group shares of residents estimated by the U.S. Census may not reflect the racial-group shares traveling on this stretch of highway. Notwithstanding, the census estimates provide some context about the local area. According to the 2000 census, Washington's population is 5.9 million and approximately 1.7 million live in King County, which has the same boundaries as WSP District 2, and includes the city of Seattle. Figure 3.3 shows the racial- and ethnic-group shares for U.S., Washington, and King County 2000 populations. The census separately considers an individual's race and ethnicity. An individual is classified into one of several racial groups as well as one of two ethnic groups, Hispanic/Latino or non-Hispanic/Latino. The first five sets of bars represent racial groups and the last set of bars represents the Hispanic/Latino ethnic group. The first set of bars, which include the non-white racial-group shares, is the sum total of the next four sets of bars, which include the shares of the non-white racial groups. The non-white population share of King County is similar to its U.S. share; however, there are some differences within non-white racial groups. The black population share of King County is lower than its U.S. share, but the Asian/Native Hawaiian/Other Pacific Islander population share of King County is higher than its U.S. share. Additionally, the Hispanic/Latino ethnic population share of King County is lower than its U.S. share.

Figure 3.3: 2000 Racial-Group and Ethnicity Shares for U.S., Washington, and King County Populations



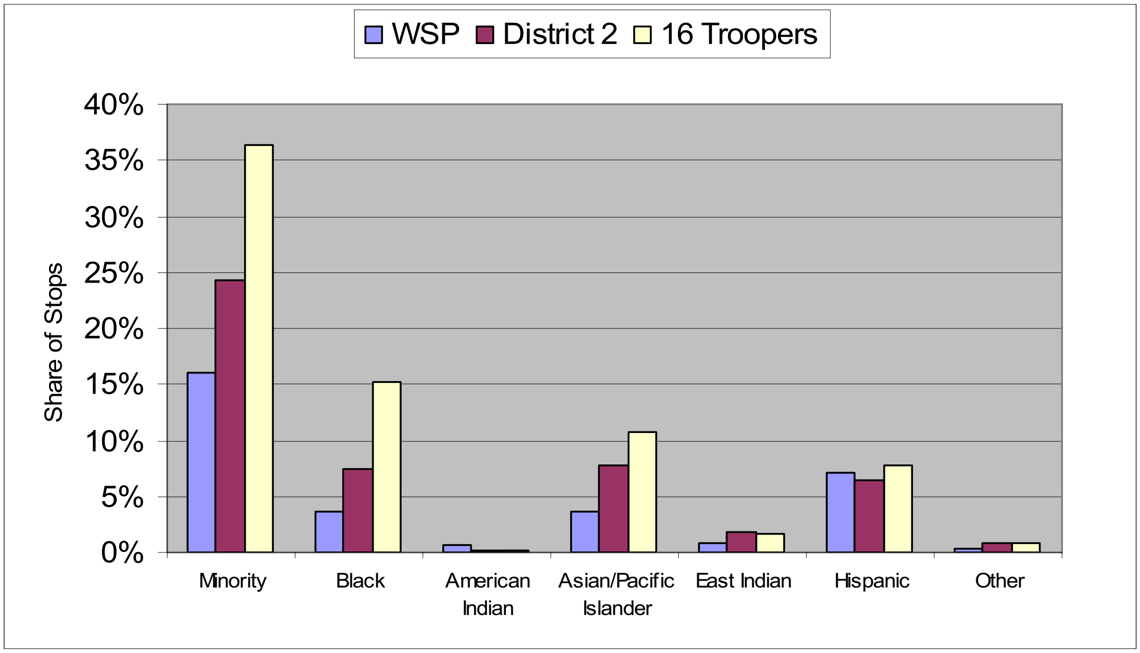
Source: U.S. Census (2000)

Figure 3.4 shows each racial group's share of stops for the WSP, District 2, and the 16 troopers for the nine-quarter period. Based on the results of Lovrich et al. (2003, 2005), for this study, I compare the outcomes of minorities (i.e., the latter seven racial groups listed above) with non-minorities (i.e., whites). Due to the small number of Pacific Islanders, I combine Asians and Pacific Islanders, which was done in Lovrich et al. (2005). The first set of bars, which include all minority motorists, is the sum total of the other sets of bars, which include the shares of each minority racial group. During this period, WSP troopers made 2.2 million stops, District 2 troopers made 350,000 stops, and the 16 troopers made 21,509 stops on Interstate 5 within the 22-mile range.[58] As compared to the WSP and District 2, these 16 troopers had a higher share of minority stops, especially blacks and Asians. This is partially explained by the census data above, which show that King County, which is the same geographical area as District 2, has a higher share of blacks and Asians as compared to the state.[59]

[58] I exclude the following contacts: contacts for commercial vehicle inspection and weighing operations; contacts involving commercial motorists or firms (i.e., wrecking yard, vehicle dealer, aircraft registration, hulk hauler, scrap processor, tow trucks); contacts involving pedestrians, hitchhikers, and bicyclists; contacts for parking violations; and contacts for non-traffic violations. These stops are excluded because this study's focus is on non-commercial motorists. A study involving commercial motorists or the other groups above would necessitate a different approach because the type of violation or reason for the contact with the WSP trooper is typically different from a non-commercial motorist contact. Additionally, I exclude emphasis patrol contacts because these contacts involve special assignments (e.g., DUI enforcement) around holidays.

[59] Because the U.S. Census separately classifies race and ethnicity, the non-white race share in Figure 3.3 does not include white Hispanics; hence, the non-white race share in that figure is not directly comparable to the minority share in Figure 3.4, which includes all Hispanics.

Figure 3.4: Shares of Stops by Race for the WSP, District 2, and the 16 Troopers



3.3 Review of the WSP Stop Analysis

Lovrich et al. (2003, 2005) analyzed traffic stop data from the WSP. In their stop analyses, they did not find that the WSP systematically used race as a factor to decide which motorists to stop, but they did have some concern of its use against minorities in particular geographical areas. However, as with many studies that estimate the use of race as a factor in stop decisions using external benchmarks, the accuracy of their benchmarks are questionable. The benchmarking methods used in each of their reports are similar, so I will primarily discuss Lovrich et al. (2005) since it analyzes traffic stops during a similar timeframe to the one used in this study.

Lovrich et al. (2005) analyzed stops that occurred between November 1, 2002 and June 30, 2004. They used the following five external benchmarks to estimate the racial-group shares of motorists at risk of being stopped within each APA: U.S. census; contacts based on calls for service and vehicles assists; contacts based on accidents; stops made during darkness; and radar- and aircraft-measured speeding stops. If the share of stopped motorists of race *R* differed by more than five percentage points from the benchmark-based share of race *R*, then Lovrich considered this difference substantively significant. I will address each benchmark in turn and present their results in Table 3.1.

Similar to many studies, Lovrich et al. (2005) compare the racial-group shares of stops with the U.S. census racial-group shares. They note that the census is not the ideal benchmark due to out-of-state motorists driving on Washington's highways and the high number of Hispanic migrant workers who are not captured by the census.

Second, they compare the racial-group shares from trooper contacts based on calls for services and trooper-initiated vehicle assists with the racial-group shares from trooper-initiated stops. This benchmark is based on the assumption that the same proportion of each racial group driving on the road is involved in calls for service or vehicle assists. However, calls for service and vehicles assists occur more with older vehicles, which may be associated with racial groups.

Third, they do the same comparison for accidents. This benchmark is based on the assumption that the same proportion of each racial group driving on the road is involved in accidents. This is a somewhat stronger assumption as compared to only including not-at-fault accidents; however, the data do not distinguish between accident types.[60]

Fourth, they compare each racial group's share of stops that occurred during daylight. This benchmark is based on troopers having a degraded ability to detect race when it is dark; hence, the stops occurring during darkness serve as a benchmark. For example, if 60 percent of white motorist stops occurred during daylight (with 40 percent at night) and 70 percent of black motorist stops occurred during daylight (with 30 percent night), then this is evidence that troopers used race as a factor to stop a higher share of minorities. However, the authors note that this assumes the racial-group shares of motorists do not change between daylight and darkness. Additionally, they assumed daylight always occurred between 7:00 a.m. and 7:00 p.m.; however, depending on the season of the year, sunrise can occur as early as 5:15 a.m. and sunset can occur as late as 9:00 p.m.

Fifth, similar to the darkness benchmark, Lovrich et al. (2005) state that a trooper typically cannot identify the race of a motorist when he measures his speed using radar/lidar or from an aircraft due to the distance between the trooper and motorist. The WSP uses both radar and lidar to measure a motorist's speed, but is increasingly using

[60] Alpert and Dunham (2004) assess non-at-fault traffic crash data as an external benchmark.

lidar, which has an operating distance of approximately 800 to 1200 feet.[61] Moreover, once a motorist's speed is measured and determined to exceed the threshold speed that warrants being stopped, the trooper will typically begin to enter traffic before the motorist passes, otherwise, due to the traffic on the freeway, it is difficult to pursue and stop the motorist. Hence, it is unlikely that a trooper would wait until the motorist passes in order to detect his race, and then decide whether to stop him. However, if the WSP employs radar and aircraft at various intensities at different locations throughout a day, week, or month, then the racial-group shares stopped by radar should not be expected to represent the average racial-group shares on the road at risk of being stopped.

Table 3.1 reports the results of the benchmarks at the WSP and APA 6 levels for each racial group.[62] The results are the difference (measured in percentage points) between the share of stopped motorists of race *R* and the benchmark share of race *R*, where a positive difference indicates the share of stopped motorists of race *R* exceeds the benchmark share of race *R*.

Table 3.1: Comparison of WSP and APA 6 Trooper Stops by Race with External Benchmark

Geographic Area and Method	Black	Native American	Asian	Hispanic	East Indian
WSP					
Calls for service	(1.5)	(0.3)	(1.1)	(0.6)	(0.2)
Collisions	(0.3)	(0.1)	(1.9)	(0.6)	0.4
Radar/Aircraft	1.4	0.2	0.4	2.5	(0.1)
APA 6					
Census	8.4	(0.7)	(1.0)	3.1	N/AV
Calls for service	(2.7)	(0.1)	3.2	0.7	0.5
Collisions	3.5	0.0	(0.7)	(0.8)	(0.8)
Daylight	(12.9)	(18.2)	(17.2)	(8.3)	(4.7)
Radar/Aircraft	3.6	0.0	2.4	(1.6)	0.0

Source: Lovrich et al., 2005

At the WSP level, the results from these benchmark comparisons do not show evidence that WSP troopers used race as a factor in deciding which motorists to stop.

[61] LIDAR, which stands for Light Detection and Ranging, is a laser speed detection system that uses a laser gun. LIDAR is preferred to radar because of its increased operating range and because the motorist's speed can be detected more quickly and accurately. For simplicity, I will refer to both radar- and LIDAR-measured stops as radar-based stops.
[62] The census and daylight benchmarks were not aggregated to the state level.

None of the percentage point differences exceeds the five-percentage point threshold. However, as stated above, these methods rely on assumptions that may not be true, making the results somewhat difficult to conclusively interpret. At the APA 6 level, with the exception of the daylight benchmark and the U.S. Census benchmark for blacks, none of the differences exceeds five percentage points. But again, the methods rely on assumptions that may not be true. The results from the daylight method provide evidence that troopers in APA 6 used race as a factor to stop a higher share of non-minority motorists. However, this is based on the assumption that the racial-group shares of motorists on the road do not change between daytime and nighttime. A study that relied on observers recording the race of each motorist and their type of violation (if any) could help answer this question; however, the cost of that study would be high. The results from the U.S. Census method provide evidence that troopers in APA 6 used race as a factor to stop a higher share of black motorists. However, as stated above, because many stops in this APA occur on a stretch of Interstate 5 that includes Seattle commuters and Sea-Tac Airport travelers, the racial-group shares represented by the U.S. Census may not reflect the racial-group shares of motorists at risk of being stopped in APA 6. In summary, due to methods' strong assumptions and the uneven results, the evidence is inconclusive whether WSP troopers used race as a factor in deciding which motorists to stop in APA 6.

3.4 Methods

This section describes the empirical method used to estimate the use of race as a factor in deciding which motorists are stopped. Based on the above discussion of the potential external benchmarks, I plan to use radar-measured speeding stops.[63] Based on interviews with the WSP and the Moose (2002) study, the radar-based stops represent an

[63] There are no aircraft-measured speeding stops in the data that I analyzed. I decided against using the U.S. Census and stops during darkness as external benchmarks. Regarding the census, this study is narrowly focused on a 22-mile stretch of a major interstate highway that crosses Sea-Tac Airport, Washington's busiest airport, and includes many commuters traveling to and from Seattle's central business district, which lies to the north. Given this situation, it is a very strong assumption that people living near or adjacent to this stretch of highway reflect the racial-group shares at risk of being stopped on this stretch of highway. Regarding darkness, I considered using the Grogger and Ridgeway (in press) method; however, because troopers only record the hour of the stop (not the minute), it severely reduces the number of stops that are known to have taken place either during darkness or daylight during clock times when both darkness and daylight stops occur.

accurate benchmark. Moose (2002) states: “It is readily accepted in the law enforcement community that uses of Radar/Laser instruments are vehicle selective, which makes them an excellent [external] benchmark.”[64]

I plan to use the Lovrich et al. (2005) methodology and improve upon it by adapting Grogger and Ridgeway’s (in press) methodology, which uses darkness as an instrument to indicate an officer’s ability to detect a motorist’s race. This will allow for the possibility that radar may have been used at different intensities over the evaluation period; hence, the racial-group shares stopped by radar would not be expected to represent the average racial-group shares at risk of being stopped, but instead would represent the average racial-group shares at risk of being stopped at the locations and time periods when radar was in use. I will also allow for the possibility that non-racial motorist characteristics (e.g., violation type, gender, and age) that are more visible during non-radar-based stops are associated with both being minority and being stopped. Lastly, I will estimate a separate model that only includes speeding stops since this is the only type of violation that can be detected using radar. This latter model allows for the possibility that minority and non-minority motorists commit different types of violations that differentially affect their probabilities of being stopped.

When a study is not based on a randomized control trial, there is often a concern that the treatment and control groups differ across characteristics that also affect the outcome of interest. These characteristics are considered confounding characteristics and are represented by the vector $\boldsymbol{x}$. In this case, it is important that motorists observed by radar possess the same characteristics that affect stop decisions as motorists observed by non-radar means. Regression analysis will control for confounding characteristics if particular assumptions are met. One major assumption is that the functional form of the regression model accurately represents the data generating process. The data generating process must be linear in the model’s parameters, the characteristics must be appropriately specified (e.g., transformed to x_1^2 or $\log(x_1)$), and the characteristics must be appropriately interacted (e.g., $x_1 \times x_2$). Moreover, when using a regression model,

[64] Note that Moose (2002) stated that Radar/Laser instruments are an excellent “internal” benchmark, meaning they are internal to the police department. To avoid confusion with my use of the term “internal benchmark,” I avoided its use here.

there is not a transparent method to evaluate whether the confounding characteristics were actually controlled for.

The propensity scoring method (Rosenbaum and Rubin, 1983) improves upon the regression model because it does not have the function form assumptions required in the regression model. Second, the method provides a transparent way to illustrate how well the confounding characteristics were controlled for. To explain the method, assume that when an officer is observing motorists in order to make a stop, he has the capacity to stop a motorist. When he stops a motorist, the stop is completed by the motorist.[65] Let S be a binary variable indicating whether a stop has been completed, where $S = 1$ indicates a stop that has been completed and $S = 0$ indicates the capacity to make a stop, but the stop has not been completed. The data only include completed stops ($S = 1$) because the data do not include observed motorists who were not stopped ($S = 0$).

Assume stop i will be completed by observing motorists using radar or a means other than radar (hereafter, "non-radar"). As discussed above, troopers have a degraded ability to detect a motorist's race when a stop will be completed using radar as compared to non-radar. In reality, a trooper's ability to detect a motorist race is not binary, but could theoretically be measured on a continuum. When non-radar observation is being used, race is not always perfectly detected, and contrastingly, when radar observation is being used, race may sometimes be partially detected. To explain this method, I will assume that the trooper's ability to detect race is binary, but will return to discuss the implications of this assumption in the discussion section.

If stop i is going to be completed using non-radar observation, then stop i is considered to be assigned to the treatment group ($V = 1$) since the motorist's race is visible during the non-radar observation. If potential stop i is going to be completed using radar observation, then stop i is considered to be assigned to the control group ($V = 0$) since race is not visible during radar observation. Let M_V indicate whether a stop involves a minority motorist when the stop is assigned to treatment level V. Depending on the treatment assignment, stop i has two potential outcomes: one outcome when $V =1$ (M_{1i}) and another outcome when $V = 0$ (M_{0i}), which is shown below.

---

[65] Although a specific motorist's race is fixed, the motorist that completes a specific stop is not fixed. This is the reason why the theoretical discussion focuses on the stop, not the motorist.

M_{1i} = 1 if stopped motorist *i* is minority if stopped by non-radar (V = 1; motorist's race is visible; treatment group)

0 if stopped motorist *i* is non-minority if stopped by non-radar (V = 1; motorist's race is visible; treatment group)

M_{0i} = 1 if stopped motorist *i* is minority if stopped by radar (V = 0; motorist's race is not visible; control group)

0 if stopped motorist *i* is non-minority if stopped by radar (V = 0; motorist's race is not visible; control group)

The objective is to estimate the difference in probabilities that stop *i* was completed by a minority motorist when stop *i* was assigned to the treatment group versus the control group. When this difference is averaged over all stops, this average difference is known as the average treatment effect (*ATE*), which is estimated using Eq. (3.1).

$$\text{Estimated } ATE = \frac{1}{N}\sum_{i=1}^{N} M_{1i} - M_{0i} \tag{3.1}$$

However, a given stop is assigned a particular treatment level; therefore, both M_{1i} and M_{0i} are not observed. Hence, the following are observed:

$M_{1i} \mid V = 1$

$M_{0i} \mid V = 0$

and the following are not observed:

$M_{1i} \mid V = 0$

$M_{0i} \mid V = 1$

In order to estimate the *ATE*, Rosenbaum and Rubin (1983) define a balancing score, $b(\boldsymbol{x})$, such that conditional on $b(\boldsymbol{x})$, the distribution of $\boldsymbol{x}$ is independent of the treatment assignment z, where z is a binary variable indicating the treatment assignment. In a randomized experiment, the independence of $\boldsymbol{x}$ and z naturally arises; however, for an observational study, this may not be the case. There are many possible balancing scores, the most trivial being $b(\boldsymbol{x}) = \boldsymbol{x}$. However, with this fine of a balance score, there will be few (if any) observations with the same $\boldsymbol{x}$ that exist in both the treatment and control groups. Hence, they define the propensity score as the coarsest function of $\boldsymbol{x}$ that is a balancing score in the sense that $e(\boldsymbol{x}) = f\{b(\boldsymbol{x})\}$ for some function f, whereby $e(\boldsymbol{x}) = \Pr(z = 1 \mid \boldsymbol{x})$.[66] They then demonstrate the *ATE* can be estimated by conditioning on the propensity score assuming that the "strong ignorability" assumption holds.

The strong ignorability assumption is that given $e(\boldsymbol{x})$, the joint distribution of (M_0, M_1) is independent of z, that is, one can ignore whether an observation was assigned to the treatment group or control group because that assignment does not affect the joint distribution of (M_0, M_1). The strong ignorability assumption is shown in Eq. (3.2), where f is the joint distribution function of (M_0, M_1).

$$[f(M_1, M_0) \perp z] \mid e(\boldsymbol{x}) \tag{3.2}$$

Remember that conditional on $e(\boldsymbol{x})$, the distribution of $\boldsymbol{x}$ is independent of the treatment assignment. If $e(\boldsymbol{x})$ contains all the relevant characteristics that affect the treatment assignment and the outcome, then the strong ignorability assumption will hold.[67] If this assumption holds, then conditional on $e(\boldsymbol{x})$, the observed distribution of M_0 for the control group can be used to estimate the unobserved distribution of M_0 for the treatment group because the two distributions are the same. Hence, the average treatment effect can be estimated.[68]

In order to estimate $e(\boldsymbol{x})$, a logistic regression function is often used, whereby the log-odds of the treatment assignment is estimated using Eq. (3.3).

$$\ln\left(\frac{\Pr(z = 1 \mid \boldsymbol{x})}{1 - \Pr(z = 1 \mid \boldsymbol{x})}\right) = \gamma_0 + \boldsymbol{x}\boldsymbol{\gamma}_1 \text{ , where } e(\boldsymbol{x}) = \Pr(z = 1 \mid \boldsymbol{x}) \tag{3.3}$$

In order to achieve a better estimate in terms of prediction error, I use a generalized boosted modeling (GBM) technique described in McCaffrey, Ridgeway, and Morral (2004) and Ridgeway (2006). The technique estimates the log-odds of being assigned to the treatment group by maximizing a Bernoulli log-likelihood function. However, unlike a logistic regression function, the technique does not require the functional form of $\boldsymbol{x}$ to

---

[66] $e(\boldsymbol{x})$ is coarser than $\boldsymbol{x}$ because $e(\boldsymbol{x})$ summarizes the k-dimensional vector $\boldsymbol{x}$ into a scalar.

[67] This assumption is the same as assuming there are no omitted variables in a regression model, that is, there is not an omitted variable that is both associated with the treatment (after controlling for the other covariates) and the outcome of interest.

[68] The above interpretation of Rosenbaum and Rubin (1983) was aided by McCaffrey, Ridgeway, and Morral (2004).

be specified, but instead, uses the variables within $\boldsymbol{x}$ to iteratively model regression trees to maximize the log-likelihood function. The initial tree is a very simple model (high bias and low variance) that grows increasing complex (which reduces bias, but increases variance). The optimal tree iteration is selected based on an algorithm that minimizes the average effect size differences across the characteristics between the treatment and control groups.[69]

Once the propensity scores are calculated, various algorithms have been used to choose which control group observations to match to each treatment observation (Deheji and Wahba, 2002). A simple approach is one-to-one matching where the control group observation with a propensity score closest to the treatment group observation is used as the match. This can be done with or without replacement. Other approaches include creating strata, nearest-neighbor averaging, or caliper averaging. I chose to use another approach where the propensity scores are converted to weights. The advantage of this approach is that the scores are applied more smoothly since all of the control group observations are used (for examples, see Rosenbaum, 1987; Hirano and Imbens, 2001; McCaffrey et al., 2004; and Riley et al., 2005; Ridgeway, in press). The exponentiation the log-odds of the treatment assignment can be considered a weight, $w(\boldsymbol{x})$, akin to probability-of-selection weights that are applied to observations in a survey sample in order to adjust for an observation's probability of being included in the sample.

The propensity score weights are applied in a manner to estimate the *ATE* on the population of interest. Using the method described in Ridgeway (in press) and Wooldridge (2002), I estimate the *ATE* on the treated stops (ATE_1). For every stop made when race is visible ($V=1$), one can imagine the counterfactual stop that would have been made had race not been visible ($V = 0$). The objective is to estimate the difference between the probability that a stop involved a minority motorist when the stop was made when race was visible (i.e., non-radar stops; $V = 1$) and the probability that the stop would have involved a minority motorist had the stop been made when race was not visible (i.e., radar stops; $V = 0$). This is ATE_1, which is represented by θ in Eq. (3.4).

[69] This algorithm as well as all the other computations in this study were implemented using R, which is a free, open source environment for statistical computing (see www.r-project.org).

$$\theta = \Pr(M_1 \mid S=1, V=1) - \Pr(M_0 \mid S=1, V=1) \tag{3.4}$$

The null hypothesis is that the probability that a stopped motorist is minority when race was visible prior to the stop ($V = 1$) is less than or equal to the probability that a stopped motorist is minority when race was not visible prior to the stop ($V = 0$). Under the null hypothesis, $\theta \le 0$.[70]

As stated above, M_{0i} is not observed when $V = 1$, but can be estimated from the data. To estimate $\Pr(M_0 \mid V = 1)$, the radar (control) stops are weighted using the propensity score weights so their weighted joint distribution of $\boldsymbol{x}$ will closely match the non-radar (treatment) stops' joint distribution of $\boldsymbol{x}$, as shown in Eq. (3.5), where f is the joint distribution function of $\boldsymbol{x}$.

$$f(\boldsymbol{x} \mid S=1, V=1) = w(\boldsymbol{x}) f(\boldsymbol{x} \mid S=1, V=0) \tag{3.5}$$

The weights are estimated using the GBM technique described above. The propensity score weighting model includes $\boldsymbol{x}$, which is composed of a vector of contextual characteristics $\boldsymbol{c}$ and a vector of motorist characteristics $\boldsymbol{m}$. $\boldsymbol{c}$ includes a set of binary variables indicating the nine quarters, a binary variable indicating weekday versus weekend, a set of binary variables indicating six four-hour time periods within a day, and a binary variable indicating whether the stop occurred south or north of Sea-Tac Airport. $\boldsymbol{m}$ includes a binary variable indicating whether the motorist was driving aggressively, a binary variable indicating the motorist's gender, and a set of binary variables indicating the motorist's age (16-25, 26-35, 36-45, or 46+ years old).

When Bayes theorem is applied to Eq. (3.5), the result is Eq. (3.6), where K is a constant that drops out of the result shown in Eq. (3.7).

$$w(\boldsymbol{x}) = K \frac{f(V=1 \mid S=1, \boldsymbol{x})}{f(V=0 \mid S=1, \boldsymbol{x})} \tag{3.6}$$

[70] A one-sided hypothesis test is used because the primary concern is the use of race as a factor to stop a higher share of minority motorists; however, if race is used as a factor to stop a higher share of non-minority motorists, that will be detected as well.

$f(V=1 \mid S=1, \boldsymbol{x})$ is the same propensity score as defined above ($e(\boldsymbol{x})$), that is, it is the probability that a stopped motorist with covariates $\boldsymbol{x}$ was stopped when $V = 1$. The propensity score weights are used to estimate the use of race as a factor in stop decisions, as defined by $\hat{\theta}$ in Eq. (3.7).

$$\hat{\theta} = \frac{\sum_{i=1}^{N} M_i V_i S_i}{\sum_{i=1}^{N} V_i S_i} - \frac{\sum_{i=1}^{N} M_i w(\boldsymbol{x}_i)(1-V_i)\, S_i}{\sum_{i=1}^{N} w(\boldsymbol{x}_i)(1-V_i)\, S_i} \tag{3.7}$$

To estimate the p-value of $\hat{\theta}$, I use the weights above in the following logistic regression model (Eq. (3.8)), where the p-value of $\hat{\theta}_1$ equals the p-value of $\hat{\theta}$.[71]

$$\ln\left(\frac{\Pr(M=1 \mid S=1, V)}{1-\Pr(M=1 \mid S=1, V)}\right) = \theta_0 + \theta_1 V \tag{3.8}$$

The propensity score weights can be evaluated to see how well they controlled for confounding characteristics. In order to assess how well the propensity score weights performed in matching the joint distribution of $\boldsymbol{x}$ between non-radar stops ($V = 1$) and weighted radar stops ($V = 0$), I evaluate the absolute and effect-size differences across characteristics. For a given characteristic, the absolute difference between the treated group and the weighted control group is informative for most characteristic values; however, the absolute difference can be less informative for some values. For example, for a binary variable such as gender, a 10-percentage-point difference between non-radar stops' and weighted radar stops' shares of male motorists is more significant if the non-radar stops' share is 1 percent versus 50 percent. An effect size difference captures this distinction, where the effect size difference is the absolute value of the difference between the mean of the characteristic for non-radar stops and the weighted mean of the characteristic for radar stops divided by the standard deviation of the characteristic for

[71] When weights are specified, by default, R estimates the standard errors using a robust estimator (i.e., a generalized Huber-Eicker-White sandwich estimator).

non-radar stops.[72] The effect size difference calculation for a binary variable x, which is the type of variable used in the propensity score model, is shown in Eq. (3.9), where $\bar{x} \mid V = 1$ is the mean of x for non-radar stops and $\bar{x}_w \mid V = 0$ is the weighted mean of x for radar stops.

$$\text{Effect size difference} = \frac{\left|\left(\bar{x} \mid V = 1\right) - \left(\bar{x}_w \mid V = 0\right)\right|}{\sqrt{(\bar{x} \mid V = 1)[1 - (\bar{x} \mid V = 1)]}} \tag{3.9}$$

Table 3.2 provides an illustration of effect size differences for different values of $\bar{x} \mid z = 1$ (which increase as one moves from the top to the bottom of the table) for different absolute differences between $\bar{x} \mid z = 1$ and $\bar{x}_w \mid z = 0$ (which increase as one moves from the left to the right side of the table).[73] The dark gray cells indicate where the effect size difference is greater than 0.2, and the light gray cells indicate where the effect size difference is greater than 0.1 but less than or equal to 0.2. Hence, because the standard deviation of $\bar{x} \mid z = 1$ increases as $\bar{x} \mid z = 1$ increases from 0.0001 to 0.5000, the same absolute difference results in a lower effect size difference as $\bar{x} \mid z = 1$ increases from 0.0001 to 0.5000. The goal is to obtain effect size differences of less than 0.2, and ideally, less than 0.1.

[72] In order to assess whether two sample proportions are statistically different, a chi-square test is normally performed on a two-way contingency table, resulting in a p-value. However, I evaluate effect-size differences instead of p-values for two reasons. First, as the variance of the weights increase, the effective sample size decreases, causing the p-value of the difference to increase. Hence, a weighting algorithm that increased the variance of the weights could artificially cause the differences to become insignificant. The second reason is due to the difficulty of calculating p-values when the cell counts are small (i.e., less than five). Although the Fisher exact p-value could normally be used in this situation, it could not be used here because the weights caused many of the cell counts to be non-integers.

[73] z is the generic treatment variable indicator defined above.

Table 3.2: Effect Size Differences for Various Treatment Group Means and Absolute Differences between Treatment and Control Group Means[74]

$\bar{x} \mid z=1$	$\lvert(\bar{x} \mid z=1)-(\bar{x}_w \mid z=0)\rvert$ 0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.11
0.0001	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00	11.00
0.0010	0.32	0.63	0.95	1.27	1.58	1.90	2.21	2.53	2.85	3.16	3.48
0.0100	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.01	1.11
0.0200	0.07	0.14	0.21	0.29	0.36	0.43	0.50	0.57	0.64	0.71	0.79
0.0300	0.06	0.12	0.18	0.23	0.29	0.35	0.41	0.47	0.53	0.59	0.64
0.0500	0.05	0.09	0.14	0.18	0.23	0.28	0.32	0.37	0.41	0.46	0.50
0.1000	0.03	0.07	0.10	0.13	0.17	0.20	0.23	0.27	0.30	0.33	0.37
0.2000	0.03	0.05	0.08	0.10	0.13	0.15	0.18	0.20	0.23	0.25	0.28
0.3000	0.02	0.04	0.07	0.09	0.11	0.13	0.15	0.17	0.20	0.22	0.24
0.4000	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22
0.5000	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	0.22

In the results section, a table is presented that shows effect size differences between the treatment group's joint distribution of $\boldsymbol{x}$ and the control group's weighted joint distribution of $\boldsymbol{x}$.

As a sensitivity analysis, I also estimate the use of race in stop decisions using the logistic regression model defined in Eq. (3.10), where $\hat{\delta}_1$ is an estimate of the use of race.[75] As discussed above, this model has strong functional form assumptions and unlike the propensity score model, one cannot assess how well the model controlled for confounding characteristics.

$$\ln\left(\frac{\Pr(M=1 \mid S=1, V, \boldsymbol{x})}{1-\Pr(M=1 \mid S=1, V, \boldsymbol{x})}\right) = \delta_0 + \delta_1 V + \boldsymbol{\delta}_2 \boldsymbol{x} \tag{3.10}$$

3.5 Results

For this analysis, I begin with the 21,509 stops displayed in Figure 3.4 and exclude stops when a trooper had relatively little discretion as to whether to make the stop. Hence, I exclude stops where a criminal violation is the primary violation that initiated the stop.[76] These stops primarily include DUI violations,[77] and to a lesser extent,

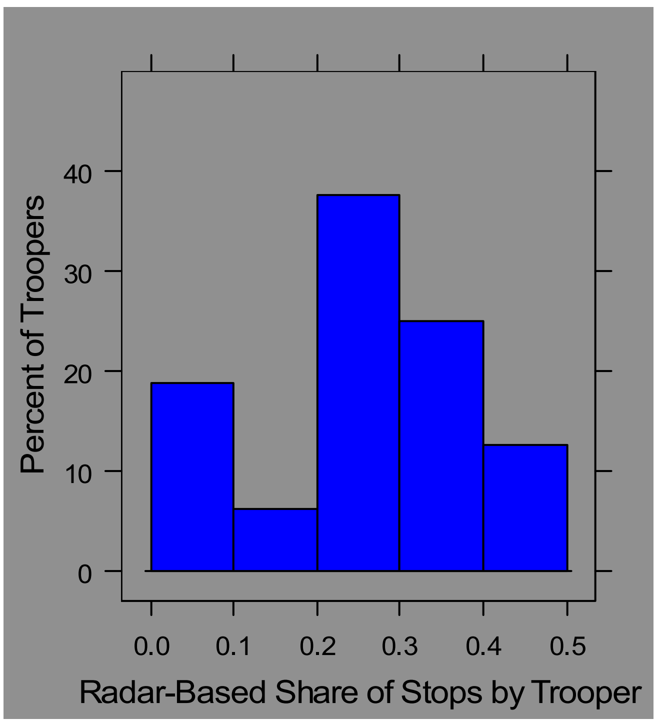
[74] The table is symmetric for $\bar{x} \mid z=1$ greater than 0.5. Note the effect size differences are rounded, but the colors are based on the more precise effect size differences not shown.

[75] So the results from the logistic regression and propensity score weighting models can be compared, the standard errors of the logistic regression model are also estimated using a robust estimator.
[76] The primary violation that initiated the stop is recorded as "violation one" in the data.

reckless driving and driving with a suspended license. I also exclude trooper-reactive stops since these stops primarily involved a trooper responding to an accident or a stranded motorist. None of the excluded stops was coded as a radar-based stop. After excluding these stops, there are 18,747 remaining stops.[78]

On average, these 16 troopers used radar for 24 percent of their stops, but did not have any aircraft-measured speeding stops. However, the troopers used radar at varying rates. The histogram in Figure 3.5 shows that the radar-based share of most troopers' stops was between 20 and 40 percent; however, the range varied from 0 to 45 percent.

Figure 3.5: Radar-Based Share of Stops by Trooper



[77] For most DUI violations, the primary violation that initiated the stop is either a lane or speeding violation, and the DUI violation is recorded as the second or subsequent violation. Based on conversations with the WSP, when a DUI is recorded as the primary violation, it is typically because the trooper responded to a motorist who was already stopped (e.g., due to an accident or being stranded). However, these contact types were not included above, so stops where a DUI is the recorded as the primary violation may have been miscoded.

[78] Of the original 21,509 stops, 36.4 percent were minority motorists. For these 18,747 stops, the minority share of motorists remains the same; therefore, dropping these stops is not, prima facie, disproportionately affecting one racial group. However, minorities represented 43.1 percent of the 364 criminal violations that were dropped.

Lovrich et al.'s (2005) test for the use of race as a factor in stop decisions involved comparing each racial group's share of stops when race was not visible ($V = 0$) to when race was visible ($V = 1$). Table 3.3 replicates this test for the stops above and shows that the minority share of stops increased eight percentage points when the stop was made when the motorist's race was more visible (i.e., when the stop was not radar-based).

Table 3.3: Minority Share of Stops for Radar-Based versus Non-Radar-Based Stops

Race	Radar-Based Stops (V=0)	Non-Radar-Based Stops (V=1)	Total
Minority	1,382	5,434	6,816
Non-Minority	3,210	8,721	11,931
Total	4,592	14,155	18,747
Percent Minority	30%	38%	36%

However, as discussed in the methods section, Table 3.3 could be misleading if the share of minority motorists on the road and the share of stops due to radar change differently throughout the evaluation period. For example, assume the 16 troopers do not use race as a factor in deciding which motorists to stop. On a particular day, they patrol from 8 a.m. to 4 p.m., and 15 troopers use radar from 8 a.m. to 12 p.m., while only one trooper uses radar from 12 p.m. to 4 p.m. Hence, the probability that a motorist is observed and stopped by radar decreases 15-fold between 12 p.m. and 4 p.m. Also assume the minority share of motorists at risk of being stopped substantially increases between 12 p.m. and 4 p.m. Hence, if the intensity of radar's use is not controlled for, then in this situation, the minority share of radar stops would be less than the minority share of non-radar stops, even though race was not used as a factor in the stop decisions. Therefore, it is important to control for contextual variables such as the time of the day, the day of the week, and the time of the year as well as the location of radar's use. This means comparing the minority shares of radar and non-radar stops within particular contexts, and then averaging the results across contexts. The table could also be misleading if non-racial motorist characteristics (e.g., violation type, gender, and age) that are more visible during non-radar-based stops are both associated with race and being stopped. These characteristics can also be controlled for within a model.[79]

Table 3.4 shows how stop characteristics are associated with radar-based stops and a stop motorist being minority. The first column of percentages shows the share of stops within each stop characteristic, both for contextual and motorist characteristics. The next two columns respectively show the radar-based and minority shares of stops within each stop characteristic. The first row summarizes the columns by including all 18,747 stops, and shows that 24 percent of the stops were made by radar and 36 percent of the stops were minority motorists. The table shows that both radar- and non-radar-based stops occur within each context; however, the radar-based share of stops varies among the contexts. The radar-based share of stops is higher on the weekend, during daytime hours, and south of Sea-Tac Airport. Radar-based stops also made up a higher share of aggressive driving stops, one-violation stops, non-criminal-violation stops, and stops involving younger motorists. The table also shows that minority motorists represented different shares of stops across these characteristics. The minority share of stops is higher during nighttime hours and north of Sea-Tac Airport. Minority stops also made up a higher share of stops involving lane violation, three or more violations, criminal violations, and younger motorists.

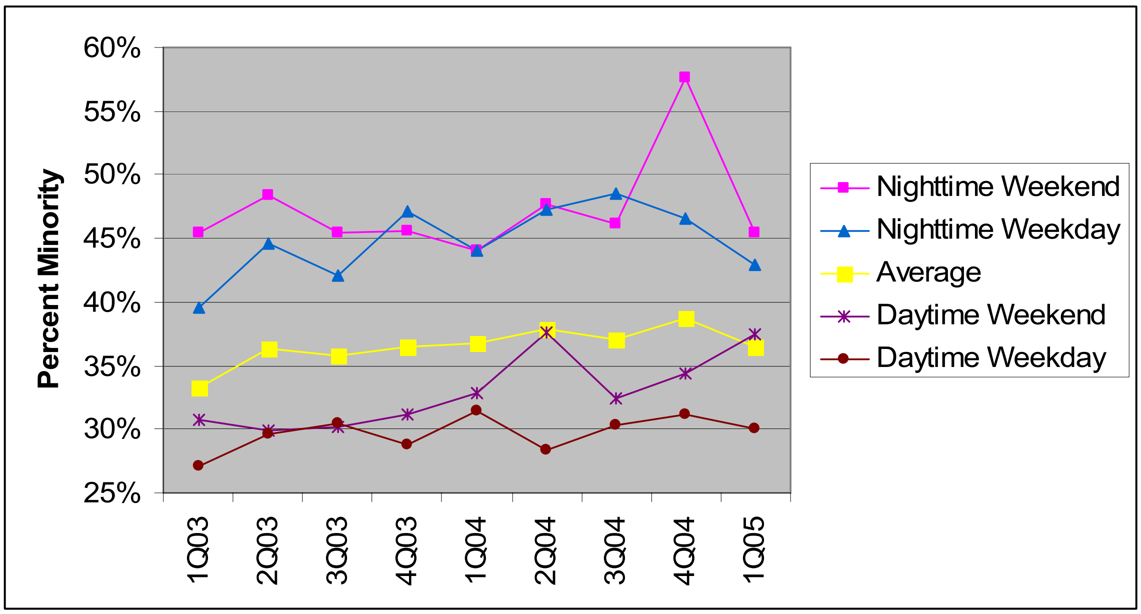
[79] On the other hand, if troopers target particular motorist characteristics (e.g., certain types of violations, gender, or age) because the characteristics are associated with minority motorists, then including these characteristics in the model masks the use of race. However, I chose to still include these characteristics for two reasons. First, when these characteristics are included, then if the result provides evidence that troopers were using race as a factor in their stop decisions, then the evidence is more conclusive since the motorist's characteristics that may be associated with race and being stopped have been controlled for. Second, while it is possible that some troopers engage in this practice (e.g., target specific violations in order to stop a particular racial group), the practice would have to be fairly sophisticated to not be detected. They would have to consistently target each racial group for the violation type. Additionally, this practice would be revealed in the descriptive statistics (e.g., seatbelt violations represent a very high share of one trooper's stops as compared to his peers).

Table 3.4: Total Shares, Radar-Based Shares, and Minority Shares within Each Stop Characteristic (all unweighted)

Stop Characteristic	Share of Stops (18,747 Stops)	Radar-Based Share of Stops	Minority Share of Stops
Total	100%	24%	36%
Quarter			
1Q03	14%	23%	33%
2Q03	12%	25%	36%
3Q03	13%	29%	36%
4Q03	12%	22%	36%
1Q04	12%	27%	37%
2Q04	12%	23%	38%
3Q04	9%	24%	37%
4Q04	9%	20%	39%
1Q05	7%	27%	36%
Day of Week			
Weekday	69%	20%	36%
Weekend	31%	35%	37%
Time of Day			
12 to 4 a.m.	22%	11%	48%
4 to 8 a.m.	12%	32%	34%
8 a.m. to 12 p.m.	18%	46%	30%
12 to 4 p.m.	16%	30%	27%
4 to 8 p.m.	15%	12%	33%
8 p.m. to 12 a.m.	16%	19%	41%
Milepost			
South of Sea-Tac Airport	74%	30%	33%
North of Sea-Tac Airport	26%	9%	45%
Primary Violation Type			
Speeding	55%	44%	34%
Equipment	11%	0%	31%
Safety Belt	6%	0%	37%
Lane Violation	17%	0%	46%
Other	11%	0%	36%
Contact Type			
Non-Aggressive Driving	88%	21%	36%
Aggressive Driving	12%	47%	38%
Number of Violations			
1	46%	29%	33%
2	30%	23%	37%
3+	24%	18%	41%
Number of Criminal Violations			
0	96%	25%	36%
1	3%	19%	52%
2+	1%	15%	52%
Motorist Gender			
Male	72%	24%	38%
Female	28%	27%	32%
Motorist Age			
16-25	34%	29%	38%
26-35	27%	27%	43%
36-45	19%	20%	34%
46+	20%	19%	27%
Motorist Race			
White	63.6%	27%	0%
Black	15.5%	20%	100%
American Indian	0.2%	21%	100%
Asian/Pacific Islander	10.7%	19%	100%
East Indian	1.5%	23%	100%
Hispanic	7.7%	20%	100%
Other	0.8%	31%	100%

Table 3.4 (above) shows the minority share of stops for a one-dimensional context (e.g., time of day); however, a stop can only defined using all the contexts. Due to the difference in the minority share of stops across the times of the day, Figure 3.6 combines the time-of-day and day-of-week contexts, and shows the share of minority stops for the combined contexts over the nine quarters. The share of minority stops was significantly higher during nighttime versus daytime, and was also slightly higher on the weekend versus the weekday. The figure also shows that the average share of minority stops stayed approximately the same over the latter eight quarters.

Figure 3.6: Minority Share of Stops by Quarter and Time Period of the Week[80]



Note that this does not necessarily mean that the share of minorities on the road necessarily increased during nighttime hours. It could be the case that the minorities on the road during nighttime hours were committing violations that made it more likely that they would be stopped as compared to minorities on the road during daytime hours. Additionally, troopers may use race as a factor to stop a higher share of minorities during nighttime hours, but may use race less (or not at all) during daytime hours.

[80] The number of stops for a particular quarter and a particular period of the week is between 144 and 1,039. The highest minority share of stops occurred during the 4Q04 during weekend nighttime hours; however, this particular context had the fewest number of stops (144), which means the minority share point estimate has the highest degree of uncertainty as compared to the other point estimates.

As seen in the above results, when estimating the use of race as a factor in stop decisions, it is important to allow for the possibility that radar may have been used at different intensities over the evaluation period. The use of race is estimated using a set of six models that include different subsets of stops, which vary based on the violation type and whether a stop was made during darkness or daylight. The first three models include all violations, and the second three models include only speeding violations. I estimate models only using speeding violations because radar is restricted to detecting speeding violations. If the racial-group shares of motorists at risk for being stopped for all violation types differ from the racial-group shares of motorists at risk for being stopped for a speeding violation, then the estimate of θ would be biased. For these models, I also exclude aggressive driving stops since the non-radar aggressive driving violation could be due to a violation other than excessive speeding.

Within each group of the three models, the first model includes all stops, the second includes daylight stops, and the third includes darkness stops.[81] I separately analyze stops during daylight and darkness because a trooper's ability to detect the race of the motorist is degraded during darkness, and is especially degraded if the trooper is using radar during darkness. Hence, when all the stops are used, the trooper is less able to detect the motorist's race for non-radar darkness stops as compared to non-radar daylight stops. Therefore, due to potential effects of the daylight and darkness on radar and non-radar comparisons, separate models are estimated that only include daylight or darkness observations.

To provide an example to assess how well the propensity score weights performed in matching the distribution of non-radar-based stop characteristics with the weighted distribution of radar-based stop characteristics, Table 3.5 shows the characteristics of non-radar stops as compared to the unweighted and weighted characteristics of radar

[81] Whether a stop occurred during daylight or darkness is based on sunrise, sunset, and civil twilight times downloaded from the U.S. Naval Observatory Web site for the city of Issaquah, which is located approximately 15 miles southeast of Seattle and is centrally located near the most populous areas of District 2. A stop is considered to have occurred during darkness if it occurred before morning civil twilight or after evening civil twilight, and a stop is considered to have occurred during daylight if it occurred between sunrise and sunset. For stops that occurred between morning civil twilight and sunrise or between sunset and evening civil twilight, they are coded as neither darkness nor daylight stops since the degree of natural lighting is ambiguous. In addition, because the time of the stop just includes the truncated hour (i.e., the minute of the stop is not reported), many additional stops near sunrise and sunset are neither coded as darkness nor daylight stops since the degree of natural lightning is also ambiguous.

stops. (This model includes all stops: all violations and both darkness and daylight stops.) With the exception of the interaction terms listed at the bottom of the table, the characteristics are sorted by their relative influence in predicting whether a motorist was stopped by non-radar versus radar.[82] The gray highlight indicates characteristics where the difference is greater than five percentage points or the effect size difference is greater than 0.1. As compared to radar stops, non-radar stops consisted of a higher proportion of nighttime stops, weekday stops, older motorists, and stops occurring north of Sea-Tac Airport. As right-side of the table illustrates, the weighted radar stops have similar characteristics to non-radar stops.

Weights reduce the effective sample size. Based on an identical sample size, a weighted mean has a larger sampling variance as compared to an unweighted mean, effectively reducing the sample size. In this case, the effective sample size of radar stops decreases from 4,592 to 1,112 stops.[83]

---

[82] The relative influence calculation is based on the non-interacted variables. The interacted variables are included in the table to evaluate how well the interacted variables matched between non-radar stops and weighted radar stops.

[83] Effective sample size (ESS) of radar stops: ESS = $\dfrac{\left(\sum_{i=1}^{N} w_i(1-V_i)\right)^2}{\sum_{i=1}^{N} w_i^{\,2}(1-V_i)}$, where V is 0 for a radar stop.

Table 3.5: Comparison between Non-Radar-Based and Radar-Based Stop Characteristics (unweighted and weighted)

Variable	Unweighted: Non-Radar-Based Stops (14,155 stops)	Unweighted: Radar-Based Stops (4,592 stops)	Unweighted: Difference (Non-Radar Minus Radar)	Unweighted: Effect Size Difference	Weighted: Weighted Radar-Based Stops (1,112 stops)	Weighted: Weighted Difference (Non-Radar Minus Radar)	Weighted: Effect Size Difference
Time							
12 to 4 a.m.	26.4%	9.8%	16.6%	0.38	26.5%	-0.1%	0.00
4 to 8 a.m.	11.0%	16.1%	-5.1%	0.16	10.5%	0.5%	0.02
8 a.m. to 12 p.m.	13.1%	34.5%	-21.4%	0.63	13.6%	-0.6%	0.02
12 to 4 p.m.	14.8%	19.2%	-4.4%	0.12	15.4%	-0.6%	0.02
4 to 8 p.m.	17.3%	7.6%	9.7%	0.26	16.4%	0.8%	0.02
8 p.m. to 12 a.m.	17.5%	12.9%	4.6%	0.12	17.6%	-0.1%	0.00
Quarter							
1Q03	13.8%	12.7%	1.0%	0.03	14.1%	-0.3%	0.01
2Q03	12.2%	12.2%	0.0%	0.00	11.3%	0.8%	0.02
3Q03	12.6%	15.6%	-3.0%	0.09	14.0%	-1.4%	0.04
4Q03	11.9%	10.6%	1.3%	0.04	13.1%	-1.2%	0.04
1Q04	11.8%	13.4%	-1.6%	0.05	10.6%	1.2%	0.04
2Q04	12.2%	11.4%	0.8%	0.02	11.3%	0.9%	0.03
3Q04	9.0%	8.9%	0.1%	0.00	9.0%	0.0%	0.00
4Q04	9.4%	7.0%	2.4%	0.08	9.5%	-0.1%	0.00
1Q05	7.2%	8.2%	-0.9%	0.04	7.2%	0.0%	0.00
Motorist Age							
16-25	31.8%	39.7%	-7.9%	0.17	32.4%	-0.6%	0.01
26-35	26.1%	29.1%	-3.0%	0.07	26.8%	-0.7%	0.02
36-45	20.4%	15.3%	5.1%	0.13	20.2%	0.2%	0.00
46+	21.7%	15.9%	5.8%	0.14	20.6%	1.1%	0.03
Contact Type							
Non-Aggressive Driving	91.3%	76.6%	14.6%	0.52	91.4%	-0.1%	0.00
Aggressive Driving	8.7%	23.4%	-14.6%	0.52	8.6%	0.1%	0.00
Milepost							
South of Sea-Tac Airport	68.2%	90.7%	-22.5%	0.48	71.1%	-2.9%	0.06
North of Sea-Tac Airport	31.8%	9.3%	22.5%	0.48	28.9%	2.9%	0.06
Day							
Weekday	73.3%	56.7%	16.6%	0.38	74.3%	-1.0%	0.02
Weekend	26.7%	43.3%	-16.6%	0.38	25.7%	1.0%	0.02
Motorist Gender							
Male	72.8%	69.3%	3.5%	0.08	72.9%	0.0%	0.00
Female	27.2%	30.7%	-3.5%	0.08	27.1%	0.0%	0.00
Gender:Age							
Male, 16-25	22.1%	25.2%	-3.1%	0.08	21.9%	0.1%	0.00
Female, 16-25	9.7%	14.5%	-4.8%	0.16	10.4%	-0.7%	0.02
Male, 26-35	19.1%	21.1%	-2.0%	0.05	20.6%	-1.4%	0.04
Female, 26-35	6.9%	7.9%	-1.0%	0.04	6.2%	0.7%	0.03
Male, 36-45	15.3%	11.5%	3.8%	0.10	14.4%	0.9%	0.02
Female, 36-45	5.1%	3.8%	1.3%	0.06	5.9%	-0.7%	0.03
Male, 46+	16.3%	11.4%	4.9%	0.13	15.9%	0.4%	0.01
Female, 46+	5.4%	4.5%	0.9%	0.04	4.7%	0.7%	0.03
Time:Day							
12 to 4 a.m., Weekday	19.5%	7.4%	12.1%	0.31	20.7%	-1.2%	0.03
4 to 8 a.m., Weekday	8.3%	7.9%	0.4%	0.02	7.4%	0.9%	0.03
8 a.m. to 12 p.m., Weekday	8.8%	15.6%	-6.7%	0.24	8.9%	-0.1%	0.00
12 to 4 p.m., Weekday	10.6%	11.8%	-1.1%	0.04	11.2%	-0.6%	0.02
4 to 8 p.m., Weekday	13.2%	4.6%	8.6%	0.25	12.7%	0.4%	0.01
8 p.m. to 12 a.m., Weekday	13.0%	9.5%	3.4%	0.10	13.3%	-0.4%	0.01
12 to 4 a.m., Weekend	6.9%	2.4%	4.5%	0.18	5.8%	1.1%	0.04
4 to 8 a.m., Weekend	2.7%	8.2%	-5.6%	0.34	3.0%	-0.3%	0.02
8 a.m. to 12 p.m., Weekend	4.2%	18.9%	-14.7%	0.73	4.7%	-0.4%	0.02
12 to 4 p.m., Weekend	4.1%	7.4%	-3.3%	0.16	4.1%	0.0%	0.00
4 to 8 p.m., Weekend	4.1%	3.0%	1.1%	0.06	3.7%	0.4%	0.02
8 p.m. to 12 a.m., Weekend	4.5%	3.4%	1.2%	0.06	4.3%	0.3%	0.01

I repeat this process for the other five models described above, and Figure 3.7 summarizes the unweighted and weighted effect size differences for all six models. The histogram on the left shows the effect size differences between the characteristics of non-radar stops and the unweighted characteristics of radar stops, and the histogram on the right shows the differences between the characteristics of non-radar stops and the weighted characteristics of radar stops. Based the aggregate number of characteristics included within the six models, each histogram includes 215 effect size differences.[84] The histogram on the right shows that the weighting eliminates the major differences.

Figure 3.7: Effect Size Comparison between Non-Radar-Based and Radar-Based Stop Characteristics (unweighted and weighted)

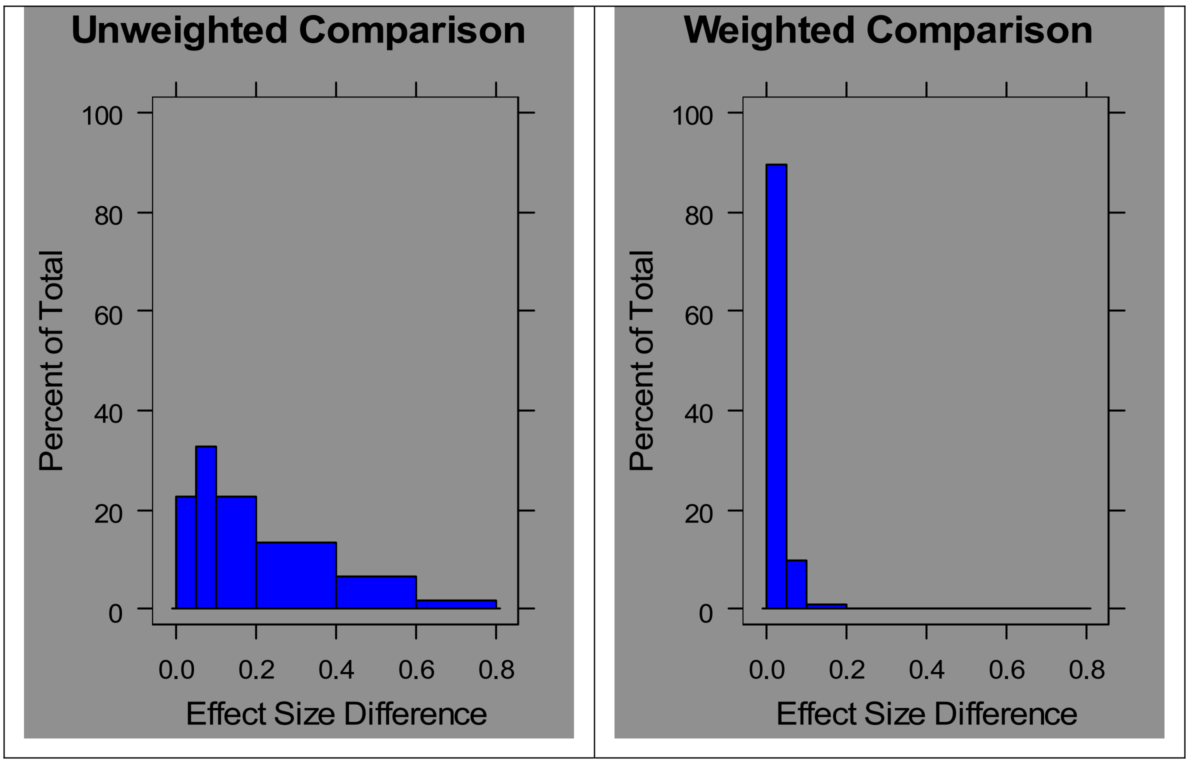


Table 3.6 shows the results from the six models. The first set of five columns includes the primary results, followed by a set of four columns that includes additional information. The first set of columns includes the following: description of the set of stops that were analyzed, the minority share of non-radar stops, the weighted minority share of radar stops, the difference between these shares (which is $\hat{\theta}$, see Eq. (3.7)), and

[84] The variable levels that are compared are listed in Table 3.4; however, some of the models did not include each level because some were not applicable when a subset of stops was used. If a variable only has two levels (e.g., gender), then one of the effect size differences is dropped since both are the same.

the one-sided p-value of $\hat{\theta}$. In the second set of columns, the first column is the unweighted minority share of radar stops, followed by the number of non-radar stops, the effective number of radar stops, and the total number of stops that are analyzed. For each model, the minority share of non-radar stops is consistently higher than the weighted minority share of radar stops; however, the one-sided p-value is only significant at the 0.05 level for the models that include all stops and darkness stops when all violations are considered. Based on this standard, the null hypothesis is rejected for these two models. The one-sided p-value represents the probability that the estimate of θ would be this or more extreme, assuming null hypothesis were true (i.e., $\theta \leq 0$).[85]

Table 3.6: Estimated Use of Race for Stop Decision

Types of Stops	Minority Share of Non-Radar Stops	Minority Share of Radar Stops (weighted)	Difference	p-value (weighted) (one-sided)	Minority Share of Radar Stops (unweighted)	Number of Non-Radar Stops	Effective Number of Radar Stops	Total Number of Stops Analyzed
All Violations								
All stops	38.4%	34.9%	3.5%	0.015*	30.1%	14,155	1,112	18,747
Daylight stops	31.0%	28.9%	2.1%	0.088	27.8%	5,893	1,071	8,962
Darkness stops	45.3%	39.8%	5.6%	0.013*	36.4%	6,750	437	7,842
Speeding Violations								
All stops	36.8%	34.3%	2.4%	0.079	28.7%	4,839	1,041	8,358
Daylight stops	29.6%	28.0%	1.6%	0.163	26.6%	1,911	1,294	4,308
Darkness stops	42.5%	38.6%	3.9%	0.077	34.7%	2,446	398	3,218

*Significantly higher minority share of non-radar stops as compared to weighted minority share of radar stops at the 0.05 level (one-sided)

The logistic regression model in Eq. (3.10) is also used to estimate the use of race for each model, and the parameter $\hat{\delta}_1$ is consistently greater than zero and the one-sided p-value is significant at the 0.05 level for all six models.

As a sensitivity analysis, I estimate four additional sets of six propensity score weighting models, where the six models within each set are defined in the same manner as the six models above (i.e., by the violation type and whether a stop was made during darkness or daylight). The second set of models (i.e., the first set of additional models) adds two binary variables to each model. One variable indicates whether more than one violation was recorded and the other indicates whether a criminal violation was recorded.

[85] Note that I observe the full census of each trooper's stops during the measurement period; however, there is still sampling uncertainty surrounding the estimates. This is because the models assume the observed stops represent a sample of the troopers' total lifetime stops. As with any sample, as the number of observations increase, the sampling uncertainty will decrease.

I did not include these binary variables in the original models because violations other than the primary violation that initiated the stop are mostly detected after the stop decision has been made (e.g., failure to possess vehicle liability insurance). However, the primary violation that initiated the stop may actually comprise of more than one violation, including a criminal violation.[86] When troopers are not using radar, they are better able to detect motorists committing multiple or criminal violations. As shown in Table 3.4, radar stops are associated with one-violation, non-criminal stops, and these stops are associated with non-minority stops. Notwithstanding, the results from this set of models are essentially the same as the results presented in Table 3.6. The null hypothesis is rejected for the models that include all stops and darkness stops when all violations are considered.

The third, fourth, and fifth sets of models estimate the use of race as a factor in stop decisions on different subsets of motorists, defined by the motorist's race, gender, and age. The third set of models excludes Asians/Pacific Islanders and motorists whose race is recorded as "other." Of the remaining minority motorists, 93 percent are either black or Hispanic and the rest are either American Indian or East Asian. These stops are separately analyzed because the majority of racial profiling complaints are from black and Hispanic motorists (e.g., see Meeks, 2000). The fourth set of models begins with the stops included in the third set, but only includes male motorists who are under 46 years old. These stops are separately analyzed because young males represent a higher share of racial profiling complaints as compared to their share of the minority population (Newport, 1999). When the use of race is estimated across all minorities, then the use of race to the detriment of young, male minorities may not be detected, especially if the share of young, male minorities within the sample of stopped minority motorists is small. The fifth set of models begins with the stops included in the fourth set, but only includes non-minority and black motorist stops. These stops are separately analyzed because racial profiling mostly affects black motorists.

[86] As discussed above, I exclude stops where a criminal violation is the primary violation that initiated the stop since these stops involved relatively less trooper discretion. However, a criminal violation that is recorded as a second or subsequent violation may have been detected prior to the stop, and these stops have not been dropped from the analysis.

The results from the third and fourth sets of models are mostly similar to the results presented in Table 3.6. The null hypothesis is rejected for the models that include all stops and darkness stops when all violations are considered. However, in the fourth set of models, the minority share of non-radar stops is less than the weighted minority share of radar stops when only speeding stops during daylight hours are considered; however, the difference is not statistically significant at the 0.05 level. The results from the fourth set of models are presented in Table 3.7, which includes the same rows and columns as Table 3.6.

Table 3.7: Estimated Use of Race for Stop Decisions of Non-Asian, Male Motorists Under 46 Years Old[87]

Types of Stops	Minority Share of Non-Radar Stops	Minority Share of Radar Stops (weighted)	Difference	p-value (weighted) (one-sided)	Minority Share of Radar Stops (unweighted)	Number of Non-Radar Stops	Effective Number of Radar Stops	Total Number of Stops Analyzed
All Violations								
All stops	34.7%	31.0%	3.7%	0.021*	26.2%	6,962	856	9,339
Daylight stops	27.8%	27.3%	0.6%	0.392	24.3%	2,773	647	4,291
Darkness stops	40.7%	34.8%	5.9%	0.034*	30.7%	3,432	274	4,061
Speeding Violations								
All stops	32.0%	29.6%	2.4%	0.148	24.8%	2,438	592	4,210
Daylight stops	23.7%	24.9%	-1.3%	0.724	23.0%	909	778	2,069
Darkness stops	38.1%	33.7%	4.5%	0.104	28.7%	1,282	242	1,710

*Significantly higher minority share of non-radar stops as compared to weighted minority share of radar stops at the 0.05 level (one-sided)

In the fifth set of models, which only include non-minority and black male motorists under 46 years old, the differences between the black share of non-radar stops and the weighted black share of radar stops have the same signs as the differences presented in Table 3.7, but their magnitudes are lower. The differences are not statistically significant at the 0.05 level.

The logistic regression model in Eq. (3.10) is also used to estimate the use of race for the four additional sets of models, and the parameter $\hat{\delta}_1$ is consistently greater than zero and its one-sided p-value is significant at the 0.05 level, except when only daylight speeding stops are considered.

[87] These models exclude Asians and motorists whose race was recorded as "other."

3.6 Discussion

The above models estimated the use of race as a factor in stop decisions, where the use-of-race estimate includes both appropriate and inappropriate uses of race. As discussed in Chapter 2, racial profiling is defined as when the use of race is inappropriate, as defined by a law or policy. For these troopers, an appropriate use of race may include using race when it is part of a particular suspect's description or part of a criminal profile. For example, if a particular suspect was fleeing a crime scene using this stretch of Interstate 5, a trooper may use the suspect's race as one factor among many to stop motorists who matched the suspect's description. While the law is more restrictive on the use of race when it is part of a criminal profile, there may be some situations when its use may be appropriate. The WSP states that approximately 75 percent of methamphetamine being consumed in Washington state originates from Mexico (Batiste, 2005). In this situation, whether a motorist's Hispanic appearance may be used as one factor among many in a stop decision depends on whether WSP has additional specific information about the methamphetamine couriers and when they were using this stretch of Interstate 5. Although the above two situations are plausible appropriate uses of race, these 16 troopers initiated the vast majority of their stops within the "core four" priorities of traffic enforcement: driving under the influence (DUI), dangerous speeding, seat belt, and aggressive driving enforcement. Hence, assuming that the use-of-race estimates above controlled for all confounding characteristics (and were statistically significant), then they mostly comprise of estimates of the inappropriate use of race.

The results provide evidence that race may have been used as a factor to stop a higher share of minority motorists; however, the evidence is not conclusive.[88] For the

[88] If the troopers used race as a factor to stop a higher share of minority motorists, it is a separate question whether this means that they stopped additional minority motorists or stopped fewer non-minority motorists. For example, in Scenario *A*, assume that a race-neutral trooper (i.e., a trooper who does not inappropriately use race as a factor in his traffic enforcement decisions) stops 100 motorists, of which, 20 are minority and 80 are non-minority. In Scenario *B*, assume that this trooper inappropriately uses race as a factor in his stop decisions, resulting in a higher share of minority motorists. If the trooper does not change the number of stops between the two scenarios, then in Scenario *B*, he could stop, for example, 25 minorities and 75 non-minorities. However, if he does change the number of stops, then in Scenario *B*, he might stop additional minorities as compared to Scenario *A* (e.g., 25 minorities and 80 non-minorities) or fewer non-minorities as compared to Scenario *A* (e.g., 20 minorities and 75 non-minorities). If the troopers in this study used race as a factor in their stop decisions, it is beyond the scope of this study to determine whether and to what degree they changed the number of their stops as compared to a situation where they did not use race as a factor in their stop decisions. This is an area of research that should be pursued.

models that included all stops, the results show that a stopped motorist had a higher probability of being minority when the stop was based on radar versus non-radar. This is especially true of nighttime stops. However, the results from the all-stop analyses may be biased if minority and non-minority motorists were committing different types of violations that differentially affected their probabilities of being stopped. When only speeding stops are considered, which are the only type of violation that can be detected using radar, there is substantial uncertainty surrounding the estimated use of race. As compared to when all stops are considered, the uncertainty in the estimated use of race increased partly because of the decreased difference between the minority share of non-radar stops and weighted minority share of radar stops, and partly because of the smaller sample size. In summary, while the results provide evidence that race may have been used as a factor to stop a higher share of minority motorists, due to the uncertainty surrounding the estimates when only speeding stops are considered, the evidence is not conclusive.[89] As more stops are analyzed from these troopers, those results will provide important evidence as to whether the above results are due to chance or are consistent with the troopers using race as a factor to stop a higher share of minority motorists.

As compared to Lovrich et al. (2005), which analyzed stops in APA 6 during a similar period, the results above are consistent with their radar/aircraft results, which showed that a higher share of minorities were stopped when radar/aircraft was not in use, but like the results from this study, there was substantial uncertainty (see Table 3.1).

For this method to determine whether race is being used as a factor in stop decisions, it is necessary for a trooper to have a degraded ability to identify the motorist's race when he is using radar (or lidar) as compared to non-radar. It is not necessary for race (or proxies for race such as the age of the vehicle) to be fully invisible when using radar, and moreover, it is not necessary for race to be fully visible when not using radar. At the extremes—race is fully invisible when using radar and fully visible when not—then the method is a quantitative estimate of the use of race. However, the estimate becomes more qualitative if the extreme conditions are not met, that is, the estimate tests

[89] The results from the logistic regression models provide evidence that race may have been used as a factor to stop a higher share of minority motorists; however, the models are based on functional form assumptions that may not be true. Because the propensity weighting score models were able to closely match the weighted characteristics of radar stops to the characteristics of non-radar stops, these models are relatively less likely to produce biased results.

whether race is being used as a factor, not the degree that it is being used.[90] However, the magnitude of the estimated use of race in the qualitative case can be interpreted as the minimum use of race. This interpretation is based on the fact that as a trooper's ability to identify a motorist's race becomes more similar for radar-based and non-radar-based stops, then if the actual influence of race is positive, the magnitude of the estimated use of race decreases.[91] Hence, the results above should be interpreted as the troopers' minimum use of race.

While there is strong support for the premise that a trooper is less able to detect the motorist's race prior to the stop decision when using radar versus when not using radar (e.g., see Moose (2002)), a trooper who desired to use race as a factor in his stop decisions could possibly do so while using radar. For example, while a motorist's race may be difficult to detect when using radar, a trooper may be able to detect vehicle characteristics (e.g., age or type of vehicle) that are associated with a motorist's race and might use these characteristics as a factor in his stop decisions. However, a trooper is less able to detect the vehicle characteristics when using radar as compared to when he is not.

The second way a trooper could use race as a factor in his stop decisions while using radar is based on when he makes his stop decision and whether he is working alone or as part of a team. When a trooper is using radar, he may work alone or as part of teams, where one trooper operates the radar and radios other troopers to make the stops. If a trooper desired to use race as a factor in his stop decisions while using radar when working alone, the trooper may wait for a speeding motorist to pass in order to identify his race before making the stop decision. However, this practice would be hard to implement on Interstate 5 due to the difficulty of catching up to a speeding motorist from a parked position versus pulling out into traffic ahead of the motorist, which is standard

[90] For a proof of using darkness as an instrument to detect the use of race, see Grogger and Ridgeway (in press), Appendix A. They demonstrate that if race is being used as a factor to stop a higher share of minority motorists and the motorist's race is less detectable during darkness as compared to during daylight, then the estimate of θ will be greater than zero. In principle, darkness is the same instrument as radar.

[91] To illustrate, consider an extreme example. Let a trooper make 200 stops, 100 using non-radar and 100 using radar. For non-radar stops, assume he perfectly detected each motorist's race prior to his stop decision, and for radar stops, assume he perfectly detected all but one of the motorists' races prior to his stop decision. Hence, the estimated influence of race will be on the order of one-hundredth of the actual influence of race, which diminishes the test's power to detect the use of race.

practice. On the other hand, when troopers are working as part of a team, where one trooper operates the radar and radios other troopers who are waiting up the interstate to intercept a speeding motorist, then the first trooper could more easily choose to identify the motorist's race before making the stop decision. While this is possible, the trooper would need the cooperation of the troopers making the stops because they would begin to recognize if a trooper consistently radioed them to stop minority motorists.

The primary purpose of this chapter was to estimate the 16 troopers' average use of race as a factor in deciding which motorists to stop. Although the estimate is important by itself, the estimate is primarily needed to better interpret the trooper-level use-of-race estimates for stop decisions, which is the focus of the next chapter. If the average use of race was the primary interest, then the methods above would have been applied to all stops within APA 6 or across the entire WSP. But instead, the methods were only applied to these 16 troopers' stops so that if a trooper is identified as using race as a factor more than his peers, then the average use-of-race estimate can help interpret that result.

Overall, the results provide evidence that race may have been used as a factor to stop a higher share of minority motorists, but due to the uncertainty surrounding the estimates, the evidence is not conclusive. However, based on the results, it is somewhat unlikely that this group of troopers on average used race as a factor to stop a higher share of non-minority motorists. Therefore, based on the analyses in the next chapter, if a trooper is identified as having stopped a significantly higher share of minority motorists than his similarly situated peers, then it is likely that he used race as a factor to stop a higher share of minority motorists as compared to a race-neutral trooper, that is, a trooper who does not inappropriately use race as a factor in his traffic enforcement decisions. On the other hand, if the above results had shown that race had been used as a factor to stop a higher share of non-minority motorists, then if a trooper is identified as having stopped a significantly higher share of minorities than his similarly situated peers, then it is ambiguous whether he used race as a factor to stop a higher share of minority motorists as compared to the race-neutral trooper. Instead, as compared to his peers, he may have least used race as a factor in deciding to stop a higher share of non-minority motorists. Hence, the results above are sufficiently conclusive in order to inform the interpretation of the trooper-level results in the next chapter.

Chapter 4: Officer-Level Stop Analysis

4.1 Introduction

Although the method in the previous chapter only focused on estimating the use of race as a factor for 16 troopers' stop decisions, the same method could be applied to estimate the use of race at the department level. The vast majority of racial profiling studies have been done at the department level; however, like the above detachment-level analysis in the previous chapter, they have two major drawbacks. First, it is difficult to identify an external benchmark that accurately represents each racial group's share of motorists violating a traffic law, while also accounting for each racial group's relative exposure to the police (which is less of an issue for studies analyzing highway stops since troopers are more uniformly distributed as compared to a city police force). Second, department-level studies only estimate the average use of race and do not capture potential variation in the use of race among officers. Therefore, even if a department-level study finds no evidence of the use of race, the actions of a few officers who are using race will likely be hidden. Moreover, if a department-level study finds evidence that race is being used, the police leadership is not able to determine whether its use varies among officers. The leadership's corrective action would differ if it knew almost all officers were using race versus a situation where only a few officers were identified as using race.

This chapter estimates each trooper's relative use of race as a factor in stop decisions in order to see if particular troopers use race more than their peers. This approach helps mitigate the drawbacks above. First, because officers serve as each other's benchmark, the often elusive external benchmark is not needed. When officers serve as each other's benchmark, the benchmark is called an internal benchmark since the benchmark is internal to the police department. Second, an officer-level analysis can be incorporated into an EI system. If a trooper uses race significantly more than his peers, this analysis will be able to identify him, which could result in further scrutiny within an EI system. However, each trooper's estimated use of race is relative to his peers, and not relative to a race-neutral trooper. These estimates are more informative when they are used in conjunction with the average use-of-race estimate completed in Chapter 3. In Chapter 2, the primary methodologies used to the estimate the relative use of race in stop decisions were discussed and critiqued, so they are not repeated here.

4.2 Data

The data for this analysis are the same data used in Chapter 3's detachment-level stop analysis for the 16 troopers, which included 18,747 stops.

4.3 Methods

This section describes the empirical method used to estimate each trooper's relative use of race as a factor in stop decisions as compared to his similarly situated peers. The estimates should take into account the degree to which the troopers are similarly situated as well as to allow for the possibility that troopers may choose to make different types of stops as compared to their peers. As a trooper's work schedule changes, his peer group of troopers will also change. Moreover, the racial-group shares of motorists at risk of being stopped during a new work shift may not be the same shares as the ones during the former work shift. Moreover, the troopers may make different types of stops that may affect the racial makeup of their stops (e.g., a trooper may focus on lane violations, which may indicate a motorist driving under the influence of alcohol).

Riley et al. (2005) developed a method that compares a particular officer's stops to other officers' stops that occurred at the same time and location; hence, the method allows for officers' schedules to change. I will build on that method and allow for the possibility that an officer and his peers differentially use non-racial motorist characteristics to influence their traffic stop decisions (e.g., an officer focuses on seatbelt violators).

The propensity score weighting method described in Chapter 3 can be adapted for this application. As in Chapter 3, assume that when an officer is observing motorists in order to make a stop, he has the capacity to stop a motorist. When he stops a motorist, the stop is completed by the motorist.[92] Let each trooper be identified by an indicator variable T_j where there are J indicator variables: T_1, T_2, T_3, … ,T_J. A trooper is either the subject trooper of interest (i.e., he is Trooper j, where $T_j = 1$) or one of the subject trooper's peers ($T_j = 0$). To illustrate, Trooper $j = 1$ is represented by the binary variable T_1. For stops completed by Trooper $j = 1$, T_1 is coded as 1, and for stops completed by his peers, T_1 is coded as 0.

In Chapter 3, the treatment assignment depended on whether the motorist was stopped by a trooper using non-radar versus radar. In this context, the treatment assignment depends on which trooper is observing a motorist in order to make a stop. If stop i is going to be completed by the subject trooper, then stop i is considered to be assigned to the treatment group ($z = 1$), and if stop i is going to be completed by the subject trooper's peers, then stop i is considered to be assigned to the control group ($z = 0$). Let M_z indicate whether a stop is completed by a minority motorist when the stop is assigned to treatment z.

For every stop that was completed by the subject trooper j ($T_j = 1$), one can imagine the counterfactual stop that would have been made had the stop been completed by the subject trooper's peers ($T_j = 0$). Therefore, depending on the treatment assignment, stop i has two potential outcomes: one outcome when $T_j =1$ (M_{1i}) and another outcome when $T_j = 0$ (M_{0i}).[93] The objective is to estimate the difference between the probability that a stop involved a minority motorist when the stop was made by the subject trooper and the probability that the stop would have involved a minority motorist had the stop been made by the subject trooper's peers. This difference is the average treatment effect on the treated stops (ATE_1), which is represented by θ_j for Trooper j in Eq. (4.1).

$$\theta_j = \Pr(M_1 \mid S = 1, T_j = 1) - \Pr(M_0 \mid S = 1, T_j = 1) \tag{4.1}$$

The null hypothesis is that the probability that a stopped motorist is minority when stopped by the subject trooper ($T_j = 1$) is less than or equal to the probability that a stopped motorist is minority when stopped by the subject trooper's peers ($T_j = 0$). Under

---

[92] Although a specific motorist's race is fixed, the motorist that completes a specific stop is not fixed. This is the reason why the theoretical discussion focuses on the stop, not the motorist.
[93] The notation is the same as the potential outcomes notation used in Chapter 3:

M_{1i} = 1 if stopped motorist i is minority if stopped by $T_j = 1$ (treatment group)
0 if stopped motorist i is non-minority if stopped by $T_j = 1$ (treatment group)

M_{0i} = 1 if stopped motorist i is minority if stopped by $T_j = 0$ (control group)
0 if stopped motorist i is non-minority if stopped by $T_j = 0$ (control group)

the null hypothesis, $\theta_j \leq 0$ for all j troopers.[94] Because θ_j is estimated for each trooper, J models are estimated. For example, Trooper 78 is the first subject trooper and the other 15 troopers serve as his peers. Then Trooper 285 becomes the subject trooper and the other 15 troopers serve as his peers, and so on.[95]

M_0 is not observed when $T_j = 1$, but can be estimated from the data. To estimate $\Pr(M_0 \mid T_j = 1)$, the subject trooper's peers' stops are weighted so their weighted joint distribution of $\boldsymbol{x}$ will closely match the subject trooper's stops' joint distribution of $\boldsymbol{x}$. This is shown in Eq. (4.2), where f is the joint distribution function of $\boldsymbol{x}$.

$$f(\boldsymbol{x} \mid T_j = 1) \approx w(\boldsymbol{x}) f(\boldsymbol{x} \mid T_j = 0) \tag{4.2}$$

The weights are estimated using the generalized boosted modeling (GBM) technique described in Chapter 3. The propensity score weighting model includes $\boldsymbol{x}$, which is composed of a vector of contextual characteristics $\boldsymbol{c}$ and a vector of motorist characteristics $\boldsymbol{m}$. Contextual characteristics are included in the model to allow for a trooper's peer group to change over time. For example, for a short period of time, a trooper may work closely with a particular group of troopers; however, if the trooper or his peers change their patrol schedule, then the trooper's peer group changes. In order to account for peer group changes, the context of each stop is included in the model. Hence, stops that occur in the same context as the subject trooper's stops are weighted more heavily as compared to stops that do not occur in the subject trooper's contact. $\boldsymbol{c}$ includes a set of binary variables indicating the nine quarters, a binary variable indicating weekday versus weekend, a set of binary variables indicating six four-hour time periods within a

[94] A one-sided hypothesis test is used so the troopers can be sorted based on a p-value.

[95] An alternative estimation strategy would be to use one logistic regression model as follows: $\ln\left(\dfrac{\Pr(M = 1 \mid S = 1, \boldsymbol{T}, \boldsymbol{x})}{1 - \Pr(M = 1 \mid S = 1, \boldsymbol{T}, \boldsymbol{x})}\right) = \delta_0 + \boldsymbol{\delta}_1 \boldsymbol{T} + \boldsymbol{\delta}_2 \boldsymbol{x}$, where $\boldsymbol{T}$ is the full set of T_j indicator variables (with one omitted). Hence, $\boldsymbol{\delta}_1$ is a vector of $J - 1$ parameters, where each parameter is the j trooper's relative use of race as compared to the omitted trooper. I did not use this strategy because I wanted the logistic regression model estimation strategy to be similar to the propensity score weighting model strategy. The algorithm used to estimate the propensity score weights requires the treatment variable to be binary, while the treatment variable in this model would be a multinomial categorical variable (i.e., each trooper would represent a different treatment level).

day, and a binary variable indicating whether the stop occurred south or north of Sea-Tac Airport. Motorists characteristics, ***m***, are also included to allow for the possibility that some motorist characteristics, which may be used by Trooper *j* differently than his peers in deciding whom to stop, may be associated with a racial group.[96] ***m*** includes a set of binary variables indicating the primary violation that initiated the stop[97] (speeding, equipment, safety belt, lane, or other), a binary variable indicating whether the motorist was driving aggressively, a binary variable indicating the motorist's gender, and a set of binary variables indicating the motorist's age (16-25, 26-35, 36-45, or 46+ years old).

Based on the weights estimated using the GBM technique, the estimate of subject trooper *j*'s ($T_j = 1$) use of race is represented by $\hat{\theta}_j$ in Eq. (4.3).

$$\hat{\theta}_j = \frac{\sum_{i=1}^{N} M_i T_{ji}}{\sum_{i=1}^{N} T_{ji}} - \frac{\sum_{i=1}^{N} M_i w(\boldsymbol{x}_i)(1 - T_{ji})}{\sum_{i=1}^{N} w(\boldsymbol{x}_i)(1 - T_{ji})} \tag{4.3}$$

To estimate the p-value of $\hat{\theta}_j$, I use the weights above in the logistic regression model defined by Eq. (4.4), where the p-value of $\hat{\theta}_{1j}$ equals the p-value of $\hat{\theta}_j$.[98]

$$\ln\left(\frac{\Pr(M = 1 \mid S = 1, T_j)}{1 - \Pr(M = 1 \mid S = 1, T_j)}\right) = \theta_{0j} + \theta_{1j} T_j \tag{4.4}$$

As a sensitivity analysis, I also estimate the relative use of race in stop decisions using the logistic regression model defined by Eq. (4.5), where $\hat{\delta}_{1j}$ is an estimate of Trooper *j*'s relative use of race.[99] As discussed above, this model has strong functional

[96] On the other hand, if troopers target particular motorist characteristics (e.g., types of violations) because the characteristics are associated with minority motorists, then including these characteristics in the model masks the use of race. The justification for including ***m*** in the model is discussed in Chapter 3.
[97] The primary violation that initiated the stop is recorded as "violation one" in the data.
[98] When weights are specified, by default, R estimates the standard errors using a robust estimator (i.e., a generalized Huber-Eicker-White sandwich estimator).
[99] So the results from the logistic regression and propensity score weighting models can be compared, the standard errors of the logistic regression model are also estimated using a robust estimator.

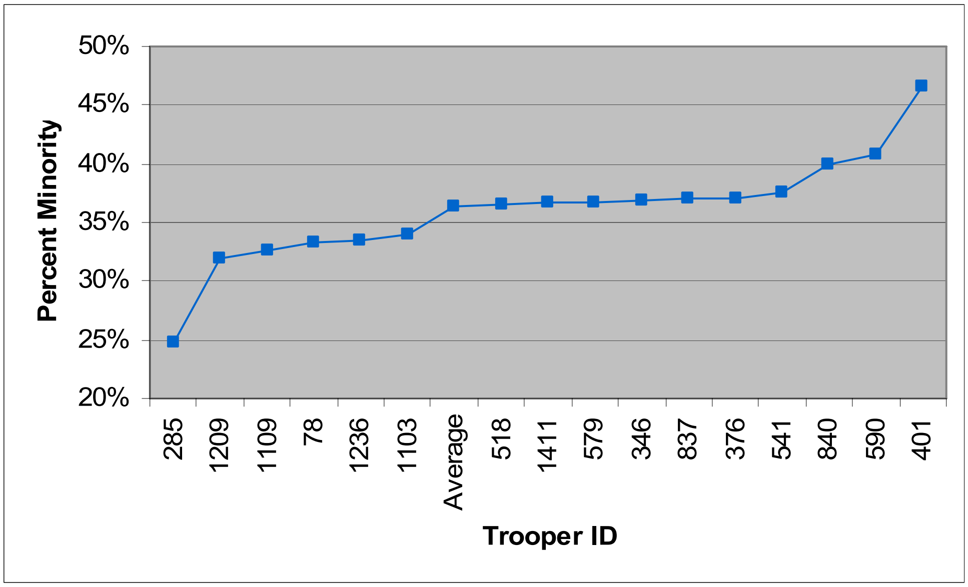
form assumptions and unlike the propensity score model, one cannot assess how well the model controlled for confounding characteristics.

$$\ln\left(\frac{\Pr(M=1 \mid S=1, T_j, \boldsymbol{x})}{1-\Pr(M=1 \mid S=1, T_j, \boldsymbol{x})}\right) = \delta_{0j} + \delta_{1j} T_j + \boldsymbol{\delta}_{2j} \boldsymbol{x} \tag{4.5}$$

4.4 Results

During the nine-quarter period evaluation period, these troopers on average each made 1,172 stops on the 22-mile stretch of Interstate 5; however, the number of stops varied widely by trooper, from a low of 670 to a high of 1,604. The wide range in the number of stops per trooper is due to troopers making different numbers of stops per shift, working different numbers of shifts, and working different percentages of their shifts on the 22-mile stretch of Interstate 5. Figure 4.1 shows the minority share of each trooper's stops. The average minority share is 36 percent, but the range varies widely from 25 to 47 percent.

Figure 4.1: Minority Share of Each Trooper's Stops[100]



Speeding violations are the most frequent violations that initiate a stop, followed by lane violations.[101] Table 4.1 shows the violation type by trooper as well as the minority share of each violation type. The troopers are sorted based on the share of their stops represented by speeding violators. As can be seen from the table, some trooper's mix of violation types differ from the average mix by a large amount. Moreover, as shown in the last row of the table, minority motorists represent 36 percent of all stops, but represent an above average share of lane violation stops and a below average share of equipment violation stops.

[100] The traffic stop dataset (and BrAC test dataset discussed in Chapters 5 and 6) are publicly available and both include the name of the trooper who made each stop or DUI arrest. While there is no formal requirement from the WSP to not reveal the troopers' identities, all RAND Corporation studies fall under an Institutional Review Board (IRB) that reviews research involving human subjects, as required by federal regulations. RAND's "Federalwide Assurance for the Protection of Human Subjects" (U.S. Department of Health and Human Services, through 2008) serves as its assurance of compliance with the regulations of 16 federal departments and agencies. According to this assurance, the IRB is responsible for review regardless of the source of funding. These federal regulations prevent RAND research from singling out specific individuals whom its research could adversely affect. Therefore, I took several steps to reduce the possibility of revealing a trooper's identity. First, I generated a random identification (ID) number to represent each trooper. Second, when tables are presented with trooper-level information that does not include results or information that support the results, I suppressed the Trooper ID since WSP personnel may be able to identify a trooper with this information. These steps significantly reduce the probability that a trooper's identity will be revealed.

[101] The primary violation that initiated the stop is recorded as "violation one" in the data.

Table 4.1: Share of Stops by the Primary Violation that Initiated the Stop by Trooper

Trooper ID (suppressed)	Speeding	Equipment	Seatbelt	Lane Violation	Other	Total
1	83%	5%	1%	2%	8%	100%
2	77%	2%	5%	11%	6%	100%
3	76%	9%	1%	4%	10%	100%
4	73%	3%	4%	17%	3%	100%
5	67%	13%	6%	10%	5%	100%
6	64%	12%	2%	8%	13%	100%
7	63%	5%	1%	18%	14%	100%
8	59%	4%	8%	12%	17%	100%
9	57%	8%	5%	18%	12%	100%
10	54%	12%	3%	14%	17%	100%
11	53%	23%	12%	6%	5%	100%
12	48%	14%	1%	35%	2%	100%
13	44%	13%	14%	13%	15%	100%
14	39%	27%	5%	14%	16%	100%
15	34%	5%	9%	47%	6%	100%
16	24%	17%	15%	22%	23%	100%
Average	55%	11%	6%	17%	11%	100%
Minority Share	34%	31%	37%	46%	36%	36%

Although these 16 troopers patrolled the same stretch of Interstate 5 between January 1, 2003 and March 30, 2005, it is important to see if they patrolled this stretch on similar dates during similar times of the day, and if not, then it is important to control for these differences. Because the troopers code each stop's location, date, and time, the stop data are very useful in determining how similarly situated these troopers were during this evaluation period. I split the 22-mile stretch of Interstate 5 into two segments: one north of Sea-Tac Airport and the other south of the airport. Table 4.2 shows that 74 percent of the stops occurred south the airport, and depending on the trooper, these stops represent between 58 and 89 percent of a trooper's stops. Minority stops represented 33 percent of stops south of the airport, slightly below the 36 percent total minority share.

Table 4.2: Share of Stops by Interstate 5 Segment by Trooper

Trooper ID (suppressed)	Share of Stops South of Sea-Tac Airport
1	89%
2	85%
3	84%
4	83%
5	81%
6	80%
7	75%
Average	74%
8	73%
9	72%
10	71%
11	70%
12	68%
13	66%
14	64%
15	63%
16	58%
Minority Share (south of Sea-Tac Airport)	33%
Minority Share (total)	36%

Table 4.3 shows that almost all troopers made stops in each quarter, and that the group of 16 troopers made more stops in the earlier quarters as compared to the later quarters.[102] Minorities represent a mostly constant share of stops, with the exception of a below average share in the 1Q03 and an above average share in the 4Q04.

[102] Due to staffing decreases, the number of stops also declined during this nine-quarter period throughout District 2, where the decrease was 19 percent. The number of stops made by the 16 troopers (including their stops off the 22-mile stretch of Interstate 5) decreased by 27 percent. However, the 16 troopers' decrease was higher for the 22-mile stretch of Interstate 5 because their share of stops on this stretch declined over the period.

Table 4.3: Share of Stops by Quarter by Trooper

Trooper ID (suppressed)	1Q03	2Q03	3Q03	4Q03	1Q04	2Q04	3Q04	4Q04	1Q05	Total
1	21%	14%	17%	10%	12%	9%	9%	5%	5%	100%
2	20%	18%	12%	12%	12%	12%	6%	8%	0%	100%
3	20%	13%	10%	11%	11%	14%	9%	7%	6%	100%
4	18%	18%	14%	9%	10%	11%	8%	8%	3%	100%
5	18%	16%	11%	9%	12%	7%	10%	11%	5%	100%
6	17%	14%	13%	9%	10%	13%	6%	8%	10%	100%
7	16%	16%	13%	10%	8%	14%	8%	8%	7%	100%
8	16%	21%	8%	10%	11%	11%	5%	9%	10%	100%
9	15%	9%	11%	10%	13%	16%	10%	6%	9%	100%
10	15%	12%	13%	11%	12%	11%	8%	11%	8%	100%
11	12%	13%	12%	9%	12%	8%	8%	15%	11%	100%
12	11%	21%	10%	10%	7%	10%	8%	13%	9%	100%
13	3%	2%	24%	17%	15%	12%	16%	9%	2%	100%
14	3%	1%	19%	14%	15%	14%	11%	10%	14%	100%
15	0%	0%	7%	23%	25%	22%	13%	5%	4%	100%
16	0%	0%	24%	19%	16%	9%	13%	4%	16%	100%
Average	14%	12%	13%	12%	12%	12%	9%	9%	7%	100%
Minority Share	33%	36%	36%	36%	37%	38%	37%	39%	36%	36%

The days of the week and times of the day that a particular trooper patrolled depends on which detachment he was assigned to. These 16 troopers primarily worked in one of four detachments, which include approximately four troopers each.[103] At any given time, there is always one detachment on patrol and sometimes two on Friday. This is because the week is split into two periods where there is an overlap on Friday: Tuesday to Friday and Friday to Monday. Each detachment is assigned to work one time shift during one period of the week. One of the detachments is considered a "ghost" detachment, which is staffed with troopers from the other three detachments when it is on duty. Due to the ghost detachment using different troopers over time, these 16 troopers not only work with the troopers in their detachment, but also work with troopers from the other two non-ghost detachments. Additionally, each detachment rotates to a new period and time shift every 28 days. Table 4.4 shows almost all of these 16 troopers patrolled Interstate 5 during different periods (both weekdays and weekends) and during different time periods, both daytime (5:00 a.m. to 8:00 p.m.) and nighttime (8:00 p.m. to 5:00 a.m.). The table also shows that some troopers worked during different periods than their peers. For example, one trooper worked exclusively during the daytime. This is important to account for since the racial-group shares of motorists at risk of being stopped may

[103] One of the detachments is considered a "ghost" detachment, which uses troopers from the other three detachments when it is on duty.

change depending on the time of the day and the day of the week. And as the last row of the table shows, the minority share of stops is significantly higher during nighttime hours.

Table 4.4: Share of Stops by Time Period of the Week by Trooper

Trooper ID (suppressed)	Daytime Weekday	Nighttime Weekday	Daytime Weekend	Nighttime Weekend	Total
1	56%	19%	18%	6%	100%
2	56%	0%	44%	0%	100%
3	55%	18%	17%	10%	100%
4	54%	18%	22%	7%	100%
5	52%	21%	23%	4%	100%
6	50%	19%	24%	7%	100%
7	48%	25%	16%	11%	100%
8	47%	27%	18%	8%	100%
9	43%	24%	26%	8%	100%
10	42%	6%	46%	6%	100%
11	37%	29%	23%	11%	100%
12	29%	38%	22%	11%	100%
13	28%	52%	3%	17%	100%
14	28%	36%	20%	17%	100%
15	27%	40%	18%	15%	100%
16	11%	61%	9%	18%	100%
Average	40%	29%	20%	10%	100%
Minority Share	30%	45%	33%	47%	36%

Table 4.5 provides an example of how well the propensity score weights performed in matching the weighted peers' distribution of stop characteristics with a subject trooper's distribution of stop characteristics. The characteristics are sorted by their relative influence in predicting whether a motorist was stopped by the subject trooper versus his peers.[104] The gray highlight indicates characteristics where the difference is greater than five percentage points or the effect size difference is greater than 0.1. As compared to his peers' stops, the subject trooper's stops consisted of a higher proportion of stops in the second quarter and north of Sea-Tac Airport, and also consisted of a lower proportion of equipment and lane violation stops. As right-side of the table

[104] The relative influence calculation is based on the non-interacted variables. The interacted variables are included in the table to evaluate how well the interacted variables matched between a subject trooper's stops and his weighted peers' stops.

illustrates, the weighted peers' stops have similar contextual and motorists characteristics as compared to the subject trooper.

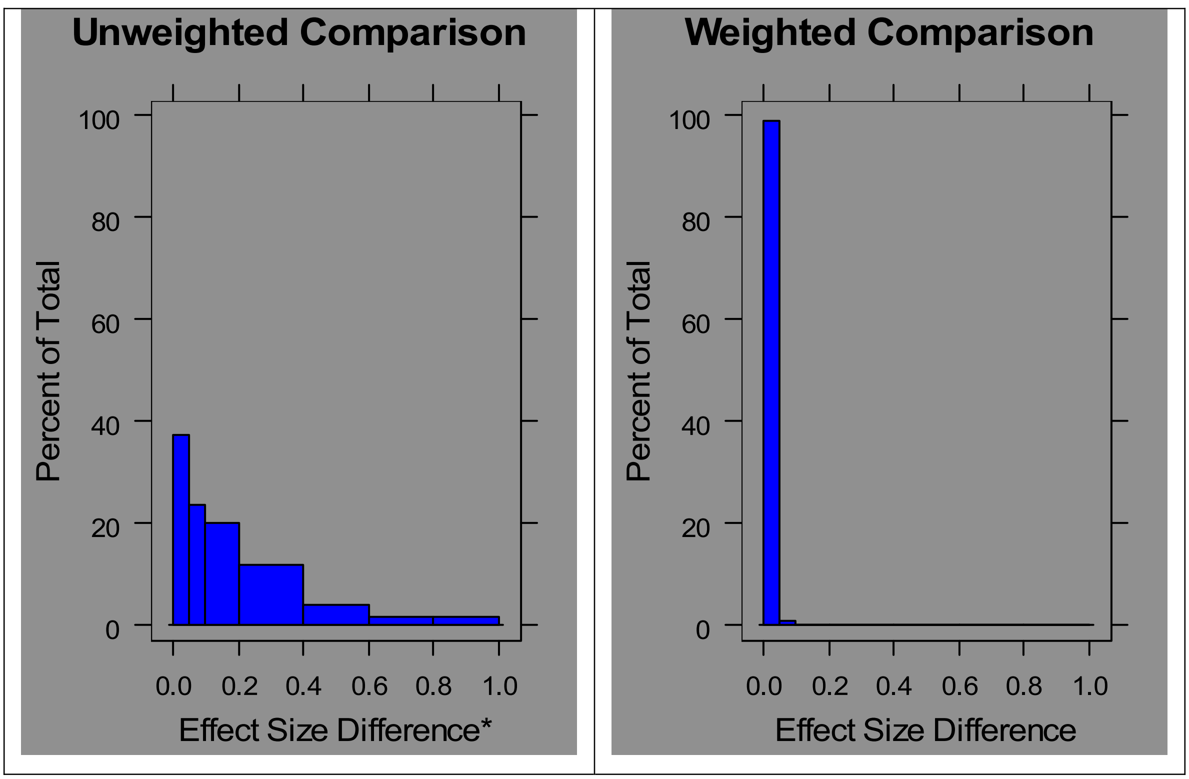
Table 4.5: Comparison between a Subject Trooper's and His Peers' Stop Characteristics (unweighted and weighted)[105]

Variable	Unweighted				Weighted		
	Subject Trooper	Peers	Difference (Subject - Peers)	Effect Size Difference	Weighted Peers	Weighted Difference (Subject - Peers)	Effect Size Difference
Quarter							
1Q03	16.0%	13.3%	2.6%	0.07	15.8%	0.2%	0.00
2Q03	21.1%	11.5%	9.5%	0.23	21.2%	-0.1%	0.00
3Q03	8.0%	13.7%	-5.7%	0.21	8.1%	-0.1%	0.00
4Q03	9.9%	11.7%	-1.8%	0.06	9.8%	0.1%	0.00
1Q04	10.5%	12.3%	-1.7%	0.06	10.4%	0.2%	0.01
2Q04	10.9%	12.1%	-1.2%	0.04	10.9%	-0.1%	0.00
3Q04	4.6%	9.3%	-4.6%	0.22	4.9%	-0.3%	0.01
4Q04	8.9%	8.8%	0.1%	0.00	8.7%	0.2%	0.01
1Q05	10.1%	7.3%	2.9%	0.09	10.1%	0.0%	0.00
Time							
12 to 4 a.m.	18.4%	22.6%	-4.3%	0.11	18.4%	-0.1%	0.00
4 to 8 a.m.	16.2%	12.0%	4.3%	0.12	15.7%	0.5%	0.01
8 a.m. to 12 p.m.	18.8%	18.3%	0.5%	0.01	19.1%	-0.4%	0.01
12 to 4 p.m.	14.6%	15.9%	-1.3%	0.04	14.6%	0.0%	0.00
4 to 8 p.m.	15.8%	14.8%	1.0%	0.03	15.9%	0.0%	0.00
8 p.m. to 12 a.m.	16.2%	16.4%	-0.2%	0.00	16.3%	-0.1%	0.00
Primary Violation Type							
Speeding	58.7%	55.3%	3.5%	0.07	59.1%	-0.3%	0.01
Equipment	4.4%	11.4%	-7.0%	0.34	4.4%	0.0%	0.00
Safety Belt	7.7%	5.6%	2.1%	0.08	7.3%	0.3%	0.01
Lane Violation	12.0%	17.3%	-5.4%	0.17	12.1%	-0.1%	0.00
Other	17.2%	10.4%	6.8%	0.18	17.1%	0.2%	0.00
Day							
Weekday	73.9%	69.0%	4.9%	0.11	74.2%	-0.3%	0.01
Weekend	26.1%	31.1%	-4.9%	0.11	25.8%	0.3%	0.01
Motorist Age							
16-25	29.5%	34.0%	-4.5%	0.10	29.6%	0.0%	0.00
26-35	27.9%	26.7%	1.2%	0.03	27.8%	0.1%	0.00
36-45	19.8%	19.1%	0.7%	0.02	19.8%	0.0%	0.00
46+	22.8%	20.1%	2.6%	0.06	22.8%	-0.1%	0.00
Milepost							
South of Sea-Tac Airport	66.4%	74.2%	-7.8%	0.16	66.4%	0.0%	0.00
North of Sea-Tac Airport	33.6%	25.8%	7.8%	0.16	33.6%	0.0%	0.00
Contact Type							
Non-Aggressive Driving	90.0%	87.5%	2.4%	0.08	89.8%	0.1%	0.00
Aggressive Driving	10.0%	12.5%	-2.4%	0.08	10.2%	-0.1%	0.00
Motorist Gender							
Male	73.7%	71.9%	1.9%	0.04	73.7%	0.0%	0.00
Female	26.3%	28.1%	-1.9%	0.04	26.3%	0.0%	0.00
Gender:Age							
Male, 16-25	20.5%	23.0%	-2.6%	0.06	20.3%	0.2%	0.00
Female, 16-25	9.1%	11.0%	-1.9%	0.07	9.3%	-0.2%	0.01
Male, 26-35	20.5%	19.6%	0.9%	0.02	20.6%	-0.1%	0.00
Female, 26-35	7.4%	7.1%	0.3%	0.01	7.2%	0.2%	0.01
Male, 36-45	15.1%	14.3%	0.8%	0.02	15.0%	0.1%	0.00
Female, 36-45	4.7%	4.8%	-0.1%	0.01	4.8%	-0.1%	0.00
Male, 46+	17.7%	14.9%	2.7%	0.07	17.8%	-0.1%	0.00
Female, 46+	5.1%	5.2%	-0.1%	0.00	5.0%	0.1%	0.00
Time:Day							
12 to 4 a.m., Weekday	13.2%	16.7%	-3.6%	0.11	13.0%	0.2%	0.00
4 to 8 a.m., Weekday	12.4%	7.9%	4.4%	0.13	11.9%	0.5%	0.02
8 a.m. to 12 p.m., Weekday	10.3%	10.5%	-0.2%	0.01	11.6%	-1.3%	0.04
12 to 4 p.m., Weekday	11.8%	10.8%	0.9%	0.03	11.2%	0.6%	0.02
4 to 8 p.m., Weekday	12.9%	10.9%	2.0%	0.06	13.0%	-0.1%	0.00
8 p.m. to 12 a.m., Weekday	13.3%	12.0%	1.3%	0.04	13.4%	-0.1%	0.00
12 to 4 a.m., Weekend	5.2%	5.9%	-0.7%	0.03	5.4%	-0.2%	0.01
4 to 8 a.m., Weekend	3.9%	4.0%	-0.2%	0.01	3.9%	0.0%	0.00
8 a.m. to 12 p.m., Weekend	8.5%	7.8%	0.7%	0.02	7.5%	1.0%	0.04
12 to 4 p.m., Weekend	2.8%	5.1%	-2.3%	0.14	3.4%	-0.6%	0.03
4 to 8 p.m., Weekend	2.9%	3.9%	-1.0%	0.06	2.8%	0.0%	0.00
8 p.m. to 12 a.m., Weekend	2.9%	4.4%	-1.5%	0.09	2.9%	0.0%	0.00

[105] The identity of the subject trooper used in this table is suppressed.

I repeat this process for the other 15 troopers whereby each trooper is the subject trooper for one model. Figure 4.2 summarizes the unweighted and weighted effect size differences for all 16 models. The histogram on the left shows the effect size differences between the characteristics of the subject trooper's stops and the unweighted characteristics his peers' stops, and the histogram on the right shows the differences between the characteristics of the subject trooper's stops and the weighted characteristics his peers' stops. Based the aggregate number of characteristics included within the 16 models, each histogram includes 756 effect size differences.[106] The histogram on the right shows that the weighting eliminates the major differences.

Figure 4.2: Effect Size Comparison between Each Trooper's and His Peers' Stop Characteristics (unweighted and weighted)



*Four of the variable levels had an unweighted effect size difference greater than 1.0, with the greatest being 1.1.

[106] The effects size comparisons between each subject trooper and his peers are based on the variable levels listed in Table 4.5. If a variable only has two levels (e.g., gender), then one of the effect size differences is dropped since both are the same. For most troopers, this results in 48 comparisons; however, if a trooper had very few or zero stops at a particular level (e.g., during one of the quarters), then those variable levels are not included in his model. This results in 12 fewer effect size comparisons.

Table 4.6 shows the results of 16 propensity score weighting models, where each trooper is represented by one model. The first set of five columns includes the primary results, followed by a set of five columns that includes additional information. The first set of columns includes the Subject Trooper's randomly generated ID, the minority share of each trooper's stops, the weighted minority share of the peers' stops, the difference between these shares (which is $\hat{\theta}_j$, see Eq. (4.3)), followed by the one-sided p-value of $\hat{\theta}_j$. In the second set of columns, the first column includes the minority share rank for each trooper, where a rank of 1 indicates the unweighted minority share of his stops is highest among the troopers. The next column is the unweighted minority share of the peers' stops, followed by the number of stops by the subject trooper, the effective number of stops by the peer troopers, and the total number of stops analyzed. Although the effective sample size varies across the models due to the different weights that are applied, for most troopers, the same 18,747 stops are analyzed in each model.[107] The troopers are sorted based on the one-sided p-value. The one-sided p-value represents the probability that the estimate of θ_j would be this or more extreme, assuming null hypothesis were true (i.e., $\theta_j \leq 0$). Because there are 16 hypothesis tests, the Bonferroni-equivalent p-value for a 0.05 significance level is 0.0031 (or 0.05/16). Based on this standard, the null hypothesis is rejected for Troopers 541 and 401.

[107] If a subject trooper had zero or few stops during particular quarters or time periods of the day, then the subject trooper's and his peers' stops during those contexts are excluded from the subject trooper's model.

Table 4.6: Estimated Use of Race by Trooper for Stop Decision

Trooper ID	Minority Share of Subject Trooper's Stops	Minority Share of Peers' Stops (weighted)	Difference	p-value (one-sided)	Minority Share Rank of Subject Trooper's Stops (unweighted)	Minority Share of Peers' Stops (unweighted)	Subject Trooper's Number of Stops	Peers' Effective Number of Stops	Total Number of Stops Analyzed
541	37.5%	31.8%	5.7%	0.0020*	4	36.3%	675	3,827	18,747
401	46.6%	42.3%	4.3%	0.0027*	1	35.6%	1,352	4,414	18,747
376	37.1%	36.0%	1.1%	0.2343	5	36.3%	1,282	7,721	17,348
346	36.8%	36.0%	0.9%	0.2552	7	36.3%	1,595	9,262	18,747
590	40.9%	39.9%	1.0%	0.2945	2	36.1%	866	3,175	18,747
579	36.7%	35.9%	0.8%	0.3108	8	36.3%	1,377	2,767	18,747
1411	36.6%	36.1%	0.5%	0.3856	9	36.3%	857	3,949	18,747
837	37.0%	36.8%	0.2%	0.4388	6	36.3%	1,214	7,251	18,747
78	33.3%	33.1%	0.2%	0.4496	12	36.6%	1,330	6,776	18,747
1109	32.7%	33.6%	-0.9%	0.7071	14	37.2%	986	3,680	13,933
518	36.6%	37.5%	-0.9%	0.7198	10	36.3%	1,289	7,484	18,747
1103	34.0%	35.4%	-1.4%	0.8142	11	36.5%	1,206	5,037	18,747
1236	33.2%	34.7%	-1.6%	0.8234	13	37.2%	911	5,090	13,933
840	40.0%	41.9%	-1.9%	0.9070	3	36.0%	1,604	4,419	18,747
1209	32.0%	35.1%	-3.1%	0.9898	15	36.8%	1,526	6,636	18,747
285	24.8%	29.2%	-4.4%	0.9902	16	31.2%	670	3,976	11,488

*Significant at the 0.0031 level (one-sided)

The logistic regression model in Eq. (4.5) is also used to estimate each trooper's relative use of race. As in the results above, the null hypothesis is rejected for Troopers 541 and 401.

As a sensitivity analysis, I estimate three additional propensity score weighting models for each trooper, which are the same as the first three sets of additional models estimated in Chapter 3. The logistic regression model in Eq. (4.5) is also used to estimate each trooper's relative use of race for the additional sets of models. The second set of models (i.e., the first set of additional models) adds two binary variables to each model. One variable indicates whether more than one violation was recorded and the other indicates whether a criminal violation was recorded. I did not include these binary variables in the original models because violations other than the primary violation that initiated the stop are mostly detected after the stop decision has been made (e.g., failure to possess vehicle liability insurance). However, the primary violation that initiated the stop may actually comprise of more than one violation, including a criminal violation.[108] If one trooper focuses on stopping motorists with multiple or criminal violation stops

[108] As discussed in Chapter 3, I exclude stops where a criminal violation is the primary violation that initiated the stop since these stops involved relatively less trooper discretion. However, a criminal violation that is recorded as a second or subsequent violation may have been detected prior to the stop, and these stops have not been dropped from the analysis.

more than his peers, then this behavior needs to be controlled for if those stops are associated with a racial group. The results from this set of models are essentially the same as the results in Table 4.6, where the null hypothesis is rejected for Troopers 541 and 401. For these troopers, the null hypothesis is also rejected using the logistic regression model (Eq. (4.5)) with the additional two binary variables.

The third and fourth sets of models estimate each trooper's relative use of race as a factor in stop decisions on different subsets of motorists, defined by the motorist's race, gender, and age. The third set of models excludes Asians/Pacific Islanders and motorists whose race is recorded as "other." Of the remaining minority motorists, 93 percent are either black or Hispanic and the rest are either American Indian or East Asian. These stops are separately analyzed because the majority of racial profiling complaints are from black and Hispanic motorists (e.g., see Meeks, 2000). The fourth set of models begins with the stops included in the third set, but only includes male motorists who are under 46 years old. These stops are separately analyzed because young males represent a higher share of racial profiling complaints as compared to their share of the minority population (Newport, 1999). When the use of race is estimated across all minorities, then the use of race to the detriment of young, male minorities may not be detected, especially if the share of young, male minorities within the sample of stopped minority motorists is small.

The results from the third and fourth sets of models are somewhat similar to the results in Table 4.6; however, there are some changes. Trooper 579's difference between his minority share of stops and the weighted non-minority share of his peers' stops increased from 0.8 percentage points in the first model to 4.9 percentage points in the fourth model. However, this difference is not statistically significant at the 0.0031 level. For Troopers 541 and 401, while the magnitude of difference between each trooper's minority share of stops and the weighted non-minority share of his peers' stops is still over three percentage points in both the third and fourth models, the differences decreased slightly from the first two models and are no longer statistically significant at the 0.0031 level.[109] The loss of statistical significance is due to the magnitude of the difference slightly decreasing as well as the sample size decreasing. For Trooper 376, while the minority share of his stops exceeded the weighted minority share of his peers' stops in Table 4.6, for the third and fourth models, the opposite is true. For each trooper, the results from the fourth model are presented in Table 4.7, which includes the same columns as Table 4.6.

Table 4.7: Estimated Use of Race by Trooper for Stop Decisions of Non-Asian, Male Motorists Under 46 Years Old[110]

Trooper ID	Minority Share of Subject Trooper's Stops	Minority Share of Peers' Stops (weighted)	Difference	p-value (one-sided)	Minority Share Rank of Subject Trooper's Stops (unweighted)	Minority Share of Peers' Stops (unweighted)	Subject Trooper's Number of Stops	Peers' Effective Number of Stops	Total Number of Stops Analyzed
579	36.2%	31.3%	4.9%	0.0064	2	32.2%	743	2,170	9,339
541	30.9%	26.4%	4.5%	0.0474	9	32.6%	324	1,604	9,339
401	42.4%	39.4%	3.0%	0.0849	1	31.7%	696	1,898	9,339
837	33.9%	32.1%	1.9%	0.1811	5	32.4%	613	3,757	9,339
346	33.5%	32.7%	0.8%	0.3246	6	32.4%	849	4,412	9,339
518	33.0%	32.5%	0.5%	0.4046	7	32.5%	703	3,890	9,339
590	35.5%	36.2%	-0.7%	0.6022	3	32.4%	448	1,292	9,339
78	28.7%	29.5%	-0.7%	0.6407	12	32.8%	588	3,471	9,339
1109	29.1%	30.5%	-1.4%	0.7217	14	34.0%	491	1,834	6,966
1411	32.3%	34.1%	-1.8%	0.7505	8	32.5%	409	1,567	9,339
1236	29.7%	31.5%	-1.8%	0.7652	13	33.9%	428	2,003	6,966
840	35.2%	36.6%	-1.4%	0.7679	4	32.3%	822	2,046	9,339
376	29.6%	31.6%	-2.0%	0.8439	11	32.7%	646	3,481	8,629
285	23.2%	25.9%	-2.7%	0.8445	16	27.7%	311	1,636	5,582
1103	30.0%	32.6%	-2.7%	0.8893	10	32.7%	574	2,102	9,339
1209	27.9%	31.1%	-3.2%	0.9497	15	32.9%	688	3,103	9,339

4.5 Discussion

The results from the first two sets of propensity score weighting models provide evidence that Troopers 541 and 401 may have used race as a factor to stop a higher share of minority motorists as compared to their similarly situated peers. When subsets of each trooper's stops are analyzed, the results also provide evidence that Trooper 579 may have used race as a factor to stop a higher share of young, male, non-Asian minorities (as compared to young, male non-minorities) as compared to his peers; however, there is a higher degree of uncertainty surrounding this result.

Based on the detachment-level stop analysis of the 16 troopers in Chapter 3, the results provided evidence that race may have been used as a factor to stop a higher share of minority motorists, but due to the uncertainty surrounding the estimates, the evidence

---

[109] However, for Trooper 401's third model, the null hypothesis is rejected using the logistic regression model in Eq. (4.5). For the third and fourth set of models using the logistic regression model, Trooper 401's third model is the only model where the null hypothesis is rejected.

[110] These models exclude Asians and motorists whose race was recorded as "other."

was not conclusive. However, based on those results, it is somewhat unlikely that this group of troopers on average used race as a factor to stop a higher share of non-minority motorists. Therefore, coupled with the results from Chapter 3, the results in this chapter not only provide evidence that the troopers identified above may have used race as a factor to stop a higher share of minority motorists as compared to their peers, but also provide evidence that they may have used race as a factor to stop a higher share of minority motorists as compared to a race-neutral trooper.

To quantify Trooper 541's and 401's uses of race, if Trooper 541 had stopped the same weighted share of minority motorists as his peers, then 38 out of 253 of his minority motorist stops would have been non-minority motorists. Using the same calculation for Trooper 401, then 58 out of 630 of his minority motorist stops would have been non-minority motorists.

Although the focus of this study is to estimate the use of race where minority motorists are adversely affected, the use of race may also adversely affected non-minority motorists.[111] The results provide evidence that Troopers 285 and 1209 may have used race as a factor to stop a higher share of non-minority motorists as compared to their peers. However, their results are not significant at Bonferroni-equivalent p-value for a 0.05 significance level is 0.9969 (or 1 – 0.05/16). Moreover, due to the results in Chapter 3, it is ambiguous whether these troopers used race as a factor to stop a higher share of non-minority motorists as compared to a race-neutral trooper. Because the results in Chapter 3 provided evidence, although it was inconclusive, that the 16 troopers may have used race as a factor to stop a higher share of minority motorists, then these two troopers may have also used race to stop a higher share of minority motorists; however, based on the results in this chapter, their uses of race to stop a higher share of minorities were the least among the 16 troopers. Therefore, it is difficult to know how these two troopers' uses of race would compare to a race-neutral trooper.

The results from the propensity score weighting model shows that the difference between a subject trooper's minority share of stops can be substantially different when

[111] Although this study did not consider the trooper's race, a number of studies do incorporate the officer's race to estimate the use of race when the officer's and motorist's races differ (e.g., see Anwar and Fang, 2006).

compared to his peers' unweighted versus weighted minority share of stops. Using a z-score or a t-test in the unweighted comparison to identify potential problem troopers would have resulted in erroneous identifications. It is important to control for differences among the troopers' patrol schedules and the type of stops they make.

This is illustrated by examining unweighted minority share rank for each trooper (see sixth column of Table 4.6). Although the subject trooper's minority share rank when compared to his unweighted peers' stops often predicts the trooper's minority share rank when compared to his weighted peers' stops, the unweighted ranks also show that the difference between the minority share of a subject trooper's stops and the unweighted minority share of his peers' stops can be substantially different from the difference between the minority share of a subject trooper's stops and the weighted minority share of his peers' stops. For example, Trooper 401 ranked 1st with the largest difference (11.0 percentage points) between the minority share of his stops and his unweighted peers' stops. When the minority share of his stops is compared to the weighted minority share of his peers' stops, he ranks 2nd with the second largest difference (4.3 percentage points); however, the disparity substantially decreases.

On the other hand, the difference for Trooper 541 substantially increases. For the unweighted comparison, the minority share of Trooper 541's stops only exceeded his peers by 1.2 percentage points; however, for the weighted comparison, the difference is 5.7 percentage points. This change could have occurred because he was patrolling when there was a lower-than-average share of minorities on the road, or he may have stopped motorists based on non-racial motorist characteristics (e.g., type of violation) that represented fewer minority violators. Based on analyzing his stops, Trooper 541 made a significantly higher proportion of his stops during the daytime as compared to his peers. During this time period, there was a lower minority share of stops as compared to the other time periods; hence, the minority share of Trooper 541's stops is expected to be lower than the average share since he was patrolling at times when the share of minorities stopped was below the average share. Additionally, speeding violations represented a significantly higher proportion of Trooper 541's stops as compared to his peers. Of the stopped motorists, non-minorities were more likely to be stopped for a speeding violation as compared to minorities. Hence, if Trooper 541 was targeting speeders, the non-minority share of his stops is expected to be higher than the average share. These two issues partially explain that while the minority share of Trooper 541's stops is just above the average minority share of his peers' stops, he stopped significantly above the average minority share given the times he patrolled and the type of violations he stopped motorists for.

The primary limitation of this chapter is whether these 16 troopers were similarly situated after controlling for the contextual and motorist characteristics that were included in the models. These 16 troopers were initially chosen as a group to be analyzed because they had patrolled the same stretch of Interstate 5 during the same time period with the same assignment. At first blush, these troopers would be considered to be similarly situated; however, the proportion of their stops that occurred during particular contexts (e.g., daytime, weekday context) varied among troopers. Moreover, the troopers made different types of stops. These differences were controlled for in the models above; however, there may be additional differences that were not captured in the data that would affect the minority share of their stops relative to their peers.

While the results provide evidence that Troopers 541, 401, and 579 may have used race as a factor to stop a higher share of minority motorists, further inquiry is required. The purpose of this analysis was to identify *potential* problem troopers, not to identify *the* problem troopers (if any exist). If the methods above were used within an EI system, the results would be a starting point to begin an inquiry, because alone, they are not conclusive (Walker, 2001). This is because there may be extenuating circumstances that explain a trooper's higher minority stop rate as compared to his peers (e.g., the trooper had a temporary assignment that differed from his peers).

WSP supervisors currently review each trooper's performance on a daily basis and already informally ask for explanations when a trooper's performance is unusual. However, reviewing the racial-group shares of a trooper's stops is not typically part of this review. The above method would allow a supervisor to review the racial-group shares of a trooper's stops over time and compare those shares to his similarly situated peers. Moreover, the above methods could also be applied to other performance indicators such as the number of stops, citations, searches, and DUI arrests per shift.

If a trooper is identified as a potential problem trooper, then the supervisor can see if there is a valid explanation. For example, assume a valid explanation was that the trooper always patrolled the highway when public events were occurring that involved a higher share of minority motorists as compared to their average share on the roadways. As explanations come to light, then when possible, they should be captured in variables that are incorporated into the trooper-level stop analysis. This is important because if those explanations are able to change the status of a potential problem trooper to a non-problem trooper, then incorporating those same explanations may have the opposite effect and change the status of a non-problem trooper to a potential problem trooper. The impact would depend on whether those explanations significantly varied among troopers and their influence on generating a higher minority share of stops. Hence, an EI system should continually evolve in order to improve its accuracy of identifying potential problem troopers.

One potential unobserved issue involves a trooper's assignment. Aggressive driving stops represented an average of 12.3 percent of 16 troopers' stops, and they represented between 5 and 14 percent of the stops for 14 of the 16 troopers. For remaining two troopers, they represented 41 and 42 percent of their stops, respectively. These two troopers may have been part of an Aggressive Driving Apprehension Team (ADAT). ADAT troopers use unmarked vehicles and focus on stopping motorists who are driving aggressively. The WSP defines aggressive driving as "The commission of two or more moving violations that is likely to endanger other persons or property, or any single intentional violation that requires a defensive reaction of another driver" (WSP Time and Activity Report Manual). If the racial-group shares of motorists who are driving aggressively differ from the racial-group shares of motorists at risk of being stopped (but who are not driving aggressively), then this would affect the racial-group shares of an ADAT trooper's stops. However, as shown in Table 3.4, minorities represented 36.4 percent of all stops and represented 38.4 percent of aggressive driving stops, which is only a two percentage point difference. Moreover, in each analysis, the model included a variable indicating whether a stop was as an aggressive driving stop. However, because ADAT troopers have a different assignment, the type of aggressive driver and the type of non-aggressive driver that they stop may be different from their non-ADAT peers who are patrolling at the same time and location.[112]

The results in this chapter will be combined with the results from the post-stop analyses in Chapters 5 and 6. When using an EI system, it is important to see if there is a pattern where a particular trooper consistently uses race more than his peers across traffic enforcement decisions. If the potential problem troopers identified above are also identified as potential problem troopers in the post-stop analysis, then this would provide additional evidence that these troopers should be further scrutinized.

[112] For purpose of not revealing trooper identities, I do not identify these two troopers; however, they are not Troopers 541 or 401, who are the only troopers who have a statistically significant estimated use of race.

Chapter 5: Detachment-Level Post-Stop Analysis

5.1 Introduction

An officer-level analysis within an EI system often involves comparing officers who are similarly situated, which includes officers who patrol the same geographical area during the same time periods with the same assignment (e.g., traffic patrol). The primary purpose of this chapter is to estimate 16 similarly situated WSP troopers' average use of race as a factor in post-stop decisions. The post-stop decisions that will be analyzed include the decision to search a stopped motorist and the decision to arrest a stopped motorist for driving under the influence (DUI) of alcohol or drugs. If the 16 troopers' average use of race in these decisions is estimated, then each trooper's relative use of race as compared to his peers can be better interpreted (and each trooper's relative use of race for post-stop decisions will be estimated in Chapter 6). In Chapter 2, the primary methodologies used to estimate the use of race in post-stop decisions were discussed and critiqued, so they are not repeated here.

5.2 Review of WSP Post-Stop Analysis

5.2.1 Search Rates (including DUI arrests)

Although an officer may issue a citation to any motorist who has violated a traffic law, other conditions must be met before he can search a motorist and his vehicle. In general, an officer may search a motorist and his vehicle if he possesses a search warrant, can establish probable cause or reasonable suspicion, or is given consent to search by the motorist. Moreover, if an officer establishes probable cause to arrest a motorist, then he may search the motorist for weapons and evidence as part of that arrest. These are known as incident-to-arrest searches.

Depending on the circumstances involved in the stop, an officer has different levels of discretion as to whether to conduct a search. If an officer has racial animosity toward a particular racial group or uses race as a statistical discriminator for criminal activity, then he might use race as a factor to search motorists from one racial group at a higher rate than similarly situated motorists from other racial groups.[113] This is especially true for relatively high-discretion searches where the search is based on an officer's degree of suspicion that the motorist is carrying contraband or is involved in some other criminal activity. On the other hand, a relatively low-discretion search is typically not based on the degree of suspicion that a motorist is carrying contraband, but is done due to an arrest or vehicle impound, regardless of the level of suspicion that a motorist is carrying contraband.

The legal basis used by the officer to conduct the search provides information about the level of discretion that the officer had in his search decision. In the WSP data, the trooper coded the legal basis he used for each search, which included one of the following six bases: incident-to-arrest, impound, search warrant, consent, canine, and frisk. An incident-to-arrest search was defined above and the other five legal bases are defined next.

An impound search is a non-evidentiary inventory search that occurs prior to a vehicle being towed, typically as the result of an accident or a DUI arrest, or because a vehicle is stranded or abandoned. The purpose of the search is not to find contraband, but is to establish a record of the personal property within the vehicle. However, if contraband is discovered, that evidence may be used against the motorist in a criminal prosecution (Loginsky, 2005). If the owner of the vehicle is present, he may refuse to permit the inventory search.

A warrant search is a search that is conducted on the basis of a search warrant. In the WSP's case, a trooper who thinks he has established probable cause to search a vehicle during a traffic stop may request a search warrant from a judge via the telephone. These requests are rare because in most cases where the trooper establishes probable cause (e.g., exigent circumstances or contraband in plain view), a search warrant is not required for him to conduct the search.

A frisk search is also known as a Terry search (see Chapter 2). If a trooper establishes reasonable suspicion that the motorist possesses a weapon that could endanger

[113] Although the term "similarly situated" has been used to describe troopers, the term can also be applied to motorists. Stopped motorists are considered similarly situated if they have the same contextual characteristics (e.g., location of stop, time of day, day of week) and motorist characteristics (e.g., number and types of violations, gender, and age).

the trooper (or establishes reasonable suspicion that a crime has or is about to occur), then a trooper may conduct a limited search, that is, frisk the motorist for weapons.

A canine search involves a trained canine sniffing for drugs or explosives. A canine search is not considered to be a search under the Fourth Amendment because of the limited loss of privacy as a result of the search, which is due to the limited nature of what a canine is trained to detect and the limited invasiveness of the search (see *United States v. Place*, 1983; *United States v. Jacobsen*, 1984; *Illinois v. Caballes*, 2005). Therefore, under federal law, a trooper does not have to establish probable cause or reasonable suspicion to conduct a canine search during a routine traffic stop as long as the canine search does not increase the duration of the stop. However, under Washington law, a canine sniff is considered to be a search (Loginsky, 2005). In the WSP's case, only specially trained troopers work with canines as part of a Serious Highway Crime Action Teams (SHCAT), and before a canine search is permitted, per WSP policy, the trooper must establish a reasonable basis for the search.

A consent search involves the trooper asking the motorist for consent to search. A trooper may ask for consent when he has not established probable cause or reasonable suspicion, but under Washington law, the trooper must advise the motorist of his right to refuse to consent to the search request (Loginsky, 2005).[114]

Even though the legal basis for the search is recorded, the degree of discretion used for a search within a particular legal basis can vary. This is especially true for an incident-to-arrest search, which is the most common legal basis justifying the search. For example, if a trooper stops a motorist and discovers that there is an outstanding warrant for his arrest involving a serious felony, the trooper has little discretion but to arrest and search the motorist for weapons and evidence. However, the chain of events that led to the arrest decision may have involved a higher degree of trooper discretion. For example, Lovrich et al. (2005) interviewed troopers who stated that if they are suspicious that a motorist is engaged in criminal activity, they may ask to motorist for consent to search, or might pointedly ask: "Where are the drugs?" Based on this questioning, some troopers stated that, surprisingly, some motorists turned over the drugs and were subsequently arrested. As with the incident-to-arrest search due to the outstanding arrest warrant, this

[114] The advisement is not required under federal law.

search is also recorded as an incident-to-arrest search; however, the events leading up to this latter search involve a greater degree of discretion. Moreover, when a trooper stops a motorist for a routine traffic violation, he has discretion over how thoroughly he scrutinizes the motorist for illegal activity such carrying contraband or driving under the influence of alcohol or drugs.

Lovrich et al. (2005) used a multinomial logistic regression model to estimate the influence of race in determining which stopped motorists were searched, including being searched as the result of a DUI arrest. Their analysis included stops that occurred between July 1, 2003 and June 30, 2004. The three outcomes included no search, high-discretion search (i.e., consent, canine, and frisk searches), and low-discretion search (i.e., incident-to-arrest, impound, and warrant searches). The covariates included motorist characteristics (race, gender, age), stop characteristics (number of violations, number of serious violations, time of the stop, whether the stopped occurred on an interstate, and the stop APA location), and trooper characteristics (gender and race). They found that stopped Native American,[115] Hispanic, and black motorists were more likely to be searched than stopped non-minority motorists for both low- and high-discretion searches, but stopped Asian/Pacific Islander and East Indian motorists were less likely to be searched than stopped non-minority motorists (see Table 5.1).

Table 5.1: Relative Odds of a Stopped Minority versus a Non-Minority Motorist Being Searched for WSP

	High-Discretion Searches		Low-Discretion Searches	
Race	Odds Ratio	p-value	Odds Ratio	p-value
Black	1.77	0.00	1.54	0.00
Hispanic	1.94	0.00	1.49	0.00
Native American	4.10	0.00	3.42	0.00
Asian/Pacific Islander	0.88	0.12	0.74	0.00
East Indian	0.49	0.00	0.44	0.00
Other	1.08	0.74	0.97	0.08

Source: based on results from Lovrich et al. (2005)[116]

[115] For this dataset, Native American and American Indian are the same racial groups.
[116] Lovrich et al. (2005) reported results for particular APAs; however, did not include APA 6. Lovrich et al. (2003) reported results for each APA. Based on their model, for APA 6, they found that black and Hispanic motorists were more likely to be searched than non-minority motorists for low-discretion searches, but black motorists were less likely to be searched than non-minority motorists for high-discretion searches. However, their evaluation period included stops that occurred between March 1 and October 31, 2002, which precedes the evaluation period for this study.

However, they concluded that although stopped Native American, Hispanic, and black motorists were more likely to be searched than stopped non-minority motorists, the data is not sufficient to conclude that the WSP troopers used race as a factor to search these motorists at a higher rate. Based on interviews with the WSP, troopers listed 29 factors that influence a decision to search, most of which were not included in the data. Some of these include the smell of alcohol or drugs, a motorist's slurred speech, or contraband in plain site. If these factors are associated with race, then the above results are biased since they do not isolate the use of race.

Lovrich et al. (2005) also calculated hit rates, which represent share of searches where contraband was found, and the results are in Table 5.2.

Table 5.2: Search Rates and Hit Rates by Race for WSP

Race	Total Stops	High-Discretion Search Rate	High-Discretion Hit Rate	Low-Discretion Search Rate	Low-Discretion Hit Rate
White (non-minority)	927,993	0.4%	22.3%	2.6%	26.6%
Black	40,033	1.0%	9.4%	5.6%	23.5%
Hispanic	75,467	0.9%	14.9%	5.6%	18.9%
Native American	6,180	1.6%	14.6%	11.8%	26.6%
Asian/Pacific Islander	36,957	0.4%	17.6%	2.3%	14.5%
East Indian*	11,674	0.2%	5.0%	1.0%	11.5%
Other*	4,225	0.4%	5.6%	2.6%	14.8%
Total	1,102,529	0.4%	19.9%	2.9%	25.0%

*Indicates 20 or fewer high-discretion searches, so the high-discretion hit rate is based on a small sample
Source: based on results from Lovrich et al. (2005)

The high-discretion hit rate is a more informative estimate of the use of race as a factor in search decisions as compared to the low-discretion hit rate. This is because a high-discretion search is based on a trooper's degree of suspicion that the motorist is carrying contraband or is involved in some other criminal activity. On the other hand, a low-discretion search is typically not based on the degree of suspicion that a motorist is carrying contraband, but is done due to an arrest or vehicle impound, regardless of the degree of suspicion that a motorist is carrying contraband. As compared to stopped non-minority motorists, the high-discretion search rates were higher for stopped black, Native American, and Hispanic motorists, but their hit rates were lower than non-minorities.

Assuming Knowles et al. (2001) model is appropriate, then the different hit rates provide evidence that these minority groups were being searched too often as compared to non-minorities in order to achieve efficient policing.[117]

5.2.2 DUI Arrest Hit Rates

As will be shown in the results below, most of the searches stem from DUI arrests. As compared to the search involved in a typical DUI arrest, the actual DUI arrest is much more burdensome on the motorist. When a motorist is arrested for a DUI, he is taken to the patrol station and asked to submit to a test, which is usually a breath test, to estimate his blood alcohol concentration (BAC). Due to Washington's implied consent law,[118] if the motorist refuses the test, his licensed is revoked for one year. The motorist is also responsible for vehicle towing costs if another driver is not available to drive his vehicle. The motorist may also bear significant legal costs if he decides to contest the charge. If the motorist is convicted of a DUI, the penalties for a first offense within seven years (for BAC below 0.15) include a \$350 to \$5,000 fine, a 24-hour jail sentence, and a 90-day driver's license suspension. The penalties for more seriously impaired motorists or repeat offenders are more severe. Similar to searches, if a trooper has racial animosity toward a particular racial group, then he might use a lower impairment threshold to arrest motorists from that racial group as compared to the impairment threshold he uses for other racial groups.

Due to the high number of fatalities and injuries from motor vehicle crashes involving alcohol and drugs, DUI enforcement is one of the WSP's "core four" mission areas within traffic enforcement. Most impaired motorists are impaired due to alcohol. In 2000, motor vehicle crashes in the U.S. resulted in 41,821 fatalities, 5.3 million injuries, and a \$231 billion economic cost (Blincoe et al., 2000). Of these crashes, alcohol was involved in 40 percent of the fatalities and 10 percent of the injuries, and accounted for 22 percent of the economic cost.[119] For the alcohol-involved crash fatalities, the

[117] See the discussion in Chapter 2 for the potential limitations of the Knowles et al. model.
[118] RCW 46.20.308: Implied Consent, Test Refusal, and Procedures.
[119] Not all crashes involving alcohol are caused by the alcohol. Blincoe et al. (2000) estimates that 81 percent of the fatalities, 70 percent of the nonfatal injuries, and 79 percent of the economic costs from crashes involving alcohol can be attributed to alcohol being the cause of the crash.

individual had a BAC at or above 0.10 percent in 79 percent of the crashes.[120] Moreover, injuries from alcohol-involved crashes were on average much more severe as compared to injuries resulting from crashes not involving alcohol.

Motor vehicle crashes involving alcohol are also costly to the state of Washington. During 2003 and 2004, Washington averaged 582 motor vehicle-related fatalities per year, of which, 43 percent involved alcohol and 89 percent of these crashes involved an individual with a BAC at or above 0.08 percent (NHTSA, 2003 and 2004). For 1996, a study estimated that alcohol-involved crashes in Washington injured 25,700 persons and cost $3.4 billion (Pacific Institute for Research and Evaluation).

Drugs also play a role in motor vehicle crashes, although to a lesser extent than alcohol. Shinar and Walsh (2003) summarize the literature that studied motorist vehicle crashes due to drug-impaired drivers. Numerous studies tested fatally injured drivers for drugs. Cannabis was the most common drug found, with a prevalence rate ranging from 7 to 37 percent with a mean of 14 percent. The prevalence rates for five other drug categories (narcotics, benzodiazepines, barbiturates, cocaine, and amphetamines) were five percent or lower. In a study that tested for drugs in 318 fatally injured drivers in the state of Washington between September 1992 and August 1993, Logan and Schwilke (1996) detected drugs in 25 percent of the drivers. The most common illicit drug prevalence rates were marijuana (11%), cocaine (3%), and amphetamines (2%), and depressant prescription medications had a prevalence rate of nine percent.

Due to DUI enforcement being one of the core four mission areas within traffic enforcement, troopers are trained to identify impaired motorists. A trooper uses various indicators to determine whether a motorist is driving under the influence. Some of the indicators include the motorist's driving behavior, alcohol or drugs in plain view, the smell of alcohol or drugs, and the motorist's demeanor and speech patterns. If a trooper has reasonable grounds (akin to reasonable suspicion in a Terry search) that the motorist is under the influence of alcohol or drugs, he may ask the motorist to take field sobriety tests and a portable breath test; however, the motorist may refuse to take these tests.[121]

---

[120] The fatalities primarily include motor vehicle drivers, passengers, and motorcycle riders, but also include pedestrians and pedalcyclists.
[121] The primary field sobriety tests include the horizontal gaze nystagmus, walk-and-turn, and one-leg stand.

Based on the trooper's observation and these tests, the trooper may arrest the motorist for a DUI if he has probable cause that the motorist is impaired (or has a BAC at or above 0.08 percent due to Washington's *per se* impaired law).[122]

For each of the trooper's decisions concerning a potential impaired motorist, the trooper has discretion. During a stop, a trooper has discretion over how thoroughly he scrutinizes the motorist for DUI indicators, and more importantly, has some discretion on the degree of suspicion he uses to decide whether to ask the motorist to take the field sobriety tests, and ultimately, whether he makes the DUI arrest. For a weakly to moderately impaired motorist, a trooper has a greater degree of discretion for these decisions as compared to the typical low-discretion incident-to-arrest search.

Lovrich et al. (2003) examined statewide breath alcohol concentration (BrAC) test data, which included the race of the motorist and whether he tested at or above 0.08 grams per 210 liters of breath (g/210L), which is the legal BrAC limit.[123] Using Knowles et al.'s (2001) reasoning to compare search hit rates among racial groups in order to identify racial animosity, Lovrich et al. (2003) compared the hit rates of DUI arrestees, where the hit rate is the share of DUI arrestees who tested at or above 0.08 g/210L (or 0.08 for short). If one racial group had a lower hit rate than the hit rates of the other racial groups, then this difference may indicate that a lower impairment threshold was used to arrest motorists from that racial group as compared to the threshold used for other racial groups. The statewide BAC data included tests from the WSP and Washington local law enforcement agencies. Based on 13,414 tests conducted by the WSP, of which 12.2 percent were minority motorists, 63.2 percent of minorities tested at or above 0.08 while 67.7 percent of non-minorities tested at or above 0.08, a difference of 4.5 percentage points, with a one-sided p-value of 0.0001 (see Table 5.3). In order for the minority and non-minority hit rates to have been the same, then troopers would have had to arrest 111 fewer minority motorists who tested below 0.08.

---

[122] Similar to other states, Washington has a *per se* law, where it is illegal to drive with a BAC at or above 0.08 percent, regardless of the level of impairment (RCW 46.61.502: Driving Under the Influence).

[123] Assuming the BrAC is 2100 times higher than the BAC, then a BrAC of 0.08 g/210L is equivalent to a BAC of 0.08 percent. For commercial vehicle operators, the legal limit is 0.04 percent. For motorists under 21 years old, the legal limit is 0.02 percent, although they are charged with a violation less serious than a DUI if they test below 0.08 percent. If an arrestee tests under the legal limit, he may still be charged with a DUI due to being sufficiently impaired at his measured alcohol concentration level (even if his impairment is only due to alcohol).

Table 5.3: WSP DUI BrAC Test Results by Race

WSP	Minority	Non-Minority	All
BrAC < 0.08	601	3,801	4,402
BrAC ≥ 0.08	1,030	7,982	9,012
Total	1,631	11,783	13,414
Proportion with BrAC ≥ 0.08	63.2%	67.7%	67.2%
Additional Minorities	111		

Source: based on results in Lovrich et al. (2003)

An even larger difference was found in other law enforcement agencies in the state (see Table 5.4). Based on 23,078 tests, of which 12.9 percent were minority motorists, 56.3 percent of minorities tested at or above 0.08 while 64.8 percent of non-minorities tested at or above 0.08, a difference of 8.5 percentage points, with a one-sided p-value 0.0000. In order for the minority and non-minority hit rates to have been the same, then officers would have had to arrest 448 fewer minority motorists who tested below 0.08.

Table 5.4: Washington Law Enforcement Agencies' (except WSP) DUI BrAC Test Results by Race

Washington Law Enforcement Agencies (except WSP)	Minority	Non-Minority	All
BrAC < 0.08	1,496	8,125	9,621
BrAC ≥ 0.08	1,929	14,953	16,882
Total	3,425	23,078	26,503
Proportion with BrAC ≥ 0.08	56.3%	64.8%	63.7%
Additional Minorities	448		

Source: based on results in Lovrich et al. (2003)

If the assumptions of Knowles's et al. (2001) model hold, which are discussed in Chapter 2, then the hit rate differences provide evidence that the WSP and Washington law enforcement agencies used a lower impairment threshold to arrest minority motorists as compared to the threshold they used to arrest non-minority motorists. A key result of Knowles's model is that the average hit rate equals the marginal hit rate, which is the rate that needs to be compared to determine whether the impairment threshold used to arrest motorists for a DUI differed among racial groups. However, the primary criticism of

Knowles's search hit rate model is even more acute when it is applied to DUI arrest hit rates.[124] For DUI arrests, the primary criticism would be that motorist characteristics $\boldsymbol{x}$ depend on whether a motorist is impaired, that is, $\boldsymbol{x}$ is endogenous to the probability that the motorist is impaired (see Anwar and Fang, 2006). This is likely to be the case. An impaired motorist will emit a higher level of suspicion of being impaired as compared to the same motorist type ($\boldsymbol{x}$, R) who is not impaired. Based on numerous studies, the impaired motorist will drive worse and perform worse on field sobriety tests (e.g., see Moskowitz and Fiorentino, 2000; Moskowitz et al., 2000; Stuster, 2001). Therefore, Knowles model may not be true when analyzing DUI arrest hit rates; therefore, a racial group's average hit rate may not equal its marginal hit rate.

Since the results above are based on each racial group's average hit rate, the evidence is less clear whether the WSP and Washington law enforcement agencies used a lower impairment threshold to arrest minority motorists as compared to the threshold they used to arrest non-minority motorists. For example, assume that a trooper uses the same impairment threshold among all racial groups (i.e., he arrests all motorists who are above a particular impairment level or arrests all motorists who exceed a particular probability of being impaired (i.e., a BAC at or above 0.08)). If the arrested motorists from group A were on average more suspicious (i.e., had higher BACs) than the arrested motorists from group B, then group A's expected average hit rate is higher than group B's expected average hit rate. Hence, in this situation, comparing each group's average hit rate is an inaccurate method to determine whether DUI arrest decision impairment thresholds differed among racial groups.

5.3 Data

For the search rate analysis, which includes DUI arrests, the data is from the same traffic stop dataset that was used in the stop analyses in Chapters 3 and 4. However, for this analysis, I use a different subset of their stops, which will be described next. For the DUI arrest hit rate analysis, I obtained a separate dataset, the BrAC test dataset, which includes each DUI arrestee's BrAC test results. That dataset will be described below.

[124] Note that Knowles et al. (2001) only applied their model to searches, not DUI arrests.

5.3.1 Data for Search Rate Analysis (including DUI arrests)

As discussed above, it is important to distinguish between low- and high-discretion searches. The Lovrich reports considered search warrant, incident-to-arrest, and impound searches to be low-discretion searches, and frisk, canine, and consent searches to be high-discretion searches. I begin with their categories and then make two changes. First, I create a new category for DUI arrests, and second, switch search warrant searches to the high-discretion category.

DUI arrests make up the majority of the incident-to-arrest searches, which make up the majority of all searches. A DUI arrest occurs for a specific violation: a motorist being impaired. Other arrests occur for various violations. Due to the high number of DUI arrests and the uniqueness of the violation, I separate DUI arrests from the other incident-to-arrest searches. And due the unique discretion involved in DUI arrests, I consider them as a separate discretion level in between a high- and low-discretion search.

When the WSP began collecting traffic stop data in 2001, the search variable was often miscoded. Lovrich et al. (2005), which analyzed search data from stops that occurred between July 1, 2003 and June 30, 2004, indicated the quality of the coding had improved, but some issues remained. The primary issue was that approximately one-third of DUI arrests were coded as no search. Per WSP policy, all DUI arrests involve an incident-to-arrest search and should be coded accordingly. For this study's nine-quarter period (January 1, 2003 to March 31, 2005), the coding improved, but a search was still not coded for 22 percent of these 16 troopers' DUI arrests. Due to this coding issue, I considered a motorist to have been arrested for a DUI if a DUI violation was recorded, even if no search was recorded.

Second, I considered search warrant searches to be high-discretion searches. There are two types of warrants, arrest and search warrants. Arrest warrants exist before the stop is made. When a trooper makes a stop, he typically checks to see if there is an outstanding arrest warrant for the motorist. If an arrest warrant exists, the arrest warrant is coded as a violation in the traffic stop data and the resulting search is coded as an incident-to-arrest search. On the other hand, as stated above, a search warrant is requested from a judge during a stop when a trooper thinks he has established probable cause to search a motorist and/or his vehicle, but requires a warrant to conduct the search. Hence, a trooper has more discretion whether to request a search warrant as compared to low-discretion searches such as an incident-to-arrest or impound search. Based on the discussion above, my search discretion levels include the following legal bases.

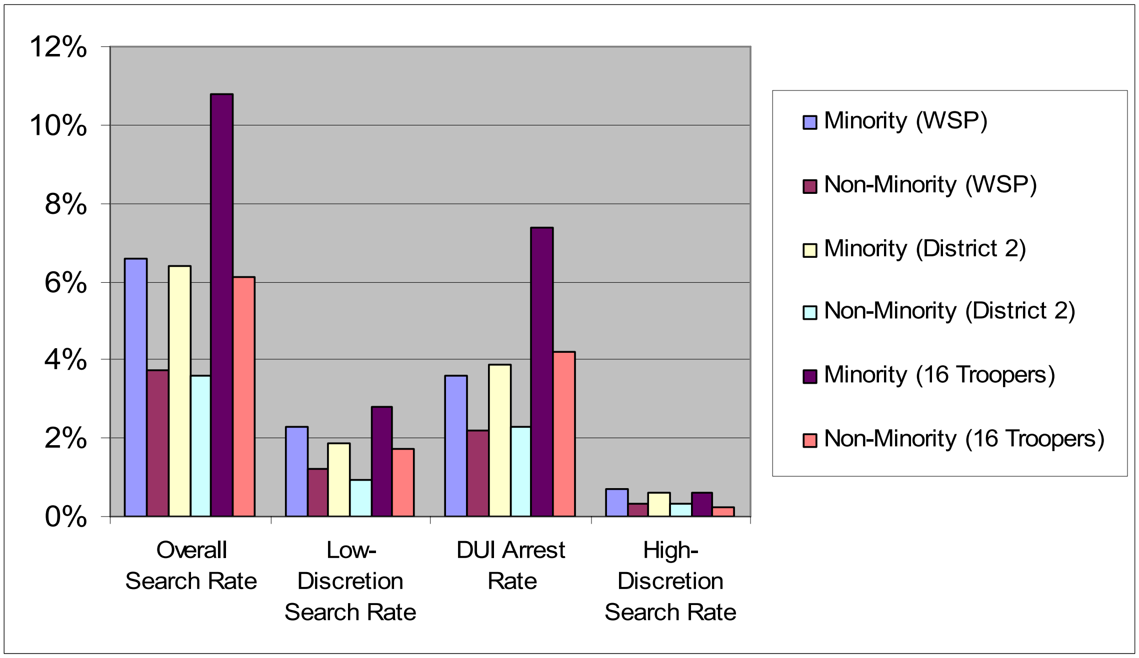
- Low Discretion (no DUI arrests)
  - Incident to arrest
  - Impound

- DUI Arrests

- High Discretion (no DUI arrests)
  - Search warrant
  - Frisk
  - Canine
  - Consent

Based on these categories, Figure 5.1 shows the WSP, District 2, and the 16 troopers' overall search rate, low-discretion search rate, DUI arrest rate, and high-discretion search rate for stopped minority and non-minority motorists during the nine-quarter period. The first set of bars, which include all searches, is the sum total of the other sets of bars, which are the search rates of each discretion level. For the WSP, District 2, and the 16 troopers, most of the searches stemmed from DUI arrests, and there were very few high-discretion searches. For each search discretion level, the search rate of stopped minority motorists was higher than the search rate of stopped non-minority motorists. As compared to the WSP and District 2, the 16 troopers had a substantially higher DUI arrest rate.

Figure 5.1: Search Rates by Search Discretion Level and Racial Group for WSP, District 2, and 16 Troopers



For the 16 troopers' searches, Table 5.5 shows the number of searches that occurred for each legal basis. Approximately 60 percent of the low-discretion searches were based on an incident to an arrest, and almost all of the high-discretion searches were frisk searches.

Table 5.5: 16-Trooper Search Count and Rate by Search Type and Racial Group[125]

Legal Basis for Search	Minority	Non-Minority	Total
Search Counts			
No Search	6,977	12,835	19,812
Low-Discretion Searches			
Incident to Arrest	197	199	396
Impound	22	36	58
DUI Arrests	577	573	1,150
High-Discretion Searches			
Frisk	44	28	72
Consent	2	2	4
Total	7,819	13,673	21,492
Search Rates			
Low Discretion	2.8%	1.7%	2.1%
DUI Arrest	7.4%	4.2%	5.4%
High Discretion	0.6%	0.2%	0.4%
Total	10.8%	6.1%	7.8%

As indicated above, a stop may have up to eight violations. To remove stops where a trooper's discretion to search is likely to be very low, I exclude stops when road rage or a felony is indicated, with the exception of a drug felony due to the higher degree of discretion involved in a drug search as compared to most felonies (e.g., hit and run). I also exclude trooper-reactive stops since most of these stops involve troopers responding to accidents, and to a lesser extent, involve calls for service and physical assists. I exclude these stops because the data do not include information about the situation. For example, the data do not include information about the seriousness of an accident and whether the motorist was at fault, which could be key predictors of whether a motorist was impaired and ultimately arrested for a DUI. Additionally, I exclude emphasis patrol contacts because these contacts often involve special assignments (e.g., DUI enforcement) around holidays. I also remove impound searches since they are non-evidentiary searches. This results in the following search counts that occurred for each legal basis (see Table 5.6).

125 In Chapter 3, 21,509 stops were analyzed, but the number analyzed in this chapter decreases to 21,492 because 17 stops did not indicate whether a search occurred.

Table 5.6: 16-Trooper Search Count and Rate by Search Type and Racial Group used in Search Rate Analysis

Legal Basis for Search	Minority	Non-Minority	Total
Search Counts			
No Search	6,231	11,437	17,668
Low-Discretion Searches			
Incident to Arrest	153	163	316
DUI Arrests	497	467	964
High-Discretion Searches			
Frisk	43	26	69
Consent	2	1	3
Total	6,926	12,094	19,020
Search Rates			
Low Discretion	2.2%	1.3%	1.7%
DUI Arrest	7.2%	3.9%	5.1%
High Discretion	0.6%	0.2%	0.4%
Total	10.0%	5.4%	7.1%

5.3.2 BrAC Data for DUI Arrest Hit Rate Analysis

DUI arrests account for the majority of searches. To analyze the arrests, I obtained the BrAC test dataset that includes each DUI arrestee's BrAC results. When a motorist is arrested for a DUI, he is typically administered two breath tests in order to estimate his BAC, and my data includes the results from the first test.[126] The data also include the arrestee's gender, whether the arrestee was a commercial vehicle driver or under 21 years old, whether the DUI arrest was the primary reason for the arrest, and the arresting trooper's name.

The BrAC test dataset included 1,820 arrests for these 16 troopers during the nine-quarter period.[127] In order to be able to compare motorists based on the 0.08 legal standard, I exclude commercial vehicle drivers and motorists under 21 years old. In order to compare motorists who were arrested primarily due to being under the influence of alcohol or drugs, I also exclude motorists where the DUI arrest primarily occurred due to

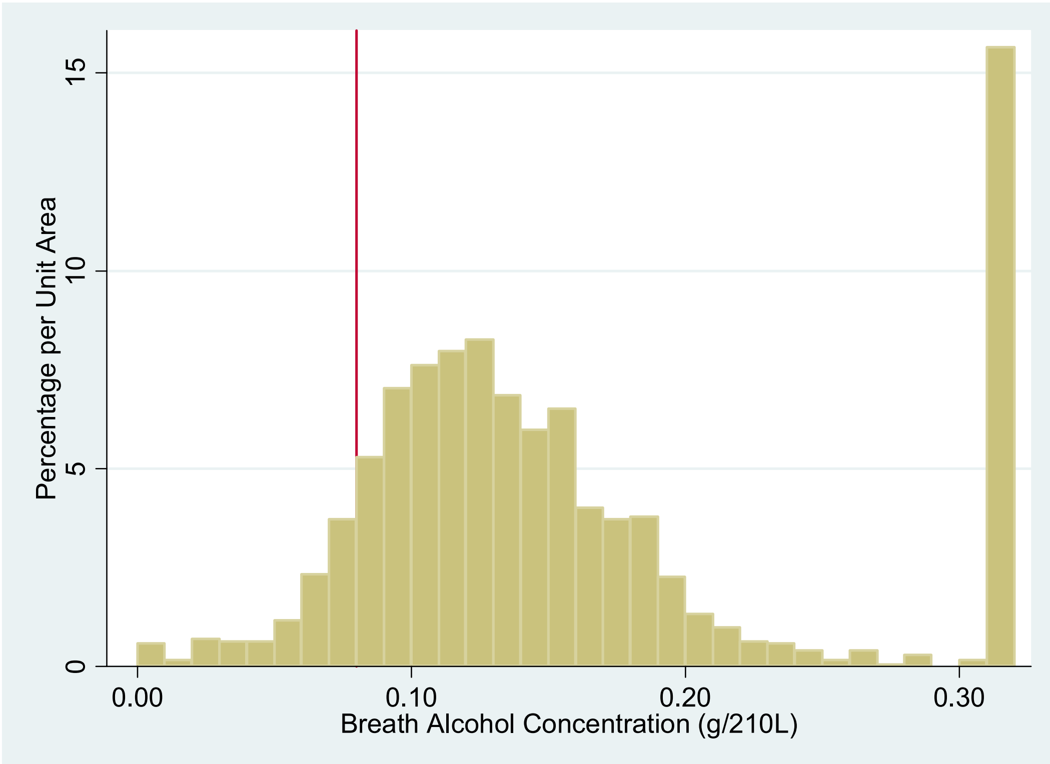
[126] The BAC is sometimes estimated from a blood or urine sample.
[127] Some of these arrests presumably occurred off of Interstate 5; however, the data do not include information about the specific location of the arrest.

another serious traffic offense (e.g., reckless or negligent driving). This results in 1,719 arrests.[128]

Figure 5.2 shows BrAC levels for the 1,719 DUI arrestees. The bar at the far right includes motorists who refused to take the breath alcohol test. The thin red line indicates Washington's legal limit of 0.08, and 90.5 percent of the arrestees tested at or above 0.08, including those who refused to test. I assume the arrestees who refused the test had a BrAC at or above 0.08 since if they believed they had a BrAC below 0.08, then a test confirming that would reduce the likelihood that they would be found guilty of a DUI.

[128] In addition, I exclude the 16 DUI arrests where the motorist was not driving the vehicle, but had physical control of the vehicle (e.g., a motorist in a vehicle parked on the shoulder of the road with the keys in the ignition). The traffic stop dataset indicated that the 16 troopers arrested 2,129 motorists for a DUI during the nine-quarter period (including stops off the 22-mile stretch of Interstate 5), and as stated above, the BrAC test dataset included 1,820 (85 percent) of the arrestees. The difference is partially due to some arrestees having their blood alcohol concentration estimated from a blood sample instead of a breath sample. For example, if a motorist is arrested for a DUI and was involved in a vehicular homicide or assault, then the motorist's blood alcohol concentration is estimated using a blood sample instead of breath sample. In addition, the difference may be partially due undercounting the number of some troopers' DUI arrests because the arresting trooper's name is sometimes ambiguous within the BrAC test dataset (e.g., DUI arrests were not included when a name only included the trooper's last name and there were at least two troopers who had that last name).

Figure 5.2: Distribution of Breath Alcohol Concentration Levels[129]



5.4 Methods

This section describes the empirical method used to estimate the use of race as a factor in deciding which stopped motorists are searched or arrested for a DUI. The first part of this section compares search and DUI arrest rates among racial groups. Due to the high number of searches stemming from DUI arrests, the second part of this section compares the shares of DUI arrestees within each racial group who tested above the legal alcohol concentration limit.

5.4.1 Search Rates (including DUI arrests)

The comparison of search rates among racial groups should take into account the possibility that there may be contextual and motorist characteristics that are both associated with a racial group and being searched or arrested for a DUI. (For brevity, I develop the method using searches, but DUI arrests are also implied.) In the stop analysis, the racial-group shares of motorists at risk of being stopped had to be estimated;

[129] In the second bar furthest from the right, motorists with a BrAC over 0.30 are binned together.

however, for the post-stop analysis, the racial-group shares of motorists at risk of being searched are known since the at risk group includes all stopped motorists, whose race is reported in the data. The primary method used to estimate the use of race as a factor in search decisions is based on Ridgeway (in press), which uses propensity score weights to estimate average treatment effects. The method is described in Chapter 3, and its application to post-stop decisions is described next.

In Chapter 3, the treatment assignment depended on whether a motorist's race was visible to the trooper prior to the stop. In this context, the treatment assignment depends on whether the stopped motorist is minority. Let $\dot{S}_z$ indicate whether a stopped motorist is searched when he is assigned to treatment z.[130] For every search decision involving a stopped minority motorist (z = 1), one can imagine the counterfactual search decision where the motorist is the exact same motorist, but is non-minority (z = 0). Therefore, depending on the treatment assignment, search decision i has two potential outcomes: an outcome when z =1 ($\dot{S}_{1i}$) and another outcome when z = 0 ($\dot{S}_{0i}$). The objective is to estimate the difference between the probability that a stopped minority motorist was searched and the probability of that minority motorist being searched had he been non-minority. This is the average treatment effect on treated stops (ATE_1), which is represented by θ in Eq. (5.1).[131]

$$\theta = \Pr(\dot{S}_1 \mid M = 1) - \Pr(\dot{S}_0 \mid M = 1) \qquad (5.1)$$

The null hypothesis is that the probability that a stopped minority motorist is searched is less than or equal to the probability that a stopped non-minority motorist is searched under the same circumstances (i.e., $\theta \leq 0$).

---

[130] The *S*-dot $(\dot{S})$ variable is used to distinguish the search variable from the stop variable (*S*).

[131] The outcomes $\dot{S}_1$ and $\dot{S}_0$ are respectively analogous to the outcomes S_1 and S_0 defined in Chapter 3:

$\dot{S}_{1i}$ = 1 if motorist *i* is searched or arrested for a DUI if minority (treatment group)
0 if motorist *i* is neither searched nor arrested for a DUI if minority (treatment group)

$\dot{S}_{0i}$ = 1 if motorist *i* is searched or arrested for a DUI if non-minority (control group)
0 if motorist *i* is neither searched nor arrested for a DUI if non-minority (control group)

$\dot{S}_0$ is not observed when $M = 1$, but can be estimated from the data. To estimate $\Pr(\dot{S}_0 \mid M = 1)$, the non-minority stops are weighted so their weighted joint distribution of stop characteristics $\boldsymbol{x}$ will closely match the minority stops' joint distribution of $\boldsymbol{x}$. This is shown in Eq. (5.2), where f is the joint distribution function of $\boldsymbol{x}$.

$$f(\boldsymbol{x} \mid M = 1) \approx w(\boldsymbol{x}) f(\boldsymbol{x} \mid M = 0) \tag{5.2}$$

The weights are estimated using the generalized boosted modeling (GBM) technique described in Chapter 3. The propensity score weighting model includes $\boldsymbol{x}$, which is composed of a vector of contextual characteristics $\boldsymbol{c}$, a vector of motorist characteristics $\boldsymbol{m}$, and a vector of binary variables representing individual troopers. $\boldsymbol{c}$ includes a set of binary variables indicating the nine quarters, a binary variable indicating weekday versus weekend, a set of binary variables indicating six four-hour time periods within a day, and a binary variable indicating whether the stop occurred south or north of Sea-Tac Airport. $\boldsymbol{m}$ includes a set of binary variables indicating the primary violation that initiated the stop[132] (speeding, equipment, safety belt, lane, criminal, or other); a binary variable indicating whether the motorist was driving aggressively; a set of binary variables indicating whether the motorist had one, two, or three or more violations; a set of binary variables indicating whether the motorist had zero, one, or two or more criminal more violations; a binary variable indicating the motorist's gender; and a set of binary variables indicating the motorist's age (16-25, 26-35, 36-45, or 46+ years old). A set of binary variables indicating the trooper who made the stop are also included because a trooper may be associated with minority stops and searching motorists. For example, if a trooper searches all motorists at a higher rate than his peers and minorities represent a higher share of his stops as compared to his peers' stops, then an indicator variable for the trooper should be included in the model, otherwise the estimate of θ would be biased.[133] As a sensitivity analysis, I will estimate a model without the trooper indicator

[132] The primary violation that initiated the stop is recorded as "violation one" in the data.
[133] With respect to the propensity score model, the trooper indicator variables should be included in the model because they influence the treatment assignment, that is, whether a stopped motorist is minority or non-minority.

variables since a trooper with a relatively high search rate might use race as a factor in deciding which motorists to stop.

Because a trooper has discretion over whether and how to record violations, then if he has racial animosity toward minorities, he may record a higher number and more serious violations for minority motorists than non-minority motorists who have the same violation(s). If this occurs when these variables are included in the model, then the use-of-race estimate will be understated. However, if minority motorists have more violations and more serious violations than non-minorities, then since these variables are positively associated with searches, then the use-of-race estimate will be overstated if they are not included in the model. I chose to include them in the model to increase the confidence in the results. Although their inclusion reduces the model's ability to detect the use of race, their inclusion reduces the probability that the use-of-race estimate will be significant when it is actually not significant. For any set of variables that are included in $\boldsymbol{x}$, the propensity score weights can be evaluated to see how well they controlled these potentially confounding characteristics. In the results section, a table is presented that compares the minority motorists' joint distribution of $\boldsymbol{x}$ with non-minority motorists' weighted joint distribution of $\boldsymbol{x}$.

Based on the weights estimated using the GBM technique, the use of race as a factor in deciding which stopped motorists are searched is estimated using Eq. (5.3).

$$\hat{\theta} = \frac{\sum_{i=1}^{N} \dot{S}_i M_i}{\sum_{i=1}^{N} M_i} - \frac{\sum_{i=1}^{N} \dot{S}_i w(\boldsymbol{x}_i)(1 - M_i)}{\sum_{i=1}^{N} w(\boldsymbol{x}_i)(1 - M_i)} \tag{5.3}$$

To estimate the p-value of $\hat{\theta}$, I use the weights above in the following logistic regression model (Eq. (5.4)), where the p-value of $\hat{\theta}_1$ equals the p-value of $\hat{\theta}$.[134]

$$\ln\left(\frac{\Pr(\dot{S} = 1 \mid M)}{1 - \Pr(\dot{S} = 1 \mid M)} \right) = \theta_0 + \theta_1 M \tag{5.4}$$

[134] When weights are specified, by default, R estimates the standard errors using a robust estimator (i.e., a generalized Huber-Eicker-White sandwich estimator).

As a sensitivity analysis, I also estimate the use of race in post-stop decisions using the following logistic regression model (Eq. (5.5)), where $\hat{\delta}_1$ is an estimate of the use of race.[135] As discussed Chapter 3, this model has strong functional form assumptions and unlike the propensity score model, one cannot assess how well the model controlled for confounding characteristics.

$$\ln\left(\frac{\Pr(\dot{S}=1 \mid M, \boldsymbol{x})}{1-\Pr(\dot{S}=1 \mid M, \boldsymbol{x})}\right) = \delta_0 + \delta_1 M + \boldsymbol{\delta}_2 \boldsymbol{x} \tag{5.5}$$

5.4.2 DUI Arrest Hit Rate Analysis

As discussed in Chapter 2, the three potential sources of racial disparities include racial animosity, statistical discrimination, and confounding characteristics. The search rate method above estimates the combined effect of racial animosity and statistical discrimination by attempting to control for confounding characteristics. In order to mitigate the effect of unobserved confounding characteristics as well as to isolate the effect of racial animosity on DUI arrests, I will adapt Knowles's et al. (2001) model that was developed to estimate racial animosity in search decisions. Knowles's model compares each racial group's search hit rate (i.e., the share of searches that result in contraband being found), and in a similar vein, my model compares each racial group's DUI arrest hit rate (i.e., the share of DUI arrestees who test over the legal alcohol concentration limit). My model will extend the DUI arrest hit rate analysis discussed above from Lovrich et al. (2003). As discussed in Chapter 2, if the hit rate is lower for the marginal motorists searched from one racial group as compared to the marginal motorists searched from another racial group, then this is evidence of racial animosity toward to the former racial group. Both the search and DUI arrest hit rate models detect racial animosity using marginal motorists (i.e., the least suspicious motorist searched and the least impaired motorist arrested for a DUI, respectively), but due to data limitations, the Knowles model must rely on average hit rates. However, the DUI data include each arrestee's breath alcohol concentration, which can be used to help identify the marginal DUI arrestees. My model begins with the following definition.

Let $IMP(R, \boldsymbol{x})$ = 1 if a stopped motorist of race R and characteristics $\boldsymbol{x}$ is impaired as defined by the state of Washington's law

= 0 if a stopped motorist of race R and characteristics $\boldsymbol{x}$ is not impaired as defined by the state of Washington's law

A trooper is unable to perfectly measure impairment, but instead estimates whether a motorist is impaired based on the motorist's characteristics $\boldsymbol{x}$ that he observes. The primary characteristics within $\boldsymbol{x}$ include the motorist's driving behavior, field sobriety test results, demeanor, and speech patterns as well as the presence of alcohol or drugs detected by sight or smell. $\boldsymbol{x}$ might also include a motorist's sex, age, height, and weight. In addition to $\boldsymbol{x}$, a trooper might use the motorist's race to estimate impairment. This use of race only includes the use of race due to statistical discrimination, not racial animosity, which will be included below.

Due to the availability of relatively direct evidence within $\boldsymbol{x}$ to establish probable cause for a DUI arrest, the use of other variables (including race) as statistical discriminators to predict impairment would likely be less effective as compared to the use of statistical discriminators to predict whether a motorist is carrying contraband. Prior to arresting a motorist for a DUI, a substantial amount of relatively direct evidence for impairment can be collected by observing the motorist's driving behavior, field sobriety test results, demeanor, and speech patterns, and detecting alcohol or drugs by sight or smell. Due to the low cost of obtaining this evidence that is a very accurate predictor of impairment, other variables (including race) are not likely to be associated with another unobserved variable that significantly improves the prediction of the motorist's impairment level. On the other hand, in a high-discretion search situation where there is relatively little direct evidence available to establish probable cause or reasonable suspicion, then indirect evidence such as the crime rate in the neighborhood, the day of the week, the time of day, and the motorist's sex, age, and race might be more effective at

[135] So the results from the logistic regression and propensity score weighting models can be compared, the standard errors of the logistic regression model are also estimated using a robust estimator.

improving the prediction of whether a motorist is carrying contraband as compared improving the prediction of whether a motorist is impaired (after the direct evidence described above is collected).[136]

Based on the motorist's characteristics $\boldsymbol{x}$ and the motorist's race, the trooper's estimate of whether the motorist is impaired can be described in probability units, that is, $\Pr(IMP = 1 \mid R, \boldsymbol{x})$. The estimated probability of impairment is estimated with error because $\boldsymbol{x}$ (and R, if it is used) are not perfect measures of impairment. For example, a motorist who violates a lane-change law may be impaired or may have been temporarily distracted (e.g., due to talking on a cellular phone). The WSP uses NHTSA's scientifically validated Standardized Field Sobriety Test (SFST), which includes the horizontal gaze nystagmus, walk-and-turn, and one-leg stand. Although the accuracy of the SFST has significantly improved over time, the test is not perfect. Stuster (2001) cites a 1998 study that found officers who used the SFST correctly identified whether a motorist had a BAC above or below 0.08 in 91 percent of the cases. Due to the error involved in estimating the probability that a motorist is impaired, a random term ε is introduced that has an expected value of zero and a distribution such that $0 < \Pr(IMP = 1) < 1$. Eq. (5.6) assumes that the trooper's estimate of a motorist's probability of impairment is an additive function of the true impairment and the random term ε. In addition, Eq. (5.6) does not include $\boldsymbol{x}$, which I integrate out. This is done for two reasons. First, the data do not include most of the variables within $\boldsymbol{x}$ described above. Second, I am interested in the share of each racial group's marginal arrestees who tested over the legal alcohol concentration limit, that is, all marginal arrestees within a racial group, not just a subset based on particular values of the variables within $\boldsymbol{x}$. Moreover, the marginal motorists will be approximately identified using breath test results, not $\boldsymbol{x}$.

$$\Pr(IMP = 1 \mid R) = IMP(R) + \varepsilon \qquad (5.6)$$ [137]

[136] In a technical sense, all the evidence collected to establish probable cause in a DUI arrest decision is used as a statistical discriminator. This evidence is associated with impairment, which would be expensive to measure on the roadways because each trooper would need to have access to a state-certified breath test machine. However, the evidence used to establish probable cause for a DUI arrest is typically a more direct measure of impairment as compared to the directness of the evidence used to establish probable cause or reasonable suspicion in a high-discretion search decision. This statement is supported by the hit rate disparities between DUI arrests and high-discretion searches, where the former is near 90 percent and the latter is near three percent.

[137] I assume the distribution of ε is independent of race. The mean of ε should not depend on race because troopers who arrest motorists for a DUI obtain the actual BrAC results. Hence, they can compare these results to the motorist's characteristics that led to the arrest, and for future potential DUI arrests, they can

Analogous to Knowles et al.'s (2001) model of a trooper's search decision, assume that a trooper arrests a motorist for a DUI if the trooper's expected benefit is greater than or equal to his cost. Let *B* equal the expected benefit, which is a function of whether a motorist is impaired and the motorist's race (i.e., $B(IMP, R)$). $B(IMP = 1, R)$ is assumed to be positive since troopers are evaluated based on their DUI arrests (and many troopers may be self-motivated to arrest impaired drivers). $B(IMP = 0, R)$ is assumed to be zero (or negative due unimpaired arrestees filing complaints against the trooper or due to the WSP's potential legal liability for arresting an unimpaired motorist, which would adversely affect the arresting trooper). Let $C(R)$ be the trooper's cost to arrest a motorist of race *R* for a DUI. The cost is primarily the trooper's time, but the cost is allowed to vary by the motorist's race, which captures the effect of racial animosity (see Becker, 1957). In order to have $C(R)$ capture the entire effect of racial animosity for DUI arrests, the expected benefit of arresting an impaired motorist will be considered to be independent of race; hence, $B(IMP = 1, R) = B(IMP = 1)$. Using the notation above, a trooper will arrest a motorist for a DUI if the expected benefit to arrest is greater than or equal to the cost to arrest, as shown in Eq. (5.7).

$$\mathrm{E}[B(IMP, R)] \geq C(R) \tag{5.7}$$

The expected benefit of a DUI arrest is probability that a motorist is impaired multiplied by the expected benefit of arresting an impaired motorist, plus the probability that a motorist is not impaired multiplied by the expected benefit of arresting an unimpaired motorist (where this benefit is zero or negative). These components of the expected benefit are shown on the left side of Eq. (5.8), which follows from Eq. (5.7) above.

---

adjust the characteristics' (including race's) influence on predicting impairment. I also assume that the variance of ε is independent of race. Although studies show that officers cannot perfectly estimate impairment, I am not aware of any study that found that the variance of the error depends on the motorist's race.

$$\Pr(IMP = 1 \mid R) \times B(IMP = 1) + (1 - \Pr(IMP = 1 \mid R) \times B(IMP = 0, R) \geq C(R) \qquad (5.8)$$

I rearrange Eq. (5.8), which results in Eq. (5.9).

$$\Pr(IMP = 1 \mid R) \geq \frac{C(R) - B(IMP = 0, R)}{B(IMP = 1) - B(IMP = 0, R)} \qquad (5.9)$$

Next, I assume the benefit of arresting an unimpaired motorist is independent of race. This assumes that $B(IMP = 0, R) = B(IMP = 0)$. Due to the historical lawsuits involving civil rights violations, it may be the case that the expected benefit is more negative when an unimpaired minority is arrested as compared to when an unimpaired non-minority is arrested. Hence, if this is the case, before making a DUI arrest, a trooper will require a higher probability that a minority motorist is impaired as compared to a non-minority motorist. Although this may be the case, the primary purpose of this method is to determine whether a lower impairment threshold was used to arrest stopped minority motorists for a DUI as compared to the impairment threshold used for stopped non-minority motorists. Hence, this assumption may reduce the probability of detecting a statistically significant lower threshold due to racial animosity; however, if it is detected, then the evidence is stronger given this assumption. Based on this assumption, Eq. (5.9) becomes Eq. (5.10).

$$\Pr(IMP = 1 \mid R) \geq \frac{C(R) - B(IMP = 0)}{B(IMP = 1) - B(IMP = 0)} \qquad (5.10)$$

The right side of Eq. (5.10) is a function of $C(R)$ and $B(IMP)$, and $C(R)$ is the only term that includes race. The null hypothesis is that $C(R) = C$ for all R.[138] Under the null hypothesis, the right side of Eq. (5.10) becomes a constant, as shown in Eq. (5.11).

$$\Pr(IMP = 1 \mid R) \geq \frac{C - B(IMP = 0)}{B(IMP = 1) - B(IMP = 0)} \qquad (5.11)$$

The following steps and assumptions are required because the data do not identify the motorists who were arrested due to being impaired solely due to drugs or due to a combination of drugs and alcohol. However, the steps and assumptions are reasonable for this context. Most impaired motorists are impaired due to alcohol, but some are impaired due to drugs or a combination of alcohol and drugs. Therefore, $\Pr(IMP=1 \mid R)$ is split into three components, where *A* represents alcohol and *D* represents drugs: $\Pr(IMP=1 \mid A=1, D=0, R)$ is the trooper's estimate of a motorist's probability of impairment solely due to alcohol, $\Pr(IMP=1 \mid A=0, D=1, R)$ is the trooper's estimate of a motorist's probability of impairment solely due to drugs, and $\Pr(IMP=1 \mid A=1, D=1, R)$ is the trooper's estimate of a motorist's probability of impairment due a combination of alcohol and drugs. Because a trooper will likely categorize a potentially impaired motorist into only one of the three categories, the sum of the three probabilities above is substituted for $\Pr(IMP=1 \mid R)$ in Eq. (5.11), resulting in Eq. (5.12).

$$\Pr(IMP=1 \mid A=1, D=0, R) + \Pr(IMP=1 \mid A=0, D=1, R) + \Pr(IMP=1 \mid A=1, D=1, R) \geq \frac{C - B(IMP=0)}{B(IMP=1) - B(IMP=0)} \tag{5.12}$$

The first term on the left side of Eq. (5.12) is the probability that a motorist is impaired only due to alcohol as defined by the Washington law. Due to Washington's *per se* law, a motorist is considered to be impaired (i.e., in violation of the law) if his BAC is at or above 0.08, regardless of his level of impairment. If an arrestee tests below 0.08, he may still be charged with a DUI due to being sufficiently impaired at his measured BAC level (even if his impairment is only due to alcohol). NHTSA has sponsored many studies to estimate the effect of BAC levels on driving impairment. Moskowitz et al. (2000) found that some impairment begins with BAC levels as low as 0.02 percent and impairment is more pronounced at higher levels. Due to the burden of obtaining a blood sample, the WSP estimates an arrestee's BAC using a breath test. Therefore, I assume

[138] The null hypothesis could also be a one-sided hypothesis, which I will use below.

when troopers are estimating the probability that a motorist is impaired only due to alcohol (i.e., $\Pr(IMP=1 \mid A=1, D=0)$), they are actually estimating the probability that the motorist will test at or above 0.08 (i.e., $\Pr(BrAC \geq 0.08 \mid A=1, D=0)$). I substitute this assumption into the left side of Eq. (5.12), which results in Eq. (5.13).

$$\Pr(BrAC \geq 0.08 \mid A=1, D=0, R) + \Pr(IMP=1 \mid A=0, D=1, R) + \Pr(IMP=1 \mid A=1, D=1, R) \geq \frac{C - B(IMP=0)}{B(IMP=1) - B(IMP=0)} \tag{5.13}$$

I assume that the probability of an arrested motorist being impaired only due to drugs (or a combination of drugs and alcohol) is independent of the arrestee's race. This means that $\Pr(IMP=1 \mid A=0, D=1, R)$ equals $\Pr(IMP=1 \mid A=0, D=1)$, and $\Pr(IMP=1 \mid A=1, D=1, R)$ equals $\Pr(IMP=1 \mid A=1, D=1)$. Although minorities, especially blacks, are arrested at a much higher rate for drug abuse violations as compared to non-minorities (FBI, 2003),[139] those differences should have a minimal effect on this data. When a DUI arrest is based on impairment due to drugs, a drug recognition expert (DRE) is brought in to collect evidence; however, the data do not indicate when a DRE is used. However, the use of DREs is rare. For the entire WSP, a DRE is brought in for approximately 100 DUI arrests per month, which represented approximately six percent of the WSP's DUI arrests for an average month during the nine-quarter period.[140] Additionally, if a DUI arrest is solely for drugs, then the arrestee

[139] Although the arrest rate differences among racial groups for drug abuse violations is clearly established, the drug usage rate differences among racial groups is less clear. As discussed above in Logan and Schwilke (1996), marijuana was the most common drug found in fatally injured drivers. Compton et al. (2004) used the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions to estimate past-year marijuana use among racial groups and found the following usage rates: blacks (4.7%), whites (4.1%), and Hispanics (3.3%). The difference between blacks and whites was not statistically significant at the 0.05 level, but the difference between blacks and Hispanics was statistically significant. However, using the 2002 National Survey on Drug Use and Health, Ramchand et al. (in press) found that black marijuana users had more risky purchasing behaviors (i.e., would purchase marijuana from strangers, in the open, and away from their home), which explains some of the arrest-usage rate disparities among racial groups. Due to purchasing marijuana further away from their homes, this behavior may increase the likelihood that blacks drive under the influence of marijuana (and possibly other drugs) at a higher rate than non-minorities. This is an area for future research.

[140] Although drug possession and impairment due to drugs sometimes coexist, when the primary suspicion is possession, then as discussed above, SHCAT troopers are called in with canines to conduct the search. Based on a 2005 roster, none of the 16 troopers is a SHCAT member. There is a possibility that one of the

will likely test below 0.01. Due to this potential issue, these tests are dropped from the data. Additionally, when the WSP is testing its breath test machines, a reading of below 0.01 will likely be recorded. Hence, tests below 0.01 are dropped to remove these tests as well. For the 16 troopers' arrests, this only resulted in dropping ten of their tests, all of which involved non-minority arrestees. For the remaining tests, if a racial group had a higher share of its arrestees who were impaired due to the drugs (or a combination of drugs and alcohol) and tested below 0.08, then the estimated impairment threshold used to arrest motorists from this racial group would be underestimated. Based on these assumptions, Eq. (5.13) becomes Eq. (5.14).[141]

$$\Pr(BrAC \geq 0.08 \mid A=1,R) \geq \frac{C - B(IMP=0)}{B(IMP=1) - B(IMP=0)} - \Pr(IMP=1 \mid A=0, D=1) \quad (5.14)$$

The right side of Eq. (5.14) is a constant, and the empirical test for the null hypothesis is whether $\Pr(BrAC \geq 0.08 \mid A=1,R)$ is equal across all racial groups. However, the hypothesis test is based on marginal arrestees, not all arrestees. For example, even when the probability-of-impairment threshold used to decide which motorists to arrest is the same among racial groups, if the average BrAC level of arrestees from one racial group is higher than another racial group, then the average hit rate is expected to be higher for the former group.[142] The hypothesis test is also consistent with the original arrest decision discussed above, where a trooper arrests a motorist if the trooper's expected benefit is greater than or equal to his cost. Hence, if a trooper changes his cost to arrest a motorist for a DUI based on the motorist's race, then this change can

---

16 troopers was a SHCAT trooper in 2003 or 2004; however, due to the lengthy training involved, the WSP stated that a SHCAT trooper tends to remain as a SHCAT trooper. Moreover, these 16 troopers had no canine searches.

[141] The $\Pr(BrAC \geq 0.08 \mid A=1, D=0, R)$ will likely be larger than $\Pr(BrAC \geq 0.08 \mid A=1, D=1, R)$ since the latter arrestee's impairment is partially based on being impaired due to drugs, which are not detected with the breath test. However, the difference will be a constant that is independent of race since the probability of impairment due a combination of drugs and alcohol is assumed to be independent of the arrestee's race. Hence, I combine these terms into $\Pr(BrAC \geq 0.08 \mid A=1, R)$.

[142] The average BrAC level may be influenced by outliers, so to be precise, if the vast majority of the distribution of BrAC levels of arrestees from one racial group is higher than the vast majority of the

only be detected at the margin, that is, the least suspicious motorists he arrested for a DUI. For example, assume a trooper has racial animosity toward minority motorists, and therefore, arrests minority motorists who have at least an 85 percent probability of testing at or above 0.08, but only arrests non-minority motorists who have at least a 90 percent probability of testing at or above 0.08. Hence, the impairment threshold difference can only be detected for motorists who had between an 85 and 90 percent probability of testing at or above 0.08. This because all motorists who had at least a 90 percent probability of testing at or above 0.08 were arrested (regardless of their race), and because no motorists who had less than an 85 percent probability of testing at or above 0.08 were arrested (regardless of their race).

Due to a trooper's imperfect ability to measure impairment, the data do not precisely identify the marginal arrestees. For example, an arrestee with a BrAC of 0.25 could have been a marginal arrestee (prior to the arrest) because his ε could have been very large and positive. Contrastingly, an arrestee with a BrAC of 0.02 could have been a marginal arrestee (prior to the arrest) because his ε could have been very large and negative. Lastly, an arrestee with a BrAC equal to 0.08 may have not been a marginal arrestee prior to the arrest. The trooper's estimated probability of that arrestee being at or above 0.08 may have been very low (e.g., $\Pr(BrAC \geq 0.08) < 0.001$) or very high (e.g., $\Pr(BrAC \geq 0.08) > 0.999$).

Notwithstanding, the most likely group of marginal arrestees includes those who tested below a BrAC level that lies just above 0.08. This is because ε is assumed to have a mean of zero. This assumption is reasonable since troopers who arrest motorists for a DUI obtain the actual BrAC results. Hence, they can compare these results to the motorist's characteristics that led to the arrest, and for future potential DUI arrests, they can adjust the characteristics' influence on predicting impairment. Therefore, I assume the marginal arrestees are those who tested below 0.12.[143] To increase the number of arrestees and as a sensitivity analysis, I also assume the marginal arrestees are those who tested below 0.15.

distribution of BrAC levels of arrestees from another racial group, then the average hit rate is expected to be higher for the former group.

[143] I do not include arrestees who refused to test as marginal arrestees.

Lastly, the above model assumes $B(IMP = 1)$ is the same for all motorists who test at or above 0.08. However, as a motorist's BAC increases, impairment significantly increases, which increases the likelihood of a crash. Therefore, the WSP's (and a trooper's) benefit of arresting an impaired motorist increases as the motorist's BAC increases. However, the trooper's perceived benefit of arresting an impaired marginal motorist will be similar among marginal motorists since the marginal motorists have similar perceived BrACs.

5.5 Results

The results from the methods above are presented in two sections. The first section presents the results from the search rate analysis, which includes DUI arrests. The second section presents the results from the DUI arrest hit rate analysis.

5.5.1 Search Rates (including DUI arrests)

Table 5.7 shows how stop characteristics are associated with all stops, minority motorist stops, search rates of all stopped motorists, and search rates of stopped minority motorists. The first column of percentages shows the share of stops by stop characteristic, both for contextual and motorist characteristics. The second, third, and fourth columns of percentages respectively show the minority share of stops, the search rate of stopped motorists, and the search rate of stopped minority motorists for each stop characteristic. The first row summarizes each column by including all 19,020 stops, and shows that 36.4 percent of all stopped motorists were minority, 7.1 percent of all stopped motorists were searched, and 10.0 percent of stopped minority motorists were searched.

Based on the second column of percentages, as compared to the minority share of stops (36.4 percent), the minority share of stops was higher during nighttime hours and north of Sea-Tac Airport. Minority stops also made up a higher share of stops involving lane violations, three or more violations, criminal violations, and younger motorists. Based on the third column of percentages, although non-drug felony stops have been removed, the number of criminal violations is still a key predictor for being searched, whether the motorist is minority or non-minority. Approximately one half of stops with one criminal violation resulted in a search, and three quarters of stops with two or more criminal violations resulted in a search. The search rate also increased for stops at night, for stops north of Sea-Tac Airport, for stops involving lane violations, and for stops when the primary violation that initiated the stop was a criminal violation. As compared to other racial groups, the search rate was highest for American Indians, Hispanics, and blacks. When the search rate for a particular stop characteristic deviated from the overall search rate, the search rate of stopped minority motorists tended to deviate in the same direction for that stop characteristic as well (see third and fourth columns of percentages).

Table 5.7: Total Shares, Minority Shares, Search Rate, and Search Rate of Stopped Minority Motorists by Stop Characteristics (all unweighted)

Stop Characteristic	Share of Stops (19,020 Stops)	Minority Share of Stops	Search Rate of Stopped Motorists	Search Rate of Stopped Minority Motorists
Total	100.0%	36.4%	7.1%	10.0%
Quarter				
1Q03	13%	33%	7%	11%
2Q03	12%	37%	8%	10%
3Q03	13%	36%	7%	10%
4Q03	12%	37%	8%	13%
1Q04	12%	37%	7%	8%
2Q04	12%	38%	8%	11%
3Q04	9%	37%	6%	7%
4Q04	9%	39%	7%	9%
1Q05	7%	36%	6%	8%
Day of Week				
Weekday	69%	36%	7%	11%
Weekend	31%	37%	7%	9%
Time of Day				
12 to 4 a.m.	23%	48%	18%	21%
4 to 8 a.m.	12%	34%	4%	5%
8 a.m. to 12 p.m.	18%	30%	2%	3%
12 to 4 p.m.	16%	27%	2%	2%
4 to 8 p.m.	15%	33%	5%	6%
8 p.m. to 12 a.m.	16%	41%	7%	9%
Milepost				
South of Sea-Tac Airport	74%	33%	6%	9%
North of Sea-Tac Airport	26%	45%	10%	13%
Violation Type				
Speeding	55%	34%	5%	8%
Equipment	11%	31%	3%	4%
Safety Belt	6%	37%	3%	3%
Lane Violation	17%	46%	13%	15%
Other	11%	36%	4%	6%
Criminal	2%	43%	81%	83%
Contact Type				
Non-Aggressive Driving	88%	36%	7%	9%
Aggressive Driving	12%	38%	11%	14%
Number of Violations				
1	46%	33%	2%	3%
2	30%	37%	7%	9%
3+	24%	41%	17%	23%
Number of Criminal Violations				
0	96%	36%	5%	7%
1	3%	50%	48%	47%
2+	1%	52%	72%	75%
Motorist Gender				
Male	72%	38%	8%	11%
Female	28%	32%	5%	6%
Motorist Age				
16-25	34%	38%	9%	11%
26-35	27%	43%	8%	12%
36-45	19%	34%	6%	9%
46+	20%	27%	4%	5%
Motorist Race				
White	63.6%	0.0%	5.4%	NaN
Black	15.5%	100.0%	11.6%	11.6%
American Indian	0.2%	100.0%	15.4%	15.4%
Asian/Pacific Islander	10.7%	100.0%	7.5%	7.5%
East Indian	1.5%	100.0%	2.8%	2.8%
Hispanic	7.8%	100.0%	11.8%	11.8%
Other	0.8%	100.0%	8.6%	8.6%

Based on the weights from propensity score weighting model, Table 5.8 shows the characteristics of stopped minority motorists as compared to unweighted and weighted stopped non-minority motorists (see first and second set of columns, respectively). With the exception of the interaction terms listed at the bottom of the table, the characteristics are sorted by their relative influence in predicting whether a stopped motorist was minority versus non-minority.[144] The gray highlight indicates characteristics where the difference is greater than five percentage points or the effect size difference is greater than 0.1. As compared to non-minority stops, minority stops consisted of a higher proportion of young, male motorists, and a higher proportion of lane violations, criminal violations, and multiple-violation stops. Minorities had a higher proportion of their stops between 12 and 4 a.m. and north of Sea-Tac Airport. As the right-side of the table shows, the weights mostly eliminate these differences.

[144] The relative influence calculation is based on the non-interacted variables. The interacted variables are included in the table to evaluate how well the interacted variables matched between minority stops and weighted non-minority stops.

Table 5.8: Comparison of Minority Motorist Stop Characteristics with Non-Minority Motorist Stop Characteristics (unweighted and weighted)

	Unweighted				Weighted		
Stop Characteristic	Minority Stops (6,926)	Non-Minority Stops (12,094)	Difference (Subject - Peers)	Effect Size Difference	Weighted Non-Minority Stops (9,178)	Weighted Difference (Minority Minus Non-Minority)	Effect Size Difference
Trooper							
78	6.5%	7.4%	-0.9%	0.04	6.4%	0.1%	0.00
285	2.4%	4.2%	-1.8%	0.11	2.5%	0.0%	0.00
346	8.5%	8.3%	0.2%	0.01	8.5%	0.0%	0.00
376	6.9%	6.7%	0.2%	0.01	6.9%	0.0%	0.00
401	9.3%	6.1%	3.2%	0.11	9.3%	0.0%	0.00
518	6.9%	6.9%	0.0%	0.00	6.8%	0.0%	0.00
541	3.7%	3.5%	0.1%	0.01	3.7%	0.0%	0.00
579	7.3%	7.2%	0.1%	0.01	7.3%	0.1%	0.00
590	5.1%	4.3%	0.8%	0.04	5.1%	0.0%	0.00
837	6.5%	6.4%	0.1%	0.00	6.4%	0.1%	0.00
840	9.9%	8.4%	1.6%	0.05	10.1%	-0.1%	0.00
1103	6.0%	6.7%	-0.7%	0.03	6.0%	0.0%	0.00
1109	4.7%	5.6%	-0.8%	0.04	4.6%	0.1%	0.00
1209	7.1%	8.7%	-1.5%	0.06	7.2%	-0.1%	0.00
1236	4.6%	5.2%	-0.6%	0.03	4.5%	0.0%	0.00
1411	4.6%	4.6%	0.0%	0.00	4.6%	0.0%	0.00
Time							
12 to 4 a.m.	30.3%	18.6%	11.7%	0.25	30.3%	0.0%	0.00
4 to 8 a.m.	11.4%	12.6%	-1.3%	0.04	11.4%	0.0%	0.00
8 a.m. to 12 p.m.	15.1%	19.8%	-4.7%	0.13	14.9%	0.2%	0.01
12 to 4 p.m.	11.7%	18.0%	-6.3%	0.20	12.2%	-0.5%	0.02
4 to 8 p.m.	13.3%	15.6%	-2.4%	0.07	13.0%	0.2%	0.01
8 p.m. to 12 a.m.	18.3%	15.3%	2.9%	0.08	18.2%	0.1%	0.00
Primary Violation Type							
Speeding	51.4%	56.3%	-4.9%	0.10	51.8%	-0.4%	0.01
Equipment	9.1%	11.7%	-2.6%	0.09	9.0%	0.1%	0.00
Safety Belt	5.6%	5.5%	0.1%	0.00	5.4%	0.1%	0.01
Lane Violation	21.2%	14.1%	7.1%	0.17	21.2%	0.0%	0.00
Other	10.6%	10.7%	-0.2%	0.01	10.5%	0.1%	0.00
Criminal	2.1%	1.6%	0.5%	0.03	2.0%	0.1%	0.01
Motorist Age							
16-25	35.2%	33.1%	2.0%	0.04	35.2%	0.0%	0.00
26-35	31.7%	24.1%	7.6%	0.16	31.4%	0.3%	0.01
36-45	18.1%	19.7%	-1.6%	0.04	18.2%	-0.1%	0.00
46+	15.1%	23.1%	-8.0%	0.22	15.2%	-0.1%	0.00
Quarter							
1Q03	12.4%	14.1%	-1.8%	0.05	12.2%	0.2%	0.01
2Q03	12.2%	12.1%	0.1%	0.00	12.2%	0.0%	0.00
3Q03	13.2%	13.5%	-0.3%	0.01	13.4%	-0.3%	0.01
4Q03	11.6%	11.6%	0.1%	0.00	11.6%	0.0%	0.00
1Q04	12.2%	12.2%	0.1%	0.00	12.2%	0.0%	0.00
2Q04	12.6%	11.8%	0.8%	0.02	12.5%	0.1%	0.00
3Q04	9.1%	8.9%	0.2%	0.01	9.1%	0.0%	0.00
4Q04	9.4%	8.5%	0.9%	0.03	9.3%	0.1%	0.00
1Q05	7.4%	7.4%	0.0%	0.00	7.5%	-0.1%	0.00
Milepost							
South of Sea-Tac Airport	67.5%	77.1%	-9.6%	0.20	67.5%	0.0%	0.00
North of Sea-Tac Airport	32.5%	22.9%	9.6%	0.20	32.5%	0.0%	0.00
Number of Violations							
1	42.0%	48.0%	-6.0%	0.12	42.3%	-0.3%	0.01
2	30.3%	29.6%	0.8%	0.02	30.4%	-0.1%	0.00
3+	27.6%	22.4%	5.2%	0.12	27.3%	0.3%	0.01
Criminal Violation							
0	94.0%	96.6%	-2.6%	0.11	94.4%	-0.4%	0.01
1	4.7%	2.7%	2.0%	0.09	4.5%	0.2%	0.01
2+	1.3%	0.7%	0.6%	0.05	1.1%	0.1%	0.01
Motorist Gender							
Male	75.2%	70.1%	5.1%	0.12	75.1%	0.1%	0.00
Female	24.8%	29.9%	-5.1%	0.12	24.9%	-0.1%	0.00
Day of Week							
Weekday	68.5%	69.8%	-1.3%	0.03	68.5%	0.1%	0.00
Weekend	31.5%	30.2%	1.3%	0.03	31.5%	-0.1%	0.00
Contact Type							
Non-Aggressive Driving	87.0%	88.1%	-1.1%	0.03	87.1%	-0.1%	0.00
Aggressive Driving	13.0%	11.9%	1.1%	0.03	12.9%	0.1%	0.00
Gender:Age							
Male, 16-25	25.5%	21.5%	4.0%	0.09	25.3%	0.2%	0.01
Female, 16-25	9.6%	11.6%	-1.9%	0.07	9.9%	-0.3%	0.01
Male, 26-35	24.0%	17.1%	6.9%	0.16	24.0%	0.1%	0.00
Female, 26-35	7.7%	7.0%	0.7%	0.02	7.5%	0.2%	0.01
Male, 36-45	13.7%	14.7%	-0.9%	0.03	14.2%	-0.4%	0.01
Female, 36-45	4.3%	5.0%	-0.7%	0.03	4.0%	0.3%	0.02
Male, 46+	11.8%	16.9%	-5.0%	0.16	11.7%	0.1%	0.00
Female, 46+	3.3%	6.2%	-2.9%	0.17	3.5%	-0.2%	0.01
Time:Day							
12 to 4 a.m., Weekday	22.2%	13.8%	8.4%	0.20	21.9%	0.4%	0.01
4 to 8 a.m., Weekday	7.5%	8.5%	-1.0%	0.04	7.4%	0.1%	0.00
8 a.m. to 12 p.m., Weekday	8.6%	11.4%	-2.9%	0.10	8.2%	0.4%	0.01
12 to 4 p.m., Weekday	7.5%	12.7%	-5.2%	0.20	8.3%	-0.7%	0.03
4 to 8 p.m., Weekday	9.5%	11.8%	-2.3%	0.08	9.5%	-0.1%	0.00
8 p.m. to 12 a.m., Weekday	13.2%	11.5%	1.7%	0.05	13.2%	0.1%	0.00
12 to 4 a.m., Weekend	8.1%	4.8%	3.3%	0.12	8.4%	-0.3%	0.01
4 to 8 a.m., Weekend	3.9%	4.1%	-0.2%	0.01	3.9%	-0.1%	0.00
8 a.m. to 12 p.m., Weekend	6.6%	8.4%	-1.8%	0.07	6.7%	-0.1%	0.01
12 to 4 p.m., Weekend	4.2%	5.3%	-1.1%	0.06	4.0%	0.2%	0.01
4 to 8 p.m., Weekend	3.8%	3.8%	0.0%	0.00	3.5%	0.3%	0.02
8 p.m. to 12 a.m., Weekend	5.0%	3.8%	1.2%	0.06	5.0%	0.0%	0.00

Based on the weights from propensity score weighting model, Table 5.9 shows the search rate results for all searches, low-discretion searches, DUI arrests, and high-discretion searches for stopped minority and non-minority motorists. The first set of five columns includes the primary results, which is followed by a set of three columns that includes additional information. The first set of columns includes the search discretion level, the search rate of stopped minorities, the weighted search rate of stopped non-minorities, the difference between these rates (which is $\hat{\theta}$), and the one-sided p-value of $\hat{\theta}$. In the second set of columns, the first column includes the unweighted search rate of stopped non-minorities, followed by the hit rates of searched minorities and searched non-minorities, respectively.

For each search discretion level, the search rate of stopped minorities exceeds the weighted search rate of stopped non-minority motorists. The results are statistically significant at the 0.05 level for all searches and high-discretion searches.[145] The one-sided p-value represents the probability that the estimate of θ would be this or more extreme, assuming null hypothesis is true (i.e., $\theta \leq 0$). For each search discretion level, the hit rate of searched minority motorists is slightly lower than the hit rate of searched non-minority motorists; however, the differences are not significant at the 0.05 level.

Table 5.9: Estimated Use of Race for Search Decision (including DUI arrests)

Search Discretion Level	Search Rate of Stopped Minorities (6,926 stops)	Search Rate of Stopped Non-Minorities (weighted) (9,178 stops)	Difference	p-value (one-sided)	Search Rate of Stopped Non-Minorities (12,094 stops)	Hit Rate of Searched Minorities	Hit Rate of Searched Non-Minorities
All	10.0%	8.8%	1.2%	0.009*	5.4%	10%	12%
Low Discretion	2.2%	1.9%	0.3%	0.141	1.3%	20%	23%
DUI Arrest	7.2%	6.5%	0.6%	0.074	3.9%	7%	9%
High Discretion	0.7%	0.4%	0.3%	0.008*	0.2%	2%	4%

*Significantly higher search rate of stopped minority motorists as compared to weighted search rate of stopped non-minority motorists at the 0.05 level (one-sided)

As a sensitivity analysis, the logistic regression model in Eq. (5.5) is also used to estimate the use of race as a factor for each search discretion level. The results are similar to the propensity score weighting model results. The parameter $\hat{\delta}_1$ is consistently greater than zero, and like the propensity score weighting model results, is statistically significant at the 0.05 level for all searches and high-discretion searches.

5.5.2 DUI Arrest Hit Rate Analysis

Figure 5.3 shows the BrAC results of DUI arrestees from the WSP, the WSP (excluding District 2), District 2 (excluding the 16 troopers), and the 16 troopers, respectively.[146] Of the 16 troopers' arrestees, 91.0 percent tested at or above 0.08 (or refused to test).[147] For the WSP and District 2, this percentage was approximately one percentage point lower.[148] Of the 16 troopers' arrestees who tested below 0.15, 84.7 percent tested at or above 0.08. For the WSP and District 2, this percentage was approximately three percentage points lower. Of the 16 troopers' arrestees who tested below 0.12, 75.4 percent tested at or above 0.08. For the WSP and District 2, this percentage was approximately four percentage points lower. Hence, as compared to other WSP troopers, these 16 troopers had a higher share of their arrestees (including their marginal arrestees) test at or above 0.08.

---

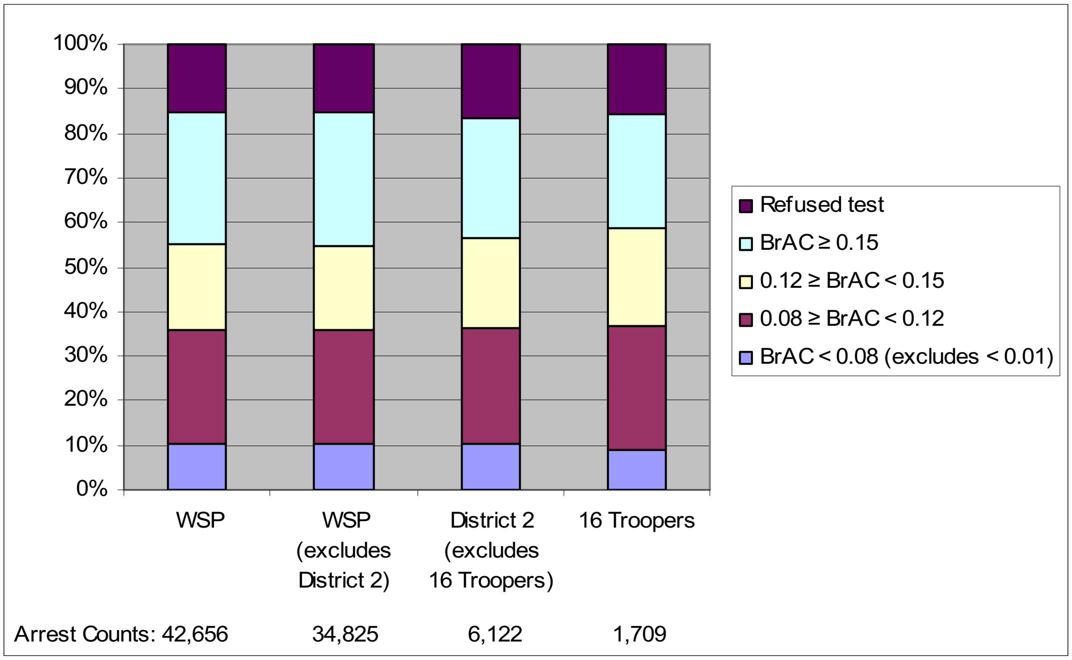
[145] The results are substantially the same when the set of binary variables indicating the trooper who made the stop are removed from the model. The primary difference is that the result for DUI arrests is statistically significant at the 0.05 level.

[146] As discussed in the methods section, the DUI arrests that are analyzed from the 16 troopers decreased from 1,719 to 1,709 arrests because BrAC tests below 0.01 are excluded.

[147] As stated above, I assume an arrestee who refuses to take the breath test had a BrAC at or above 0.08. For brevity, when I report the share that tested at or above 0.08, I do not also explicitly state "or refused to test"; however, that assumption is implied unless stated otherwise.

[148] The share of WSP arrestees who tested at or above 0.08 (or refused to test) is significantly higher than the share reported in Lovrich et al. (2003) (see Table 5.3). However, their data covered an earlier time period and their results may have categorized test refusals differently. If refusals are removed from the WSP sample, then 88.0 percent of the WSP arrestees tested at or above 0.08, which is very similar to the results reported in the WSP's "Stateside Summary of Breath Alcohol Data" for 2003 to 2005, which also excluded tests below 0.01. (If tests below 0.01 are included, then 86.6 percent of the WSP arrestees tested at or above 0.08.)

Figure 5.3: BrAC Test Results of DUI Arrestees for WSP, District 2, and 16 Troopers



For the DUI arrests of the 16 troopers, Table 5.10 shows the number of arrestees by racial group for three groups of arrestees: all arrestees, arrestees with a BrAC less than 0.15, and arrestees with a BrAC less than 0.12.[149]

Table 5.10: Number of the 16 Troopers' DUI Arrestees by Racial Group and BrAC Level

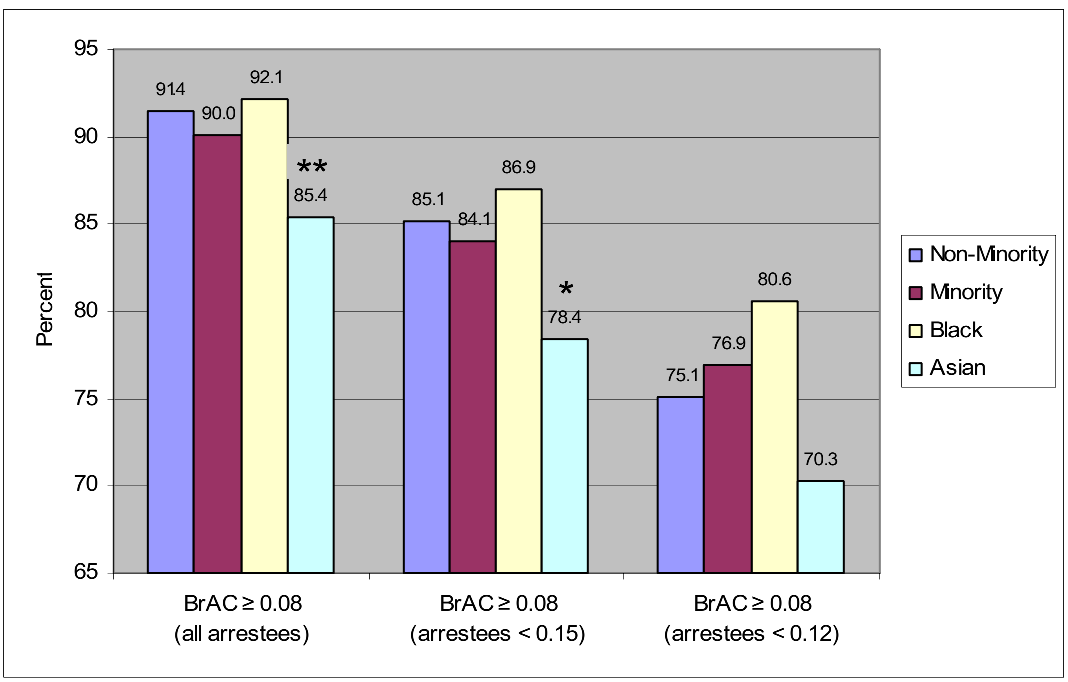
Racial Group	Arrestees (All)	Arrestees (BrAC < 0.15)	Arrestees (BrAC < 0.12)
Non-Minority	1,157	666	397
Minority	481	301	208
Black	330	199	134
Asian	151	102	74
Total	1,638	967	605

Figure 5.4 includes the 16 troopers' DUI arrestees from Table 5.10 and shows the share of the arrestees that tested at or above 0.08 by racial group and BrAC level. With

[149] The DUI arrest count decreased from 1,709 to 1,638 because the data did not include the race of 57 arrestees. I also excluded American Indians and Hispanics because there were only seven arrestees from each racial group. The number of Hispanic arrestees is understated because some of the breath test machines did not have a racial group code for Hispanic.

the exception of Asians, the shares of arrestees that tested at or above 0.08 are similar among racial groups. For the three groups of arrestees (i.e., all arrestees, arrestees with a BrAC less than 0.15, and arrestees with a BrAC less than 0.12) respectively, the shares of Asian arrestees who tested at or above 0.08 are 6.0, 6.7, and 4.8 percentage points lower than the respective non-minority shares, and the first two differences are statistically significant. The one-sided significance level represents the probability that the estimated minority hit rate would be below the estimated non-minority hit rate by this difference or more when the hit rates were actually the same (or when the minority hit rate was actually higher than the non-minority hit rate). The one-sided p-values are calculated using a chi square test. The uncertainty surrounding the estimated hit rate difference is due the finite sample size. Although the data includes all DUI arrests for these troopers, the data is considered to be a sample (e.g., a sample of their lifetime DUI arrests).

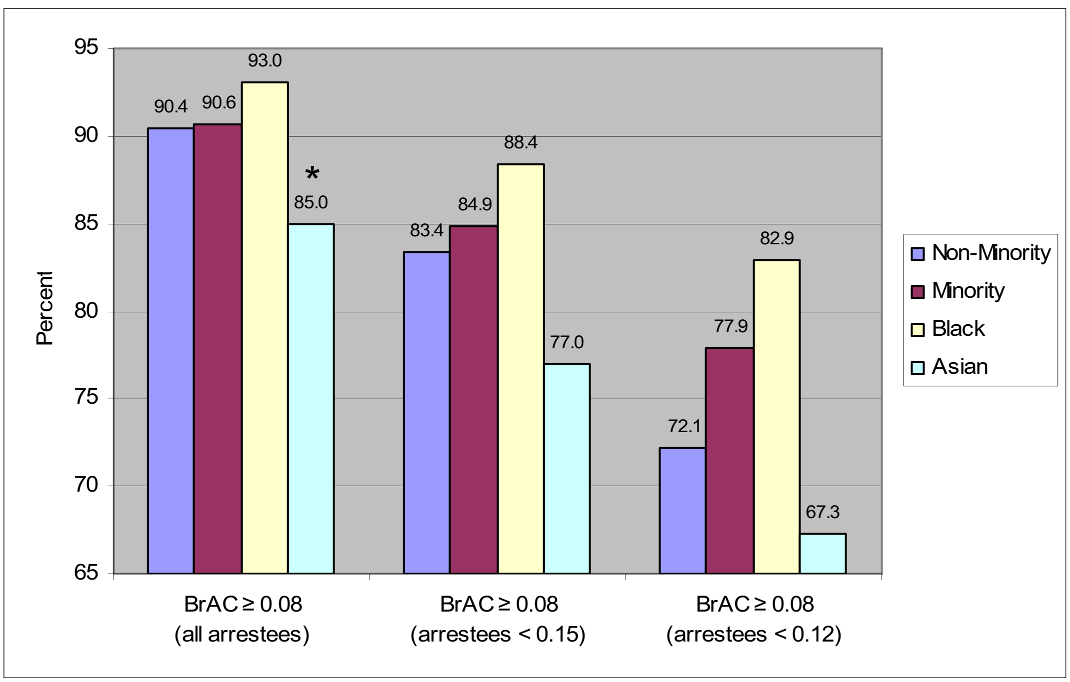
Figure 5.4: BrAC Test Results of 16 Troopers' DUI Arrestees by Race



*Significantly lower hit rate for arrestees represented by the bar as compared to non-minority arrestees at the 0.05 level (one-sided)
**Significantly lower hit rate for arrestees represented by the bar as compared to non-minority arrestees at the 0.01 level (one-sided)

The use of race may vary across age and gender, especially for young males (Newport, 1999). The DUI arrest data include the arrestee's gender, but do not include the arrestee's age. Of the troopers' 1,638 arrestees analyzed above, 78 percent are male. Figure 5.5 shows the shares of 16 troopers' male DUI arrestees who tested at or above 0.08 by race. The figure includes the three groups of arrestees that were included in Figure 5.4 above. The results are similar to the results that included all arrestees. For arrestees who tested at or above 0.12, with the exception of Asians, the shares of arrestees who tested at or above 0.08 are similar among racial groups. For the three groups of arrestees respectively, the shares of Asian arrestees who tested at or above 0.08 are 5.5, 6.3, and 4.8 percentage points lower than the non-minority shares, and the first difference is statistically significant (see figure).[150] On the other hand, when only arrestees who tested below 0.12 are considered, the share of non-minority arrestees who tested at or above 0.08 was 10.8 percentage points lower than the respective black arrestee share, and this difference is statistically significant at the 0.05 level (two-sided p-value = 0.025).[151]

Figure 5.5: BrAC Test Results of 16 Troopers' DUI Male Arrestees by Race



*Significantly lower hit rate for arrestees represented by the bar as compared to non-minority arrestees at the 0.05 level (one-sided)

[150] The percentages displayed in the figure are rounded; however, the percentage point differences between Asians and non-minorities are based on more precise percentages.
[151] The equivalent one-sided p-value is 0.988.

5.6 Discussion

The results from the search rate analysis, which includes DUI arrests, and the results from the DUI arrest hit rate analysis are respectively discussed next.

5.6.1 Search Rates (including DUI arrests)

The above models estimated the use of race as a factor in search decisions, where the use-of-race estimate includes both appropriate and inappropriate uses of race. As discussed in Chapter 2, racial profiling is defined as when the use of race is inappropriate, as defined by a law or policy. For these troopers, an appropriate use of race may include using race when it is part of a particular suspect's description or part of a criminal profile. For example, if a particular suspect was fleeing a crime scene using this stretch of Interstate 5, a trooper may use the suspect's race as one factor among many to stop and possibly search motorists who matched the suspect's description. While the law is more restrictive on the use of race when it is part of a criminal profile, there may be some situations when its use may be appropriate. As stated in Chapter 3, the WSP states that approximately 75 percent of methamphetamine being consumed in Washington state originates from Mexico (Batiste, 2005). In this situation, whether a motorist's Hispanic appearance may be used as one factor among many in a stop decision depends on whether WSP has additional specific information about the methamphetamine couriers and when they were using this stretch of Interstate 5. Although the above two situations are plausible appropriate uses of race, these 16 troopers initiated the vast majority of their stops within the "core four" priorities of traffic enforcement: driving under the influence (DUI), dangerous speeding, seat belt, and aggressive driving enforcement. Hence, assuming that the use-of-race estimates above control for all confounding characteristics, then they mostly comprise of estimates of the inappropriate use of race.

The results presented in Table 5.9 show that minority motorists were searched at a higher rate after controlling for available characteristics that differed between stopped minority and non-minority motorists. The results also show that the search rate disparity between stopped minority and non-minority motorists decreases when available characteristics are controlled for. While these results provide evidence that race may have been used as a factor in the troopers' search decisions that resulted in the search rate of stopped minority motorists to be higher than the search rate of similarly situated stopped non-minority motorists, the evidence is not conclusive due to the model not including unobserved characteristics that influence the decision to search or arrest a motorist for a DUI that also may be associated with a racial group.[152] Some of the characteristics include a trooper spotting contraband in plain site or a motorist making a furtive movement as the trooper approaches the vehicle. And some impairment indicators include the motorist's driving behavior, field sobriety test results, and the smell of alcohol or drugs. Additionally, these characteristics may also influence whether a motorist is stopped in the first place; hence, stopped motorists may not have the same risk of being searched.

Searches from each discretion level are discussed next. For low-discretion searches, which are all non-drug felony incident-to-arrest searches, the remaining arrests are likely to involve relatively little trooper discretion. However, if the trooper used race as a factor in deciding which motorists to arrest, which led to the search, then controlling for the seriousness and number of violations recorded may mask the use of race in the search decision.[153] For DUI arrests, the data do not include the key indicators used to

---

[152] In the stop analysis, the use of race as a factor adversely affecting minority motorists in stop decisions was described as stopping a higher share of minorities. In the case of searches, the analogous description is that the search rate of stopped minority motorists is higher than the search rate of similarly situated stopped non-minority motorists (i.e., a higher share of stopped minority motorists are searched as compared to the share of similarly situated stopped non-minority motorists who are searched). If the situation was described as searching a higher share of stopped minorities or a lower share of stopped non-minorities, then this would imply the overall search rate changed as compared to a situation where race was not being used as a factor in search decisions. Analogous to the stop analysis, if the troopers used race as a factor in their search decisions that adversely affected stopped minority motorists, then it is a separate question whether this means that they searched additional stopped minority motorists or searched fewer stopped non-minority motorists. For example, assume a trooper stops 100 minorities and 100 non-minorities. In Scenario *A*, assume he is race neutral and searches 10 minorities and 5 non-minorities. (Assume the minority search rate is higher due to reasons that justify the higher rate.) In Scenario *B*, assume that the trooper uses race as a factor in his search decisions, resulting in an even larger search rate disparity between minority and non-minority motorists. If the trooper does not change the number of searches between the two scenarios, then in Scenario *B*, he could, for example, search 12 minorities and 3 non-minorities. However, if he does change the number of searches, then in Scenario *B*, he might search additional minorities as compared to Scenario *A* (e.g., search 15 minorities and 5 non-minorities) or fewer non-minorities as compared to Scenario *A* (e.g., search 10 minorities and 3 non-minorities). If the troopers in this study used race as a factor in their search decisions, it is beyond the scope of this study to determine whether and to what degree they changed the number of their searches as compared to a situation where they did not use race as a factor in their search decisions. This same issue applies to DUI arrests. This is an area of research that should be pursued.

[153] However, I chose to still include these characteristics in the model because when these characteristics are included, then if the result provides evidence that troopers were using race as a factor in their stop decisions, then the evidence is more conclusive since the motorist's characteristics that may be associated with race and being searched have been controlled for.

establish probable cause to arrest a motorist for a DUI such as the motorist's driving behavior and his field sobriety test results. If these indicators differ among racial groups, then the DUI arrest rates among racial groups are not expected to be the same. Although high-discretion searches are very rare (45 minority and 27 non-minority searches), these searches normally have the most potential at identifying a use-of-race disparity that is mostly within the trooper's control. However, most of these searches were frisk searches, and the data do not include the factors that the trooper used to establish reasonable suspicion that the motorist may have been carrying a weapon that could endanger him or was involved in a crime that had or was about to occur. The data indicated that contraband was found in one of the searches involving a minority and one of the searches involving a non-minority (and the hit rate difference is not statistically significant at the 0.05 level). For the searches that did not yield contraband, it is important to identify the factors used to establish reasonable suspicion. Therefore, while there is evidence that stopped minorities were searched at a higher rate than similarly situated stopped non-minorities, due to unobserved factors that affect search decisions that may also differ among racial groups, the evidence is inconclusive whether race was inappropriately used as a factor in high-discretion search decisions.

The low-discretion search and DUI arrest results above are similar to the search analysis results in Lovrich et al. (2003) for APA 6; however, that study found that black motorists were less likely to be searched than non-minority motorists on a high-discretion basis. However, that study analyzed stops within APA 6 during an earlier period. For the entire WSP, Lovrich et al. (2005) found that minorities were subjected to higher search rates than non-minorities after controlling for available characteristics that differed between the groups (see Table 5.1). However, based on interviews with the WSP, they also concluded that the results were inconclusive due to unobserved factors that affect search decisions that may also differ among racial groups.

Besides attempting to determine whether race may have been used as a factor in a search decision, it is important to quantify the estimated use. For this discussion, the unobserved characteristics that influence search decisions are assumed to be unassociated with a racial group of stopped motorists. The estimate will be made for high-discretion searches since the trooper had the highest degree of discretion in deciding whether to search these motorists. If the troopers had searched 21 fewer stopped minority motorists (which represents 47 percent of stopped minorities who were searched on a high-discretion basis), then the high-discretion search rate of stopped minorities would have been the same as the high-discretion search rate of similarly situated stopped non-minorities, which is the weighted high-discretion search rate of stopped non-minority motorists. Although this represents a small number of minority motorists over a period of two years and three months, these 16 troopers represent approximately two percent of the WSP's traffic enforcement troopers and their high-discretion searches that were analyzed represent approximately one percent of the WSP's high-discretion searches. Hence, if the estimated use of race as a factor in high-discretion search decisions by these troopers is representative of all troopers, the impact is much more substantial. (But as stated in the beginning of this paragraph, this is a hypothetical exercise because the estimated use of race by these 16 troopers in high-discretion search decisions may be biased due to confounding characteristics.)

5.6.2 DUI Arrest Hit Rate Analysis

The DUI arrest hit rate analysis circumvents many of the issues stemming from unobserved variables that are associated with race that also affect DUI arrest decisions. Additionally, unlike the results discussed above, this use-of-race estimate only includes the use of race due to racial animosity and does not include the use of race due to statistical discrimination.

In general, the results from the DUI arrest hit rate analyses do not provide evidence that a lower probability-of-impairment threshold was used to arrest minority motorists as compared to the threshold used to arrest non-minority motorists. However, when certain subsets of arrestees are compared, there is some evidence that different thresholds were used. When comparing black and non-minority DUI arrestees, the results do not provide any evidence that a lower probability-of-impairment threshold was used to arrest black motorists as compared to the threshold used to arrest non-minority motorists. In fact, when only marginal male arrestees are considered, the results provide evidence that the probability-of-impairment threshold used to arrest non-minority male motorists was statistically lower than the threshold used to arrest black male motorists. When comparing Asian and non-minority DUI arrestees, the results show that the shares of

Asian arrestees who tested at or above 0.08 are consistently (and often statistically) lower than the shares of non-minority arrestees who tested at or above 0.08. However, when only arrestees who tested below 0.12 are considered, the differences are not statistically different from zero at the 0.05 level. This is due to a smaller percentage point difference between the non-minority and Asian shares of arrestees who tested at or above 0.08, and also due to the smaller number of observations.

The results from the marginal arrestees (i.e., those who tested below 0.12) are the most useful in estimating the use of race as a factor in these decisions because those arrestees provide evidence whether the probability-of-impairment thresholds used to arrest motorists differed among racial groups. This is because if the average BrAC level of arrestees from one racial group is higher than another racial group, then the average hit rate is expected to be higher for the former group (even when the probability-of-impairment threshold used to decide which motorists to arrest is the same for the two racial groups). In this case, 66 percent of non-minority and 51 percent of Asian arrestees tested at or above 0.12 (or refused to test), respectively, indicating a higher average BrAC level for non-minority arrestees. Based on the results that only include arrestees who tested below 0.12, there is some evidence that the probability-of-impairment threshold used to arrest Asian motorists was lower than the threshold used to arrest non-minority motorists; however, the evidence is not conclusive due to the high one-sided p-value (0.193). This p-value means that if the true percentage point difference was zero or had opposite sign (i.e., the share of non-minority arrestees who tested at or above 0.08 was less than the share of Asian arrestees who tested at or above 0.08), then one would observe this positive difference or one more extreme 19.3 percent of the time. As more data is collected, then the tests of additional motorists who tested below 0.12 can be analyzed to see whether the results are consistent with the above results, or if they point toward race not being used as a factor that results in a relatively lower probability-of-impairment threshold being used to arrest Asian motorists for a DUI.

Although the major concern with racial profiling does not typically involve Asian motorists, Lovrich et al. (2005) stated that based on focus groups with WSP troopers, a stop tended to be more confrontational involving Asians as compared to non-Asians. They noted that this may be due to cultural and language differences. Moreover, Lovrich states that based on their analysis and focus groups with WSP troopers, Asians are involved in a higher share of accidents as compared to their share of motorists. When a trooper contacts a motorist due to an accident, he will closely scrutinize the motorist for being impaired. Although the evidence for impairment is still substantially based on BrAC test results, since an accident occurred, then this could be used as incriminating evidence that the motorist was legally impaired at a BrAC below 0.08.

As discussed in the methods section, the estimated hit rate disparity only measures a disparity in BrAC test results. Hence, these results assume that of the arrestees who tested below 0.08, the shares of those arrestees who were impaired due to drugs (or a combination of alcohol and drugs) did not differ among racial groups, otherwise the results above would be biased. Given that DUI arrests due to drug impairment is rare (approximately six percent), this assumption should not significantly bias the results. However, the data still likely include most of the arrestees who were impaired from drugs—only 10 arrestees (0.6 percent) were dropped due to the arrestee testing below 0.01. But if DUI arrests due to drug impairment followed drug abuse arrest rates (i.e., disproportionately involved minority motorists (especially blacks)), then the above estimated probability-of-impairment threshold used to arrest minority motorists would underestimate the actual threshold by more than the underestimate of threshold used to arrest non-minority motorists.

As in the search rate discussion, besides attempting to determine whether race may have been used as a factor in a DUI arrest decision, it is important to quantify the estimated use. For this discussion, the estimated hit rate difference between Asians and non-minority marginal arrestees is assumed to be statistically significant and the difference is assumed to not be due to different drug impairment rates. The magnitude of the hit rate disparity can better be interpreted as the reduction in the number of marginal Asian DUI arrestees who tested below 0.08 that would be needed to equalize the hit rates of marginal Asian and non-minority arrestees. Of the 74 Asian marginal arrestees (i.e., those who tested below 0.12), 22 tested below 0.08. If the troopers had arrested five fewer Asian motorists who tested below 0.08, then the shares of Asian and non-minority marginal motorists who tested at or above 0.08 would have been the same.[154] Although the reduction represents a small number of motorists over a period of two years and three months, these 16 troopers represent approximately two percent of the WSP's traffic enforcement troopers and their DUI arrests were responsible for approximately 4 and 13 percent of the WSP's DUI arrests and Asian DUI arrests, respectively. Hence, if these troopers are representative of all troopers, the impact is more substantial; however, the impact is still somewhat small as compared to the number of Asian motorists traveling on Washington state highways. (But as stated in the beginning of this paragraph, this is a hypothetical exercise because the DUI arrest hit rate difference between marginal Asian and non-minority arrestees was not statistically significant.)

5.6.3 Summary

The primary purpose of this chapter was to estimate the 16 troopers' average use of race as a factor in deciding which stopped motorists to search or arrest for a DUI. The methods evaluated search rates and DUI arrest hit rates. Although the estimates from these methods are important in themselves, the estimate from the search rate analysis is primarily needed to better interpret the trooper-level use-of-race estimates for search decisions, which is the focus of the next chapter.[155] If the average use of race was the primary interest, then the methods above would have been applied to all stops within APA 6 or across the entire WSP. But instead, the methods were only applied to these 16 troopers' stops so that if a trooper is identified as having used race as a factor in his search decisions more than his peers, then the average use-of-race estimate can help interpret that result.

While the search rate results provide evidence that race may have been used as a factor to search a higher share of stopped minority motorists as compared to similarly situated stopped non-minority motorists, the evidence is not conclusive due to the possibility that unobserved factors that affect search decisions differed among racial groups. However, based on these results, it is somewhat unlikely that this group of

[154] If all 151 Asian arrestees are considered, of which 22 tested below 0.08, then if the troopers had arrested 10 fewer of these motorists, then the shares of Asians and non-minority motorists who tested at or above 0.08 would have been the same.

[155] The DUI arrest hit rate analysis in the next chapter does not involve a peer-to-peer comparison, so the average DUI arrest hit rate analysis in this chapter is less needed to inform the results in the next chapter.

troopers on average used race as a factor to search a lower share of stopped minority motorists as compared to similarly situated stopped non-minority motorists. Therefore, based on the analyses in the next chapter, if a trooper is identified as having a minority-to-non-minority search rate ratio that is significantly higher than his peers' ratio, then it is likely that he used race as a factor to search a higher share of stopped minority motorists as compared to stopped non-minority motorists with respect to a race-neutral trooper, that is, a trooper who does not inappropriately use race as a factor in his traffic enforcement decisions. On the other hand, if the above results had shown that race had been used as a factor to search a lower share of stopped minority motorists as compared to similarly situated stopped non-minority motorists, then if a trooper is identified as having a minority-to-non-minority search rate ratio that is significantly higher than his peers' ratio, then it is ambiguous whether he used race as a factor to search a higher share of stopped minority motorists as compared to stopped non-minority motorists with respect to a race-neutral trooper. Instead, as compared to his peers, he may have least used race as a factor in deciding to search a lower share of stopped minority motorists as compared to stopped non-minority motorists. (The above discussion also applies to DUI arrests.) Hence, the search rate results above are sufficiently conclusive in order to inform the interpretation of the trooper-level search rate results in the next chapter.

Chapter 6: Officer-Level Post-Stop Analysis

6.1 Introduction

Although the methods in the previous chapter only focused on estimating the use of race as a factor for 16 troopers' post-stop decisions, the same methods could be applied to estimate the use of race at the department level. The vast majority of racial profiling studies have been done at the department level; however, like the detachment-level analysis in the previous chapter, they have two major drawbacks. First, the data do not include important variables that influence whether a motorist is searched or arrested for a DUI. Second, department-level studies only estimate the average use of race and do not capture potential variation in the use of race among officers. Therefore, even if a department-level study finds no evidence of the use of race, the actions of a few officers who are using race will likely be hidden. Moreover, if a department-level study finds evidence that race is being used, the police leadership is not able to determine whether its use varies among officers. The leadership's corrective action would differ if it knew almost all officers were using race versus a situation where only a few officers were identified as using race.

This chapter estimates each trooper's relative use of race as a factor in post-stop decisions in order to see if particular troopers use race more than their similarly situated peers. As was done in Chapter 5, the search and DUI arrest decisions are analyzed. The trooper-level analysis approach helps mitigate the drawbacks above. First, the unobserved confounding characteristics are likely to be similar across similarly situated troopers; hence, each trooper's relative use of race can be estimated more accurately than the 16 troopers' combined average use of race. Second, a trooper-level analysis can be incorporated into an EI system. If a trooper uses race significantly more than his peers, this analysis will be able to identify him, which could result in further scrutiny within an EI system. However, each trooper's estimated use of race is relative to his peers, and not relative to a race-neutral trooper. Hence, these estimates are more informative when they are used in conjunction with the average use-of-race estimates completed in Chapter 5.

6.2 Data

For the trooper-level search rate analysis, which includes DUI arrests, I use the same observations from Chapter 5's detachment-level search rate analysis. These observations are similar to the observations used in the trooper-level stop analysis in Chapter 4; hence, the figures and tables that described the stop characteristics that were presented in that chapter are not repeated here. For the DUI arrest hit rate analysis, I use the same observations from Chapter 5's detachment-level DUI arrest hit rate analysis.

6.3 Methods

This section describes the empirical method used to estimate each trooper's relative use of race in deciding which stopped motorists are searched or arrested for a DUI. The first part of this section compares each trooper's search and DUI arrest rates among racial groups. Due to the high number of searches stemming from DUI arrests and potential for unobserved confounding characteristics in the search rate analysis, the second part of this section compares the share of each trooper's non-minority DUI arrestees who tested above the legal alcohol concentration limit with the share of his minority DUI arrestees who tested above the legal alcohol concentration limit.

6.3.1 Search Rates (including DUI arrests)

The methods used to estimate each trooper's relative use of race should take into account the degree to which the troopers are similarly situated as well as to allow for the possibility that a trooper may choose to make different types of stops as compared to his peers. As seen in Chapter 4, troopers' schedules shift over time. They patrol different times of the day on different days of the week. The racial-group shares of motorists at risk of being stopped may change across these contexts, and probability of a stopped motorist being searched or arrested for a DUI changes across these contexts (e.g., more DUI arrests occur at night). Moreover, the troopers may make different types of stops that may affect their search and DUI arrest rates (e.g., a trooper may focus on lane violations, which may indicate a motorist driving under the influence). Thus, troopers may search or arrest stopped motorists for a DUI, regardless of their race, at a different average rates.

The propensity score weighting method described in Chapter 3 can be adapted for this application. However, there are two distinct approaches to estimate each trooper's relative use of race in deciding which motorists to search or arrest for a DUI. The first approach is similar to Chapter 5, where weights are applied to non-minority motorists, while the second approach is similar to Chapter 4, where weights are applied to a subject trooper's peers' stops. Each approach is described in turn.

6.3.1.1 Approach I: Search Rates (including DUI arrests)

This approach uses the method in Chapter 5 where stopped non-minority motorists are weighted so the joint distribution of their stop characteristics (***x***) closely match the joint distribution of stopped minority motorist's characteristics. In Chapter 5, this approach was used to evaluate the combined stops of the 16 troopers, but for this approach, each trooper's stops will be separately analyzed. Using the potential outcomes notation from Chapter 5 and the trooper notation from Chapter 4, subject trooper *j*'s ($T_j = 1$) use of race is represented by θ_j in the Eq. (6.1), where $\dot{S}$ indicates whether a search or DUI arrest occurred, and *M* indicates whether the stopped motorist was minority. There are *J* models and each model only includes the subject trooper's stops.[156]

$$\theta_j = \frac{\Pr(\dot{S}_1 \mid M = 1, T_j = 1)}{\Pr(\dot{S}_0 \mid M = 1, T_j = 1)} \qquad (6.1)$$

In Eq. (6.1), θ_j is in the form of a ratio instead of a difference because a ratio is more informative when the search rates vary widely by trooper. The null hypothesis for each trooper is that the probability that a stopped minority motorist is searched is less than or equal to the probability that a stopped non-minority motorist is searched under the same circumstances (i.e., $\theta_j \leq 1$).

---

[156] The outcomes $\dot{S}_1$ and $\dot{S}_0$ are respectively analogous to the outcomes S_1 and S_0 defined in Chapter 3:

$\dot{S}_{1i}$ = 1 if motorist *i* is searched or arrested for a DUI if minority (treatment group)
0 if motorist *i* is neither searched nor arrested for a DUI if minority (treatment group)

$\dot{S}_{0i}$ = 1 if motorist *i* is searched or arrested for a DUI if non-minority (control group)
0 if motorist *i* is neither searched nor arrested for a DUI if non-minority (control group)

The *S*-dot $(\dot{S})$ variable is used to distinguish the search variable from the stop variable (*S*).

$\dot{S}_0$ is not observed when $M = 1$, but can be estimated from the data. To estimate $\Pr(\dot{S}_0 \mid M = 1)$ for each trooper, each trooper's non-minority stops are weighted so their weighted joint distribution of $\boldsymbol{x}$ will closely match the trooper's minority stops' joint distribution of $\boldsymbol{x}$, as shown in the Eq. (6.2), where f is the joint distribution function of $\boldsymbol{x}$.

$$f(\boldsymbol{x} \mid M = 1, T_j = 1) = w_j(\boldsymbol{x}) f(\boldsymbol{x} \mid M = 0, T_j = 1) \tag{6.2}$$

The weights are estimated using the GBM technique described in Chapter 3. The propensity score weighting model includes the same characteristics $\boldsymbol{x}$ that were included in the combined trooper post-stop analysis in Chapter 5, with the exception of the set of binary variables indicating the trooper who made the stop since each model only includes one trooper's stops. $\boldsymbol{x}$ is composed of a vector of contextual characteristics $\boldsymbol{c}$ and a vector of motorist characteristics $\boldsymbol{m}$. $\boldsymbol{c}$ includes a set of binary variables indicating the nine quarters, a binary variable indicating weekday versus weekend, a set of binary variables indicating six four-hour time periods within a day, and a binary variable indicating whether the stop occurred south or north of Sea-Tac Airport. $\boldsymbol{m}$ includes a set of binary variables indicating the primary violation that initiated the stop[157] (speeding, equipment, safety belt, lane, criminal, or other); a binary variable indicating whether the motorist was driving aggressively; a set of binary variables indicating whether the motorist had one, two, or three or more violations; a set of binary variables indicating whether the motorist had zero, one, or two or more criminal more violations; a binary variable indicating the motorist's gender; and a set of binary variables indicating the motorist's age (16-25, 26-35, 36-45, or 46+ years old).

Based on the weights estimated using the GBM technique, each trooper's use of race as a factor in deciding which stopped motorists are searched or arrested for a DUI is estimated using Eq. (6.3).

[157] The primary violation that initiated the stop is recorded as "violation one" in the data.

$$\hat{\theta}_j = \frac{\dfrac{\sum_{i=1}^{N} \dot{S}_i M_i T_{ji}}{\sum_{i=1}^{N} M_i T_{ji}}}{\dfrac{\sum_{i=1}^{N} \dot{S}_i w(\boldsymbol{x}_i)(1-M_i)T_{ji}}{\sum_{i=1}^{N} w(\boldsymbol{x}_i)(1-M_i)T_{ji}}} \tag{6.3}$$

To estimate the p-value of $\hat{\theta}_j$, I use the weights above in the following logistic regression model (Eq. (6.4)), where the p-value of $\hat{\theta}_{1j}$ equals the p-value of $\hat{\theta}_j$.

$$\ln\left(\frac{\Pr(\dot{S}=1 \mid T_j = 1, M)}{1-\Pr(\dot{S}=1 \mid T_j = 1, M)}\right) = \theta_{0j} + \theta_{1j} M \tag{6.4}$$

6.3.1.2 Approach II: Search Rates (including DUI arrests)

The second approach uses a similar method to one presented in Chapter 4, where weights are applied to the subject trooper's peers' stops.[158] As in Chapter 4, the stops of the peer troopers will be weighted so the weighted joint distribution of the characteristics of their stops closely match the joint distribution of the characteristics of the subject trooper's stops. For this application, the matching will be done within minority and non-minority stops, which will be discussed below.

Using the potential outcomes notation from Chapter 4, subject trooper *j*'s ($T_j = 1$) use of race is represented by θ_j in Eq. (6.5).[159]

---

[158] This method is similar to the method used in by Guest (2005) for *Lacy v. Villeneuve*, except he used a logistic regression model.

[159] The outcomes $\dot{S}_1$ and $\dot{S}_0$ are respectively analogous to the outcomes S_1 and S_0 defined in Chapter 4:

$\dot{S}_{1i}$ = 1 if motorist *i* is searched or arrested for a DUI if he was stopped by *Tj* = 1 (treatment group)
0 if motorist *i* is neither searched nor arrested for a DUI if he was stopped by *Tj* = 1(treatment group)

$\dot{S}_{0i}$ = 1 if motorist *i* is searched or arrested for a DUI if he was stopped by *Tj* = 0 (control group)
0 if motorist *i* is neither searched nor arrested for a DUI if he was stopped by *Tj* = 0 (control group)

$$\theta_j = \frac{\dfrac{\Pr(\dot{S}_1 = 1 \mid T_j = 1, M = 1)}{\Pr(\dot{S}_1 = 1 \mid T_j = 1, M = 0)}}{\dfrac{\Pr(\dot{S}_0 = 1 \mid T_j = 1, M = 1)}{\Pr(\dot{S}_0 = 1 \mid T_j = 1, M = 0)}} \tag{6.5}$$

The minority-to-non-minority (M:NM) relative risk of being searched is the search rate for stopped minority motorists divided by the search rate of stopped non-minority motorists. The null hypothesis is that the M:NM relative risk of being searched is less for motorists stopped by the subject trooper as compared to motorists stopped by his peers. Under the null hypothesis, $\theta_j \le 1$ for all *j* troopers.[160] Because θ_j is estimated for each trooper, *J* models are estimated. For example, Trooper 78 is the first subject trooper and the other 15 troopers serve as his peers. Then Trooper 285 becomes the subject trooper and the other 15 troopers serve as his peers, and so on.

$\dot{S}_0$ is not observed when $T_j = 1$, but can be estimated from the data. To estimate $\Pr(\dot{S}_0 \mid T_j = 1)$, the subject trooper's peers' stops are weighted so their weighted joint distribution of $\boldsymbol{x}$ will closely match the subject trooper's stops' joint distribution of $\boldsymbol{x}$. This is shown in Eq. (6.6), where *f* is the joint distribution function of $\boldsymbol{x}$.

$$f(\boldsymbol{x} \mid T_j = 1) \approx w(\boldsymbol{x}) f(\boldsymbol{x} \mid T_j = 0) \tag{6.6}$$

The propensity score model includes the characteristics $\boldsymbol{x}$ from the first approach, but also includes *M*, a binary variable indicating whether the stopped motorist is minority. The justification for including *M* follows. Because I am estimating a trooper's relative use of race as compared to his peers, the model only needs to control for contextual and motorist characteristics that differ among motorists within same racial group who are stopped by the subject trooper as opposed to his peers. To illustrate, assume there are two groups of stopped motorists: one minority and the other non-minority. Assume they are stopped by the subject trooper ($T_j = 1$) who decides which motorists to search, resulting M:NM relative risk of being searched. Now assume these same motorists were instead stopped by the peer troopers ($T_j = 0$), who decide which

[160] A one-sided hypothesis test is used so the troopers can be sorted based on a p-value.

motorists to search. Again, a M:NM relative risk of being searched is calculated. The M:NM relative risks of being searched when $T_j = 1$ and $T_j = 0$ can be compared, and if they differ, this is evidence that the troopers use race differently in determining whom to search. Note that the absolute use of race cannot be estimated using this comparison, only the trooper's relative use race as compared to his peers.

This means that minority motorists stopped by the subject trooper's peers need to be weighted so their weighted joint distribution of characteristics closely match the joint distribution of characteristics of minority motorists stopped by the subject trooper. The same is true for stopped non-minority motorists. However, non-minority motorists stopped by the subject trooper do not have to be weighted so they have similar weighted characteristics to the subject trooper's stopped minority motorists. The same is true for the peers' stopped non-minority and minority motorists.

These ideas are formally expressed in the equations below, where *f* is the joint distribution function of $\boldsymbol{x}$, $a \in \{0,1\}$, and $b \in \{0,1\}$. Eq. (6.7) denotes that the weighted motorists' characteristics in the numerator (denominator) of the denominator of Eq. (6.5) need to closely match the motorists' characteristics in the numerator (denominator) of the numerator of Eq. (6.5).

$$f(\boldsymbol{x} \mid T_j = 1, M = b) \approx w(\boldsymbol{x}) f(\boldsymbol{x} \mid T_j = 0, M = b) \tag{6.7}$$

On the other hand, Eq. (6.8) denotes that the unweighted (weighted) motorists' characteristics in the numerator of the numerator (denominator) of Eq. (6.5) do not need to closely match the motorists' characteristics in the denominator of the numerator (denominator) of Eq. (6.5).

$$f(\boldsymbol{x} \mid T_j = a, M = b) \neq w(\boldsymbol{x}) f(\boldsymbol{x} \mid T_j = a, M \neq b) \tag{6.8}$$

In summary, the weighted characteristics of the motorists stopped by the peer troopers should closely match the characteristics of the motorists stopped by the subject troopers. This should occur across all motorists and also within racial groups since *M* is included in the propensity score model and each regression tree in the GBM model uses four-level interactions. Hence, expression (6.7) should be true. On the other hand, expression (6.8) will likely be true (i.e., the left and right side expressions will not likely be equal) since weights will only be applied to the peer troopers' stops, not the subject trooper's stops.

Based on the weights estimated using the GBM technique, the estimate of subject trooper *j*'s ($T_j = 1$) use of race is represented by $\hat{\theta}_j$ in Eq. (6.9).[161]

$$\hat{\theta}_j = \frac{\left(\dfrac{\dfrac{\sum_{i=1}^{N} \dot{S}_i T_{ji} M_i}{\sum_{i=1}^{N} T_{ji} M_i}}{\dfrac{\sum_{i=1}^{N} \dot{S}_i T_{ji} (1-M_i)}{\sum_{i=1}^{N} T_{ji} (1-M_i)}} \right)}{\left(\dfrac{\dfrac{\sum_{i=1}^{N} \dot{S}_i w_i (1-T_{ji}) M_i}{\sum_{i=1}^{N} w_i (1-T_{ji}) M_i}}{\dfrac{\sum_{i=1}^{N} \dot{S}_i w_i (1-T_{ji})(1-M_i)}{\sum_{i=1}^{N} w_i (1-T_{ji})(1-M_i)}} \right)} \tag{6.9}$$

[161] In Chapters 3, 4 and 5, I used a logistic regression model as a sensitivity analysis to the propensity score weighting model. However, in this application, there is a much greater concern that it may produce biased results, so I am not using it here. Cepeda et al. (2003) run simulations to compare the performances of logistic regression and propensity score models across models that vary the number of successes and the number of variables, where a success is defined by a one for a zero-one indicator dependent variable. They compute the ratio of the number of successes to the number of variables included in each model. Their simulations show that the logistic regression model parameter estimates are biased when this ratio is below seven. In this study, the number of searches per trooper that are analyzed ranges from approximately 25 to 250, with most under 150. In this analysis, a logistic regression model would include between 30 to 35 variables, which results in a ratio of under five for most troopers. Hence, the parameter estimate of each trooper's relative use of race from the logistic regression models is likely to be biased. (Note that their method to estimate and apply propensity scores differs from the method used in this study. They used a logistic regression model to estimate the propensity scores, and based on the propensity score results, divided the observations into five strata. As stated in Chapter 3, the GBM approach results in better propensity scores in terms of prediction error and the use of weights, as opposed to strata, applies the scores more smoothly.)

To estimate the p-value of $\hat{\theta}_j$, I use the weights above in the following logistic regression model (Eq. (6.10)), where the p-value of $\hat{\theta}_{3j}$ equals the p-value of $\hat{\theta}_j$.[162]

$$\ln\left(\frac{\Pr(\dot{S}=1 \mid T_j, M)}{1-\Pr(\dot{S}=1 \mid T_j, M)}\right)=\theta_{0j}+\theta_{1j}T_j+\theta_{2j}M+\theta_{3j}T_jM \tag{6.10}$$

6.3.1.3 Discussion of Approaches I and II

The two approaches presented above have relative advantages and disadvantages. Approach I is an absolute estimate of the use of race by a trooper (i.e., is an estimate that is relative to a race-neutral trooper), while Approach II is a relative estimate of a trooper's use of race, that is, relative to his peers' use of race. Approach I assumes unobserved characteristics (from the analyst's perspective) (e.g., motorist's driving behavior, actions, demeanor, field sobriety test results, and speech patterns) that affect search and DUI arrest decisions are not associated with race. Approach II weakens this assumption. It assumes unobserved characteristics within a racial group that affect search and DUI arrest decisions are not associated with a trooper. This is a weaker assumption since unobserved characteristics within a racial group that affect search and DUI arrest decisions are likely to be less dissimilar as compared to unobserved characteristics among racial groups that affect search and DUI arrest decisions.

However, when troopers search or arrest motorists for a DUI at different average rates (across all racial groups), Approach II introduces an assumption that is not present within Approach I. Assume there are two categories of troopers: high- and low-search-rate troopers. If a motorist is highly suspicious of carrying contraband or being under the influence of alcohol or drugs, assume that both high- and low-search-rate troopers search or arrest this motorist for a DUI; however, if a motorist is only moderately suspicious, then assume that only high-search-rate troopers search or arrest this motorist for a DUI.

[162] When weights are specified, by default, R estimates the standard errors using a robust estimator (i.e., a generalized Huber-Eicker-White sandwich estimator).

Hence, Approach II assumes that the racial-group distribution is the same between highly and moderately suspicious motorists, which may not be the case.

Alternatively, each trooper's estimated use of race from Approach I could be considered as an estimate of his relative use of race instead of his absolute use of race. That is, each trooper's estimated use of race could be compared his peers. Then Approach I's assumption that unobserved characteristics that affect search and DUI arrest decisions are not associated with race can be switched to Approach II's weaker assumption that unobserved characteristics within a racial group that affect search and DUI arrest decisions are not associated with a trooper. However, these unobserved characteristics may differ between highly suspicious and moderately suspicious motorists within a racial group.

In summary, both approaches will be used and compared; however, Approach I is favored because the search rates vary widely among troopers. This limits the credibility of Approach II, which assumes the racial-group distribution is the same between highly and moderately suspicious motorists. Moreover, for Approach I, I will treat each trooper's estimated use of race as a relative estimate as compared to his peers, which weakens the assumption that unobserved characteristics that affect search and DUI arrest decisions are not associated with race.

6.3.2 DUI Arrest Hit Rate Analysis

As discussed in Chapter 2, the three potential sources of racial disparities include racial animosity, statistical discrimination, and confounding characteristics. The search rate method above estimates the combined effect of racial animosity and statistical discrimination by attempting to control for confounding characteristics. In order to mitigate the effect of unobserved confounding characteristics as well as to isolate the effect of racial animosity on DUI arrests, I will use the DUI arrest hit rate method developed in Chapter 5.

The null hypothesis is that each trooper's cost to arrest a motorist for a DUI does not vary by race. This is expressed in the following expression: $[C_j(R) \mid T_j = 1] = C_j$ for all R.[163] When all troopers' DUI arrests were combined in Chapter 5, the empirical test

[163] Note that this does not require C to be the same among troopers.

for the null hypothesis was whether $\Pr(BrAC \geq 0.08 \mid A=1, R)$ was equal among all racial groups. For this application, this test is done for each trooper's DUI arrestees, that is, it tests whether $\Pr(BrAC \geq 0.08 \mid A=1, R) \mid T_j = 1$ is equal across all racial groups for each trooper's DUI arrestees.[164] The hypothesis test is based on the marginal DUI arrestees, that is, the least suspicious motorists who were arrested for a DUI. Therefore, I assume the marginal arrestees are those who tested below 0.12.[165] To increase the number of arrestees and as a sensitivity analysis, I also assume the marginal arrestees are those who tested below 0.15.

6.4 Results

The results of the above methods are presented in two sections. The first section presents the results from the search rate analysis, which includes DUI arrests. The second section presents the results from the DUI arrest hit rate analysis.

6.4.1 Search Rates (including DUI arrests)

Table 6.1 shows each trooper's search rate for each search discretion level, and troopers are sorted by the M:NM search rate ratio. The troopers' combined average search rate is 7.1 percent, but the range varies widely by trooper from 0.1 to 16.8 percent. DUI arrests account for most of the searches and high-discretion searches account for very few of the searches. For low-discretion searches, Trooper 837 accounted for 19 percent of these searches while he only accounted for 6 percent of the stops. For high-discretion searches, Trooper 840 accounted for 71 percent of these searches while he only accounted for 9 percent of the stops. Although the focus of this study is on a trooper's use of race, police leadership can use this data to better understand why the search rates vary so widely for these similarly situated troopers, especially Trooper 840's high-discretion search rate.

[164] The null hypothesis could also be a one-sided hypothesis, which I will use below.
[165] I do not include arrestees who refused to test as marginal arrestees.

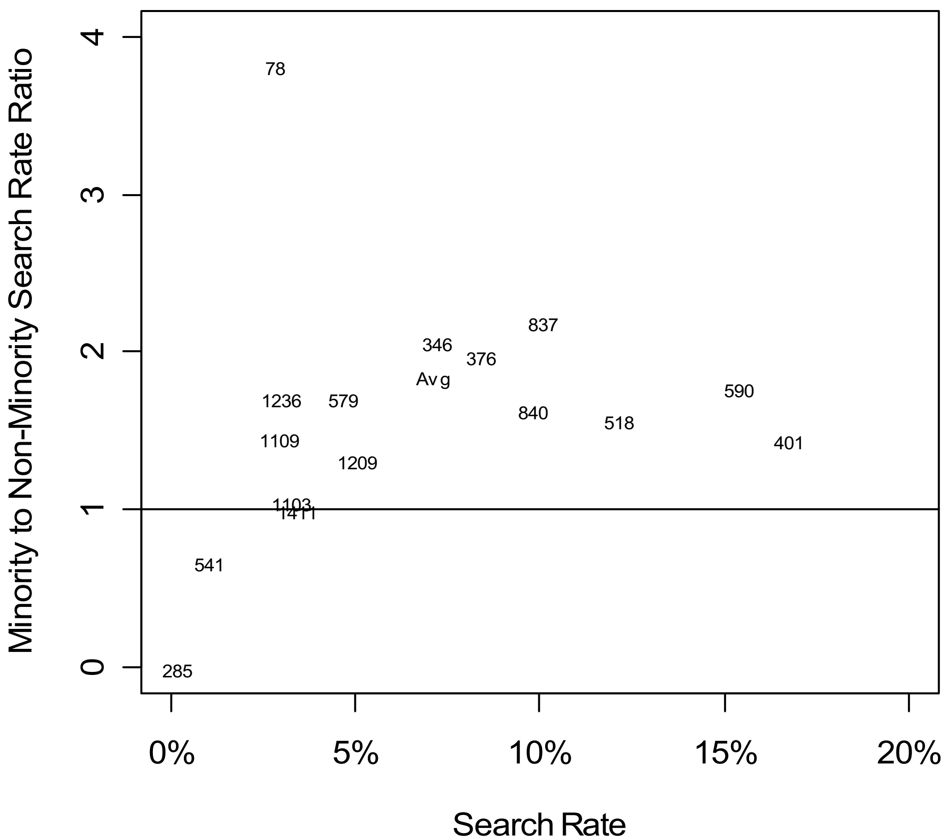
Table 6.1: Search Rates by Search Discretion Level and M:NM Search Rate Ratio by Trooper (all unweighted)[166]

Trooper	Stops	Search Rate	Low-Discretion Search Rate	DUI Arrest Rate	High-Discretion Search Rate	Number of Minorities Searched	Number of Non-Minorities Searched	Minority to Non-Minority Search Rate Ratio
78	1,339	2.8%	1.5%	1.3%	0.0%	25	13	3.8
837	1,219	10.1%	4.8%	4.6%	0.7%	69	54	2.2
346	1,600	7.2%	2.3%	4.4%	0.6%	63	52	2.1
376	1,285	8.4%	1.7%	6.7%	0.0%	58	50	2.0
590	877	15.4%	1.7%	13.7%	0.0%	74	61	1.8
1236	942	3.0%	0.3%	2.7%	0.0%	13	15	1.7
579	1,380	4.6%	1.5%	3.0%	0.1%	32	32	1.7
840	1,701	9.8%	1.4%	5.5%	3.0%	88	79	1.6
518	1,304	12.1%	2.3%	9.8%	0.0%	75	83	1.6
1109	998	2.9%	0.6%	2.3%	0.0%	12	17	1.5
401	1,379	16.8%	2.7%	14.1%	0.0%	129	102	1.4
1209	1,543	5.1%	1.5%	3.5%	0.1%	30	48	1.3
1103	1,227	3.3%	0.9%	2.4%	0.0%	14	26	1.0
1411	875	3.4%	1.1%	2.3%	0.0%	11	19	1.0
541	678	1.0%	0.0%	0.7%	0.3%	2	5	0.7
285	673	0.1%	0.0%	0.1%	0.0%	0	1	0.0
Total	19,020	7.1%	1.7%	5.1%	0.4%	695	657	1.8

Figure 6.1 plots each trooper's M:NM search rate ratio by his search rate, and also plots the 16 troopers' average ratio and rate. (The figure includes both searches and DUI arrests.) The average M:NM search rate ratio is 1.8; however, the rate varies widely by trooper, ranging from 0 to 3.8. When a trooper has a low search rate, the M:NM search rate ratio is not very meaningful since a few additional searches of one racial group can significantly change the ratio. Overall, there does not appear to be a relationship between a trooper's M:NM search rate ratio and his search rate.

[166] All high-discretion searches are frisk searches except for three consent searches.

Figure 6.1: Plot of Each Trooper's Minority:Non-Minority Search Rate Ratio by Search Rate



The results from Approaches I and II are respectively presented next. Due to the low search rates of Troopers 285 and 541, they are not included in the results below. For Approach I, to assess how well the propensity score weights performed in matching the distribution of each trooper's minority stop characteristics with the weighted distribution of each trooper's non-minority stop characteristics, Figure 6.2 summarizes the unweighted and weighted effect size differences for each trooper's model, which totals 14 models. The histogram on the left shows the effect size differences between the characteristics of the subject trooper's minority stops and the unweighted characteristics his non-minority stops, and the histogram on the right shows the effect size differences between the characteristics of the subject trooper's minority stops and the weighted characteristics his non-minority stops. Based the aggregate number of characteristics included within the 14 models, each histogram includes 748 effect size differences.[167] The histogram on the right shows that the weighting eliminates the major differences.

Figure 6.2: Effect Size Comparison between Each Trooper's Minority and Non-Minority Stop Characteristics (unweighted and weighted)

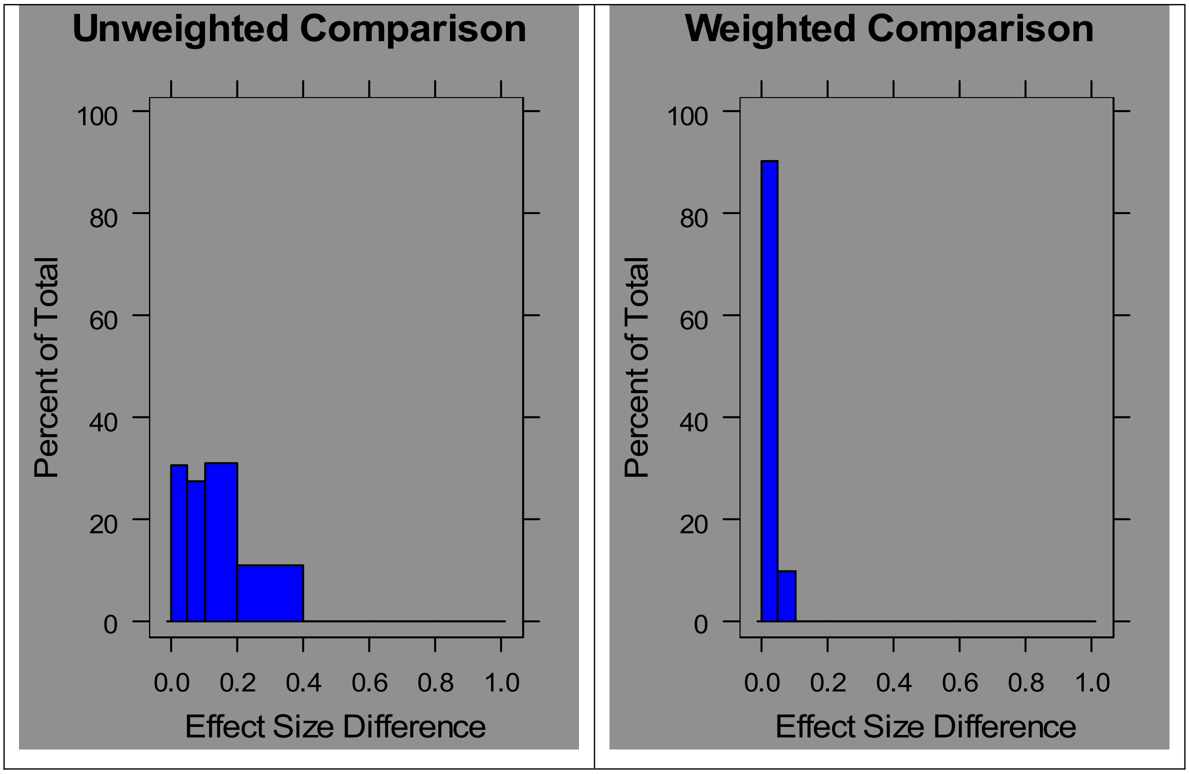


Table 6.2 shows the results of the propensity score weighting model for all searches, low-discretion searches, DUI arrests, and high-discretion searches by trooper. Due to the difficulty of obtaining results with adequate precision for low-search-rate troopers, only troopers with at least 20 minority and 20 non-minority searches within the particular search discretion level are in included in the table.[168] The first set of five

[167] The effects size comparisons between each trooper's minority and non-minority stops are based on the variables described above (which are also listed in Table 5.8; note that the binary indicator variable for each trooper is not included here). If a variable only has two levels (e.g., gender), then one of the effect size differences is dropped since both are the same. For most troopers, this results in 55 comparisons; however, if a trooper had very few or zero stops at a particular level (e.g., during one of the quarters), then those variable levels are not included in his model. This results in 22 fewer effect size comparisons.

[168] Although Trooper 840 only had 19 high-discretion non-minority searches (and 32 high-discretion minority searches), I still analyzed his high-discretion searches due to the high number relative to his peers.

columns includes the primary results, which is followed by a set of five columns that includes additional information. The first set of columns includes the search discretion level, the search rate of stopped minorities, the weighted search rate of stopped non-minorities, the ratio of these rates (which is $\hat{\theta}$ for Approach I, see Eq. 6.3), and the one-sided p-value of $\hat{\theta}$. In the second set of columns, the first column includes the M:NM search rate ratio rank for each trooper, where a rank of 1 indicates the highest M:NM search rate ratio. The next column is the unweighted search rate of stopped non-minority motorists, followed by the number of minority stops, the effective number of non-minority stops, and the total number of stops analyzed.

For each search discretion level, the troopers are sorted based on the one-sided p-value. For all searches, because there are 14 hypothesis tests, the Bonferroni-equivalent p-value for a 0.05 significance level is 0.0036 (or 0.05/14). Based on this standard, the null hypothesis is not rejected for any particular trooper. For Trooper 837's low-discretion searches, the null hypothesis is not rejected. For DUI arrest decisions, because there are eight hypothesis tests, the Bonferroni-equivalent p-value for a 0.05 significance level is 0.0063 (or 0.05/8). Based on this standard, the null hypothesis is rejected for Trooper 837. For Trooper 840's high-discretion searches, the null hypothesis is rejected (one-sided p-value = 0.0089).

Table 6.2: Estimated Use of Race by Trooper for Search Decision (Approach I)

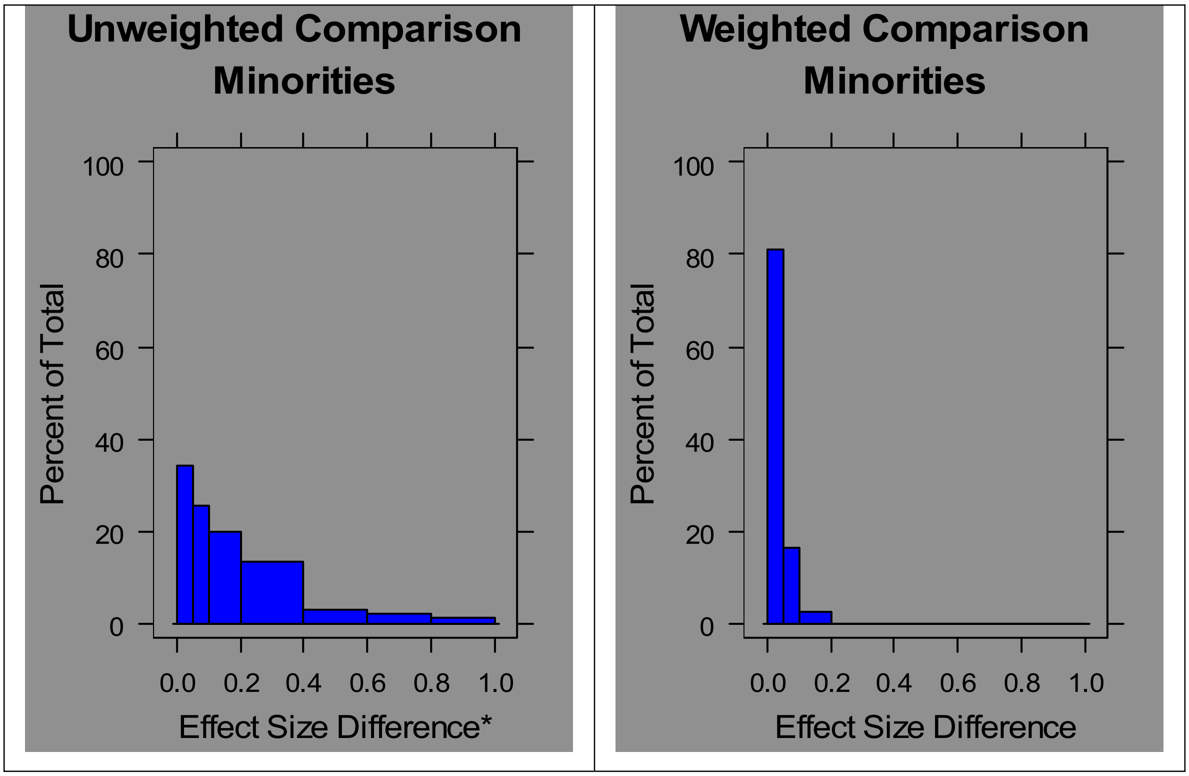
Trooper ID	Search Rate of Stopped Minorities	Search Rate of Stopped Non-Minorities (weighted)	Minority to Non-Minority Search Rate Ratio (weighted)	p-value (one-sided)	Minority to Non-Minority Search Rate Ratio Rank	Search Rate of Stopped Non-Minorities	Number of Minority Stops	Effective Number of Non-Minority Stops	Number of Stops Analyzed
All Searches									
837	15.4%	9.3%	1.7	0.0076	2	7.0%	448	441	1,219
78	5.6%	2.6%	2.1	0.0217	1	1.5%	448	607	1,339
840	12.8%	9.8%	1.3	0.0418	8	7.8%	688	796	1,701
579	6.3%	4.0%	1.6	0.0490	7	3.7%	509	611	1,380
590	20.8%	17.2%	1.2	0.1432	5	11.7%	356	313	877
376	12.2%	9.8%	1.2	0.1534	4	6.2%	475	537	1,283
401	20.1%	18.0%	1.1	0.2007	11	13.9%	643	532	1,379
346	10.7%	9.4%	1.1	0.2433	3	5.2%	591	727	1,600
1236	4.2%	3.5%	1.2	0.3259	6	2.4%	312	461	937
1109	3.7%	3.4%	1.1	0.4108	10	2.5%	326	464	998
1209	6.1%	6.5%	0.9	0.6029	12	4.6%	494	722	1,543
518	15.8%	16.5%	1.0	0.6120	9	10.0%	475	527	1,304
1411	3.4%	4.0%	0.9	0.6603	14	3.4%	320	405	875
1103	3.4%	4.7%	0.7	0.7665	13	3.2%	416	405	1,227
Low-Discretion Searches									
837	6.7%	5.1%	1.3	0.2127	4	3.8%	448	441	1,219
DUI Arrests									
837	7.6%	3.7%	2.1	0.0055*	2	2.9%	448	441	1,219
590	19.7%	15.6%	1.3	0.1130	4	9.6%	356	313	877
376	9.9%	8.5%	1.2	0.2613	3	4.8%	475	537	1,283
401	16.6%	15.8%	1.1	0.3505	11	11.8%	643	532	1,379
840	6.4%	6.1%	1.1	0.4067	12	4.8%	688	796	1,701
346	6.1%	6.2%	1.0	0.5330	5	3.4%	591	727	1,600
1209	4.5%	4.7%	1.0	0.5606	9	3.1%	494	722	1,543
518	12.6%	13.7%	0.9	0.6772	8	8.2%	475	527	1,304
High-Discretion Searches									
840	4.7%	2.3%	2.1	0.0089*	2	1.9%	688	796	1,701

*Significant at the 0.05 Bonferroni-equivalent level (one-sided)

The following results are from Approach II. To assess how well the propensity score weights performed in matching the weighted peers' distribution of stop characteristics with a subject trooper's distribution of stop characteristics, Figure 6.3 summarizes the unweighted and weighted effect size differences of the 14 model's minority stops, and Figure 6.4 summarizes the non-minority stops. In Figure 6.3, the histogram on the left shows the effect size differences between the characteristics of the subject trooper's minority stops and the unweighted characteristics his peers' minority stops, and the histogram on the right shows the effect size differences between the characteristics of the subject trooper's minority stops and the weighted characteristics his peers' minority stops. Based the aggregate number of characteristics included within the 14 models, each histogram includes 748 effect size differences.[169] The histogram on the right shows that the weighting eliminates the major differences.

[169] These comparisons are the same as the ones in Figure 6.2.

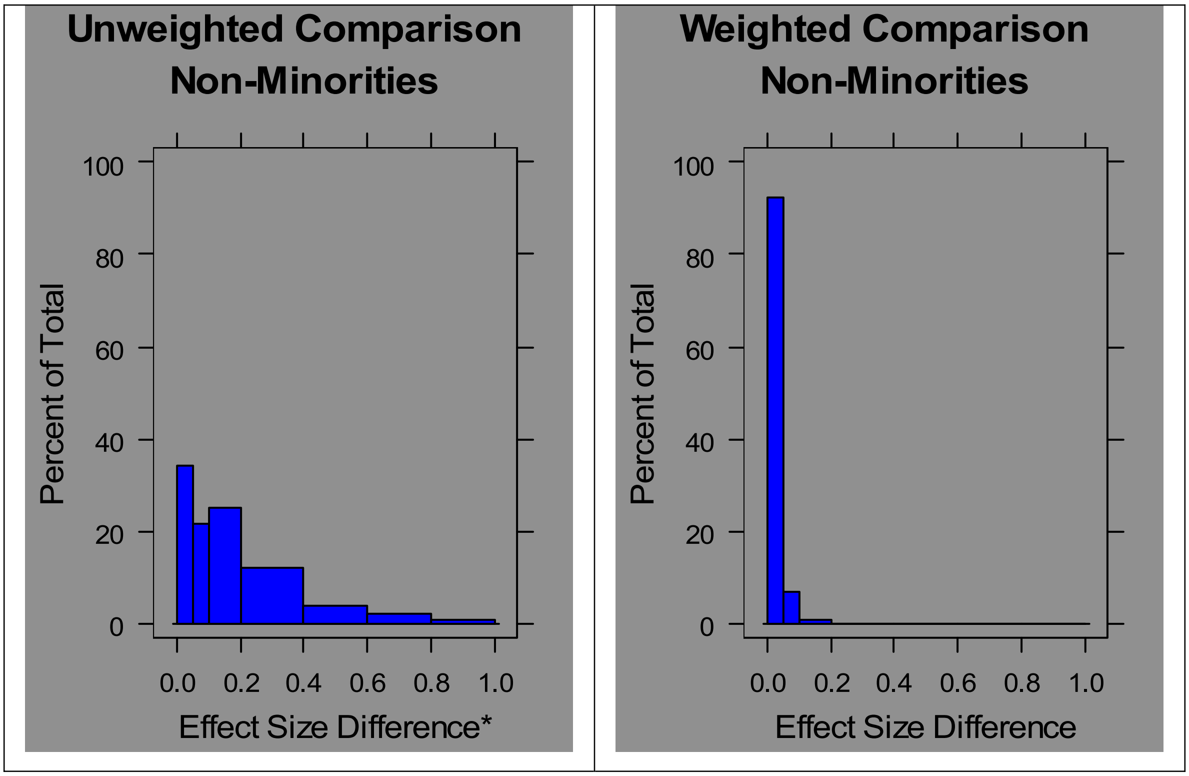
Figure 6.3: Effect Size Comparison between Each Trooper's and His Peers' Minority Stop Characteristics (unweighted and weighted)



*Three of the variable levels had an unweighted effect size difference greater than 1.0, with the greatest being 1.3.

Figure 6.4 includes the analogous histograms to the ones in Figure 6.3 for non-minority stops. Again, the weighting removes the major differences.

Figure 6.4: Effect Size Comparison between Each Trooper's and His Peers' Non-Minority Stop Characteristics (unweighted and weighted)



*Three of the variable levels had an unweighted effect size difference greater than 1.0, with the greatest being 1.4.

Table 6.3 is structured in the same way as Table 6.2. It shows the results of the propensity score weighting model for all searches, low-discretion searches, DUI arrests, and high-discretion searches by trooper. The first set of five columns includes the primary results, which is followed by two columns that include the primary results from Table 6.2, followed by a set of five columns that include additional information. The first set of columns respectively include the search discretion level, the subject trooper's M:NM search rate ratio, his peers' weighted M:NM search rate ratio, the ratio of these ratios (which is $\hat{\theta}_j$ for Approach II, see Eq. 6.9), and the one-sided p-value of $\hat{\theta}_j$. The second set of columns includes the primary results from Table 6.2: the subject trooper's weighted M:NM search rate ratio (which is $\hat{\theta}_j$ for Approach I, see Eq. 6.3) followed by the one-sided p-value of $\hat{\theta}_j$. In the third set of columns, the first column includes the M:NM search rate ratio rank for each trooper, where a rank of 1 indicates the highest

M:NM search rate ratio. The next column is the subject trooper's peers' unweighted M:NM search rate ratio, followed by the number of minority stops, the effective number of peer stops, and the total number of stops analyzed.

For each search discretion level, the troopers are sorted based on the one-sided p-value. For all searches, because there are 14 hypothesis tests, the Bonferroni-equivalent p-value for a 0.05 significance level is 0.0036 (or 0.05/14). Based on this standard, the null hypothesis is not rejected for any particular trooper, which is consistent with the results from Approach I. However, the p-values between the approaches did not always align. For example, when a trooper had a somewhat low p-value using Approach I (e.g., under 0.05), the p-value from Approach II tended to be higher. For Trooper 837's low-discretion searches, the null hypothesis is not rejected, which is consistent with his result using Approach I. For all DUI arrest decisions, because there are eight hypothesis tests, the Bonferroni-equivalent p-value for a 0.05 significance level is 0.0063 (or 0.05/8). Based on this standard, the null hypothesis is not rejected for any particular trooper. Using Approach I, the null hypothesis was rejected for Trooper 837; however, using this approach, the one-sided p-value for Trooper 837 is 0.0870, which is somewhat low, but not statistically significant. For Trooper 840's high-discretion searches, the null hypothesis is not rejected. Using Approach I, the null hypothesis was rejected; however, using this approach, the one-sided p-value for Trooper 837 is 0.1740, which is not significant.

Table 6.3: Estimated Use of Race by Trooper for Search Decision (Approach II)

Trooper ID	Subject Trooper's Minority to Non-Minority Search Rate Ratio	Peers' Minority to Non-Minority Search Rate Ratio (weighted)	Ratio of Subject Trooper's and Peers' M:NM Search Rate Ratios (weighted)	p-value (one-sided)	Approach I: Minority to Non-Minority Search Rate Ratio (weighted)	Approach I: p-value (one-sided)	Minority to Non-Minority Search Rate Ratio Rank of Subject Trooper	Peers' Minority to Non-Minority Search Rate Ratio (unweighted)	Number of Subject Trooper's Stops	Effective Number of Peers' Stops	Total Number of Stops Analyzed
All Searches											
78	3.8	2.0	1.9	0.0325	2.1	0.0217	1	1.8	1,339	6,618	19,020
837	2.2	1.8	1.2	0.1416	1.7	0.0076	2	1.8	1,219	7,243	19,020
376	2.0	1.7	1.1	0.2363	1.2	0.1534	4	1.9	1,283	7,487	17,608
401	1.4	1.6	0.9	0.5847	1.1	0.2007	11	1.8	1,379	4,287	19,020
579	1.7	1.8	0.9	0.6149	1.6	0.0490	7	1.9	1,380	2,324	19,020
840	1.6	1.9	0.9	0.6189	1.3	0.0418	8	1.9	1,701	2,050	19,020
346	2.1	2.3	0.9	0.6891	1.1	0.2433	3	1.8	1,600	8,075	19,020
1236	1.7	2.1	0.8	0.7014	1.2	0.3259	6	1.8	937	3,736	14,146
590	1.8	2.0	0.9	0.7548	1.2	0.1432	5	1.8	877	1,828	19,020
518	1.6	1.8	0.9	0.8019	1.0	0.6120	9	1.9	1,304	6,243	19,020
1109	1.5	2.2	0.7	0.8578	1.1	0.4108	10	1.8	998	3,727	14,146
1411	1.0	1.7	0.6	0.9125	0.9	0.6603	14	1.9	875	4,362	19,020
1209	1.3	2.0	0.7	0.9555	0.9	0.6029	12	1.9	1,543	4,582	19,020
1103	1.0	1.9	0.6	0.9558	0.7	0.7665	13	1.9	1,227	4,412	19,020
Low-Discretion Searches											
837	1.8	1.8	1.0	0.4821	1.3	0.2127	4	1.6	1,219	7,243	19,020
DUI Arrests											
837	2.7	1.8	1.5	0.0870	2.1	0.0055*	2	1.8	1,219	7,243	19,020
376	2.1	1.8	1.1	0.2943	1.2	0.2613	3	1.9	1,283	7,487	17,608
590	2.0	1.9	1.1	0.4021	1.3	0.1130	4	1.8	877	1,828	19,020
401	1.4	1.4	1.0	0.4402	1.1	0.3505	11	1.9	1,379	4,287	19,020
346	1.8	2.1	0.9	0.6896	1.0	0.5330	5	1.9	1,600	8,075	19,020
840	1.3	1.8	0.7	0.7818	1.1	0.4067	12	1.9	1,701	2,050	19,020
518	1.5	1.9	0.8	0.8241	0.9	0.6772	8	1.9	1,304	6,243	19,020
1209	1.5	2.1	0.7	0.9113	1.0	0.5606	9	1.9	1,543	4,582	19,020
High-Discretion Searches											
840	2.5	1.2	2.1	0.1740	2.1	0.0089*	2	2.9	1,701	2,050	19,020

*Significant at the 0.05 Bonferroni-equivalent level (one-sided)

Overall, the results from Approaches I and II are somewhat consistent in ordering the troopers; however, the p-values from Approach II tend to be less statistically significant. Each approach relies on different assumptions; however, as stated above, Approach I is favored because the search rates vary widely among troopers.

6.4.2 DUI Arrest Hit Rate Analysis

Table 6.4 includes three groups of DUI arrestees: all arrestees, arrestees with a BrAC less than 0.15, and arrestees with a BrAC less than 0.12. For each group of arrestees, the table shows each trooper's number of arrestees and minority share of his arrestees. The troopers are sorted based on the minority share of each trooper's arrestees who tested below 0.12, which represent the marginal group of arrestees (i.e., the least suspicious group of motorists who were arrested). The last row represents the combined number of arrests and the combined minority share of arrests for all troopers. For arrestees who tested below 0.12, the average minority share among all troopers is 34 percent, but the shares vary widely by trooper from 10 to 58 percent. However, some of the differences are not statistically significant due to the low DUI arrest count for some of the troopers.

Table 6.4: Number of the 16 Troopers' DUI Arrestees by Racial Group and BrAC

	Arrestees (All)		Arrestees (BrAC < 0.15)		Arrestees (BrAC < 0.12)	
Trooper ID	Number	Minority Share	Number	Minority Share	Number	Minority Share
1209	65	49%	37	51%	24	58%
78	31	32%	15	33%	9	56%
840	108	33%	64	39%	42	43%
376	180	26%	102	31%	63	37%
1109	41	24%	21	33%	11	36%
1236	71	15%	32	25%	22	36%
837	84	27%	54	33%	39	36%
401	306	33%	201	35%	131	36%
590	154	38%	107	35%	80	35%
518	287	25%	168	25%	95	29%
579	63	22%	37	24%	17	29%
1103	65	22%	41	20%	26	27%
1411	55	31%	28	29%	15	27%
346	123	28%	59	22%	31	10%
Total*	1,638	29%	967	31%	605	34%

*Troopers 285 and 541 are not individually included in the table because they had fewer than 10 DUI arrests, but their arrests are included in the total.

Figure 6.5 shows the share of each trooper's DUI arrestees who tested at or above 0.08. The figure includes the same three groups of arrestees included in Table 6.10. For all arrestees, the average share that tested at or above 0.08 was 91.0 percent, but the shares varied widely by trooper, ranging from 81.5 to 96.8 percent. The troopers are sorted based on each trooper's share of marginal arrestees (i.e., those who tested below 0.12) who tested at or above 0.08. The average share was 75.7 percent, but these shares also varied widely by trooper, ranging from 61.3 to 86.3 percent.

Figure 6.5: Share of DUI Arrestees Who Tested At or Above 0.08 by Trooper[170]

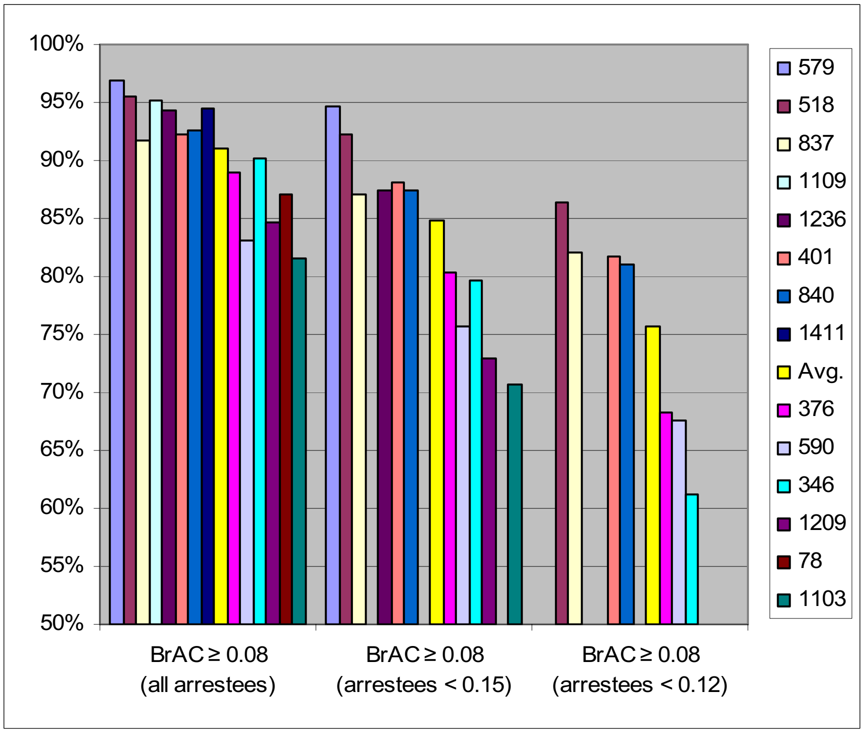


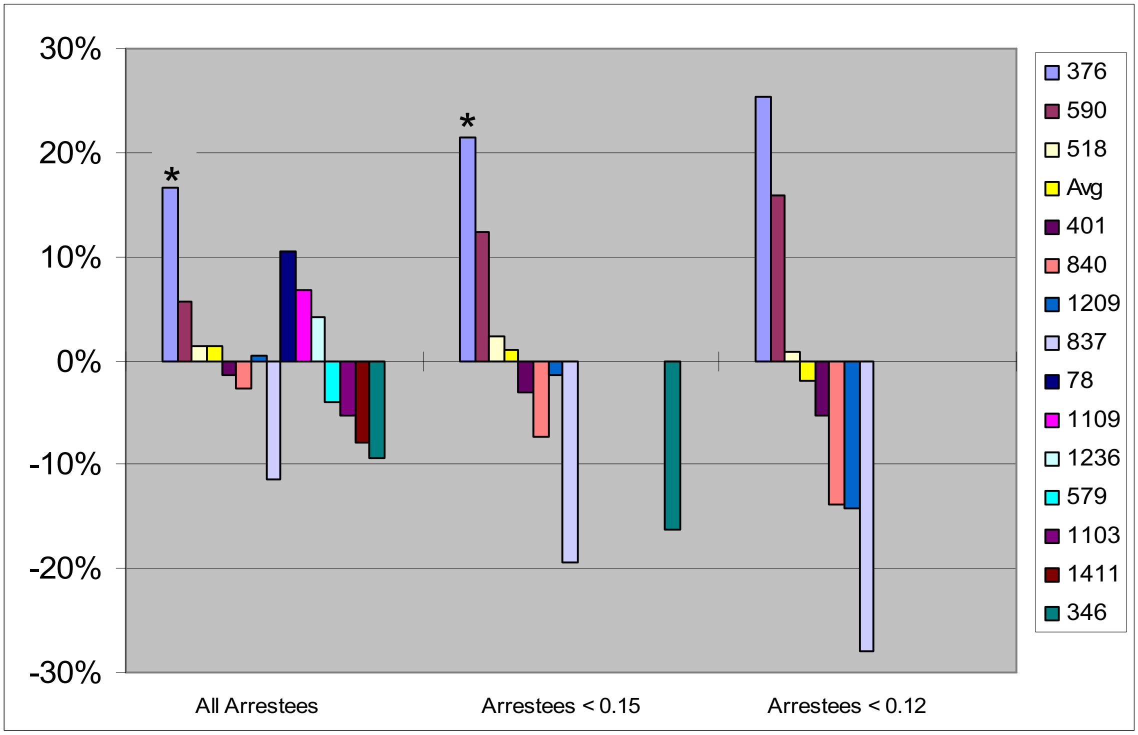
Figure 6.6 shows the share of each trooper's non-minority DUI arrestees who tested at or above 0.08 minus the share of his minority DUI arrestees who tested at or above 0.08. This difference is the trooper's hit rate difference between non-minority and minority DUI arrestees. A positive difference provides evidence that the trooper used a lower probability-of-impairment threshold to arrest minority motorists as compared to the threshold he used to arrest non-minority motorists. The statistical significance of the difference is calculated using a chi square test, resulting in a one-sided p-value.[171] The figure includes the same three groups of arrestees included in Table 6.4 and Figure 6.5. A

[170] A trooper is included in the figure if he had 30 or more DUI arrests for the particular set of arrestees. This threshold is used in order to remove troopers who had relatively few DUI arrests, where the trooper's share of arrestees who tested at or above 0.08 would be estimated with a relatively large standard error. For the results of all arrestees, all troopers listed in the legend are included. For the results of arrestees who tested below 0.15, Troopers 1109, 1411 and 78 are excluded. For the results of arrestees who tested below 0.12, Troopers 579, 1109, 1236, 1411, 1209, 78 and 1103 are excluded. However, the averages include the arrests of all troopers, including the troopers not presented in the figure.
[171] When the number of arrestees who test below 0.08 or at or above 0.08 (for either racial group being compared) decreases to approximately five or fewer arrestees, the chi square test is less reliable; therefore, the one-sided p-value is also calculated using a Fisher exact test.

trooper's difference is excluded if he had fewer than either 10 minority or 10 non-minority DUI arrests. This threshold is used in order to remove troopers who had relatively few DUI arrests, where the trooper's hit rate difference would be estimated with a relatively large standard error. Based on the number of troopers analyzed in each group of arrestees, the Bonferroni-equivalent p-value for a 0.05 significance level is 0.0036, 0.0063, and 0.0071, respectively. The troopers are sorted by the hit rate difference for arrestees who had a BrAC less than 0.12.

Based on the p-value standards above, the null hypothesis is rejected for Trooper 376 when all arrestees or when only arrestees who tested below 0.15 are included; however, the null hypothesis is not rejected when only arrestees who tested below 0.12 are included. Trooper 376's share of all arrestees who tested at or above 0.08 is 88.9 percent; however, the shares of his non-minority and minority arrestees who tested at or above 0.08 are 93.2 and 76.6 percent, respectively, a difference of 16.6 percentage points (one sided p-value 0.0009). When only marginal DUI arrestees (i.e., BrAC < 0.15 or 0.12) are included, then his hit rate differences increase to 21.5 and 25.3 percentage points, respectively. The respective one-sided p-values are 0.0055 and 0.0188, where latter p-value increased due to the smaller sample size. Regarding Trooper 837, he arrested 23 minority motorists for a DUI and each of them tested at or above 0.08; therefore, the non-minority hit rate is less than the minority hit rate for each group of arrestees in the figure.

Figure 6.6: Share of Non-Minority DUI Arrestees Minus Share of Minority DUI Arrestees Who Tested At or Above 0.08 by Trooper[172]



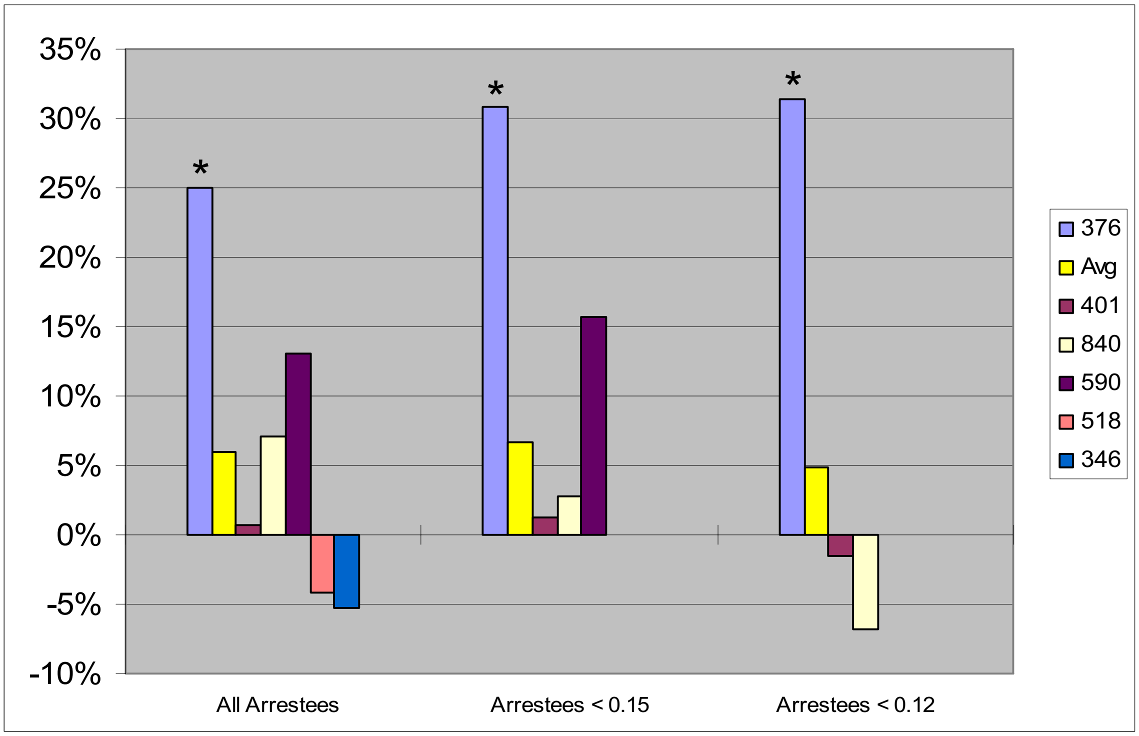
*Significantly higher hit rate for non-minority arrestees as compared to minority arrestees at the 0.05 Bonferroni-equivalent level (one-sided)

Figure 6.7 is the same as Figure 6.6, except that it excludes black arrestees in order compare the hit rates of non-minority and Asian arrestees. This comparison is made due to the results in Chapter 5, which showed that Asians arrestees had a lower share of arrestees who tested at or above 0.08 than non-minority arrestees (although the difference was not statistically significant for arrestees who tested below 0.12). Due to excluding black arrestees, fewer troopers had at least 10 minority DUI arrests; therefore, those troopers were dropped from the analysis. Based on the number of troopers analyzed in each group of arrestees, the Bonferroni equivalent p-value for a 0.05 significance level is 0.0083, 0.0125, and 0.0167, respectively. The troopers are sorted by the hit rate difference for arrestees who had a BrAC less than 0.12. These results are similar to the above results when all minorities were included, where the null hypothesis is rejected for

[172] For the results of all arrestees, all troopers listed in the legend are included. For the results of arrestees who tested below 0.15, Troopers 78, 1109, 1236, 579, 1103 and 1411 are excluded. For the results of arrestees who tested below 0.12, the same six troopers are excluded as well as Trooper 346. However, the averages include the arrests of all troopers, including the troopers not presented in the figure.

Trooper 376 when all arrestees or when only arrestees who tested below 0.15 are included. But in this case, the null hypothesis is also rejected when only arrestees who tested below 0.12 are included.[173]

Figure 6.7: Share of Non-Minority DUI Arrestees Minus Share of Asian DUI Arrestees Who Tested At or Above 0.08 by Trooper[174]



*Significantly higher hit rate for non-minority arrestees as compared to Asian arrestees at the 0.05 Bonferroni-equivalent level (one-sided)[175]

In order to compare the hit rates of non-minority and black arrestees, Asian arrestees are excluded. The results are similar to the results when black arrestees are excluded. The hit rate of Trooper 376's non-minority arrestees exceeded the hit rate of

[173] When all arrestees or when only arrestees who tested below 0.15, the differences are also statistically significant when the one-sided p-value is calculated using the Fisher exact test; however, when only arrestees who tested below 0.12 are included, the one-sided p-value is 0.0392. For consistency, I use the one-sided chi square p-value, which is 0.0162, but since the cell counts approach five (e.g., only six Asian arrestees tested at or above 0.08), the actual one-sided p-value may be slightly higher.

[174] For the results of all arrestees, all troopers listed in the legend are included. For the results of arrestees who tested below 0.15, Troopers 518 and 346 are excluded. For the results of arrestees who tested below 0.12, Troopers 590, 518 and 346 are excluded. However, the averages include the arrests of all troopers, including the troopers not presented in the figure.

[175] As shown in Chapter 5 (but not shown in this figure), when all troopers' arrests are considered, the average hit of Asian arrestees was significantly lower than the average hit rate of non-minority arrestees (see Figure 5.5).

his black arrestees by more than any other trooper. The differences for the three groups of arrestees are 9.2, 12.1, and 17.5 percentage points, respectively; however, the differences are not statistically significant at the 0.05 Bonferroni-equivalent level.

6.5 Discussion

The results from the search rate analysis, which includes DUI arrests, and the results from the DUI arrest hit rate analysis are respectively discussed next. However, when the search rate analysis for a particular trooper focuses on his DUI arrests, the results from the trooper's DUI arrest hit rate analysis is also discussed.

6.5.1 Search Rates (including DUI arrests)

As expected from the detachment-level results in Chapter 5, the results from Approach I presented in Table 6.2 show that stopped minority motorists were searched or arrested for a DUI by many troopers at a higher rate than similarly situated stopped non-minority motorists. When all searches are considered, no particular trooper had a weighted M:NM search rate ratio that significantly differed from one (using the Bonferroni-equivalent p-value for a 0.05 significance level). Due to this uncertainty, the results do not provide conclusive evidence that a particular trooper used race as a factor to search or arrest stopped motorists for a DUI at a higher rate than similarly situated stopped non-minority motorists. Moreover, the models do not include unobserved characteristics that influence the decision to search a motorist that also may be associated with a racial group. Some of these characteristics include a trooper spotting contraband in plain site or a motorist making a furtive movement as the trooper approaches the vehicle. And some impairment indicators include the motorist's driving behavior, field sobriety test results, and the smell of alcohol or drugs. Additionally, these characteristics may also influence whether a motorist is stopped in the first place; hence, stopped motorists may not have the same risk of being searched. While Approach II only assumes unobserved characteristics within a racial group that affect search and DUI arrest decisions are not associated with a trooper, the results do not provide statistically significant evidence that a trooper used race as a factor in his search or DUI arrest decisions that resulted in his M:NM search rate ratio to be higher than his similarly situated peer troopers' M:NM search rate ratio. However, these results may be biased because the approach assumes that the racial-group distribution is the same between highly and moderately suspicious motorists, which may not be the case.

However, when only particular search discretion levels are considered, the results are statistically significant for Trooper 837's DUI arrests and for Trooper 840's high-discretion searches. But when unobserved confounding factors or Trooper 837's DUI arrest hit rate analysis results are considered, then the evidence is inconclusive as to whether these troopers inappropriately used race as a factor in their DUI arrest and high-discretion search decisions, respectively. The results from Approach I provide statistically significant evidence that Trooper 837 may have used race as a factor to arrest minority motorists for a DUI at a higher rate than similarly situated non-minority motorists, and the results from Approach II provide some evidence that Trooper 837 may have used race as a factor that resulted in his M:NM DUI arrest rate ratio to be higher than his similarly situated peer troopers' M:NM DUI arrest rate ratio.[176] However, the data do not include the key indicators used to establish probable cause to arrest a motorist for a DUI such as the motorist's driving behavior, alcohol or drugs in plain view, the smell of alcohol or drugs, and the motorist's demeanor and speech patterns. If these indicators differed among racial groups, then the DUI arrest rates among racial groups are not expected to be the same. The results from Trooper 837's DUI arrest hit rate analysis estimate whether he used a lower probability-of-impairment threshold to arrest minority motorists as compared to the threshold he used to arrest non-minority motorists. However, the results do not provide evidence that he used a lower impairment threshold to arrest minority motorists. In fact, for all three categories of DUI arrestees (all arrestees, arrestees testing below 0.15, and arrestees testing below 0.12), the share of minority arrestees who tested at or above 0.08 was greater than the share of non-minority arrestees who tested at or above 0.08. This is because all 23 minority motorists that he arrested for a DUI tested at or above 0.08.

The results from Approach I provide evidence that Trooper 840 may have used race as a factor to search minority motorists on a high-discretion basis at a higher rate than similarly situated non-minority motorists. However, all of his high-discretion searches that were analyzed were frisk searches, and the data do not include the factors that he used to establish reasonable suspicion that the motorist may have been carrying a weapon that could endanger him or was involved in a crime that had or was about to occur. Therefore, this evidence is not conclusive until more is known about the factors that led him to establish reasonable suspicion for his searches. As stated above, Trooper 840 accounts for almost three-fourths of all the high-discretion searches, which is interesting in and of itself. At first blush, it seems as though Trooper 840 searches too many motorists on a high-discretion basis; however, because his search rate is being compared to his peers (who are not an objective standard), it may also be the case that his peers search too few motorists on a high-discretion basis, thus putting themselves at unnecessary risk of being harmed by a motorist with a weapon. However, this may not be the case because none of Trooper 840's high-discretion searches resulted in finding contraband. Hence, again, it is important to understand the factors that led him to establish reasonable suspicion for his high-discretion searches.

As seen in the sixth column of Table 6.2, the subject trooper's unweighted M:NM search rate ratio rank often predicts the trooper's weighted M:NM search rate ratio rank. However, the analysis also shows that the difference between a subject trooper's search rate of minority motorists can be substantially different when compared to his unweighted versus weighted search rate of non-minority motorists. For example, Trooper 346 ranked 3rd with a large unweighted M:NM search rate ratio of 2.1 (due to a 5.5 percentage point search rate difference between the two racial groups). When his search rate of minority motorists is compared to his weighed search rate of non-minority motorists, the ratio substantially decreases to 1.1 (or a 1.3 percentage point search rate difference between the two racial groups). A relatively simpler analysis using a z-score or t-statistic would not have identified this decrease. The decrease is due to the characteristics of his minority versus his non-minority stops. As compared to the proportions of his non-minority stops, minorities had a higher proportion of their stops at night, for lane violations, involving young motorists. These characteristics are predictive of searches, especially incident-to-arrest searches stemming from a DUI arrest.

---

[176] While the results from Trooper 837's low-discretion searches provide some evidence that he have used race as a factor to search minority motorists on low-discretion basis at a higher rate than similarly situated

Additionally, an EI system will have more confidence in the results from a propensity score weighting or logistic regression model since they attempt to control for differences in the characteristics between a subject trooper's and his peers' stopped motorists.

6.5.2 DUI Arrest Hit Rate Analysis

The results from the DUI arrest hit rate analysis provide evidence that some troopers may have used a lower probability-of-impairment threshold to arrest minorities for a DUI as compared to the threshold used for non-minorities. However, with the exception of Trooper 376, the evidence is not conclusive due to the uncertainty surrounding the estimates, which is primarily due to the small sample size of each trooper's marginal arrestees (i.e., arrestees who tested below 0.12). Based on analyzing Trooper 376's marginal arrestees, the results provide statistically significant evidence that Trooper 376 used a lower probability-of-impairment threshold to arrest Asians as compared to the threshold he used to arrest non-minorities. Because the DUI arrest hit rate analysis is based on an outcome (i.e., the motorist's breath test), the results are less affected by unobserved factors that may be associated with a racial group that also affect DUI arrest decisions. Hence, the DUI arrest hit rate results are favored over results from search rate results. However, impairment due to drugs is one unobserved factor that may bias the DUI arrest hit rate results, which would occur if minority arrestees who tested below 0.08 were impaired due to drugs at a different rate than non-minority arrestees who tested below 0.08.

6.5.3 Summary

The results show that stopped minority motorists were searched or arrested for a DUI by many troopers at a higher rate than similarly situated stopped non-minority motorists. However, either due to the uncertainty surrounding the estimates or due to unobserved confounding factors, the results are not conclusive as to whether they inappropriately used race as a factor in their search and DUI arrest decisions.[177] For

---

non-minority motorists, there is a substantial uncertainty surrounding the estimate.

[177] It turns out that the results from Chapter 5 are not substantially used to interpret the trooper-level results in this chapter because when all searches are considered, no particular trooper has a statistically significant result. When high-discretion searches are considered, Trooper 840 has a statistically significant result showing that he may have used race as a factor to search minority motorists on a high-discretion basis at a higher rate than similarly situated non-minority motorists; however, his result from this chapter is

Trooper 840, the evidence is inconclusive due to potential unobserved confounding factors. The first step of an EI system intervention is to determine whether unobserved factors explain his use-of-race estimate. The results also show that some troopers may have used a lower probability-of-impairment threshold to arrest minorities for a DUI as compared to the threshold used to arrest non-minorities. However, with the exception of Trooper 376, the results are not conclusive due to the uncertainty surrounding the estimates, which is primarily due to the small sample size of each trooper's marginal arrestees. Similar to Trooper 840, the first step of an EI system intervention is to determine whether unobserved factors explain Trooper 376's use-of-race estimate.

The results in this chapter will be combined with the results from the stop analyses in Chapters 3 and 4. When using an EI system, it is important to see if there is a pattern where a particular trooper consistently uses race as a factor more than his peers across traffic enforcement decisions. If the potential problem troopers identified above were also identified as potential problem troopers in the stop analysis, then this would provide additional evidence that these troopers should be further scrutinized. Chapter 7 will summarize the results from Chapters 3 through 6.

---

essentially the same result from Chapter 5 since he made almost three-fourths of the high-discretion searches.

Chapter 7: Summary of Results and Policy Recommendations

The primary objective of this study is to contribute to the racial profiling literature that analyzes traffic enforcement data at the individual officer level. The study discusses the key differences between estimating the use of race as a factor in traffic enforcement decisions at the department versus officer level, and summarizes the key implications of incorporating traffic enforcement racial profiling analyses into an EI system. Second, the study improves upon existing methods to estimate the use of race in stop, search, and DUI arrest decisions at both the department and officer levels. It also assesses how well different empirical methods are able to identify officers who potentially racial profile. The methods are applied to the traffic stops of 16 Washington State Patrol troopers who patrolled South Seattle during 2003 to 2005 in order to estimate the troopers' average use of race as well as each trooper's relative use of race as compared to his 15 peers. This chapter summarizes and discusses the use-of-race estimates, discusses the study's limitations, and concludes with policy recommendations.

7.1 Summary of Results and Study Limitations

7.1.1 Detachment-Level Results

The results from the stop and search rate analyses provide some evidence that the 16 troopers used race as a factor in their traffic enforcement decisions that resulted in stopping and searching (or arresting for a DUI) a higher share of minority motorists. The results from the DUI arrest hit rate analysis provide some evidence that the troopers used a lower impairment threshold to arrest Asian motorists as compared to the threshold used to arrest non-minority motorists. However, due to unobserved factors that may be both associated with these decisions and a racial group, as well as the uncertainty surrounding the use-of-race estimates, there is not sufficient evidence to conclude that these troopers on average inappropriately used race as a factor in these decisions. The results of each of these decisions are further discussed next.

The results from the stop analyses provide some evidence that race was used as a factor to stop a higher share of minority motorists; however, the evidence is not conclusive due to the uncertainty surrounding the estimates. As more of these troopers' traffic stop data are collected and analyzed, those results will provide important evidence as to whether these results are due to chance or are consistent with the troopers using race as a factor to stop a higher share of minority motorists.

However, based on these results, it is somewhat unlikely that this group of troopers on average used race as a factor to stop a lower share of minority motorists. Therefore, if a trooper is identified as having stopped a significantly higher share of minority motorists than his similarly situated peers, then it is likely that he used race as a factor to stop a higher share of minority motorists as compared to a race-neutral trooper, that is, a trooper who does not inappropriately use race as a factor in his traffic enforcement decisions. Hence, the results are sufficiently conclusive in order to inform the interpretation of the trooper-level stop results.

The results from the search rate analysis, which includes DUI arrests, provides evidence that race may have been used as a factor in the troopers' search decisions that resulted in the search rate of stopped minority motorists to be higher than the search rate of similarly situated stopped non-minority motorists. The search rate differences are statistically significant when all searches and when only high-discretion searches are considered. However, there are unobserved factors that influence the decision to search a motorist that may also be associated with a racial group, which would bias these results. For DUI arrests, which trigger the vast majority of the searches, the data do not include the key indicators used to establish probable cause to arrest a motorist for a DUI such as the motorist's driving behavior, field sobriety test results, demeanor, and speech patterns as well as the presence of alcohol or drugs detected by sight or smell. If these indicators differed among racial groups, then the DUI arrest rates among similarly situated racial groups are not expected to be the same. For high-discretion searches, which were mostly frisk searches, the data do not include the factors the trooper used to establish reasonable suspicion that the motorist may have been carrying a weapon that could endanger the trooper or was involved in a crime that had or was about to occur. Therefore, while there is evidence that stopped minority motorists were searched at a higher rate than similarly situated stopped non-minority motorists, due to unobserved factors that affect search decisions that may have differed among racial groups, the evidence is inconclusive whether race was inappropriately used as a factor in search decisions.

However, based on these results, it is somewhat unlikely that this group of troopers on average used race as a factor to search a relatively lower share of stopped minority motorists as compared to similarly situated stopped non-minority motorists. Therefore, if a trooper is identified as having a minority-to-non-minority search rate ratio that is significantly higher than his peers' ratio, then it is likely that he used race as a factor to search a relatively higher share of stopped minority motorists as compared to stopped non-minority motorists with respect to a race-neutral trooper.[178] Hence, these results are sufficiently conclusive in order to inform the interpretation of the trooper-level search rate analyses results.[179]

The results from the DUI arrest hit rate analyses did not provide evidence that a lower probability-of-impairment threshold was used to arrest minority motorists as compared to the threshold used to arrest non-minority motorists. However, when certain subsets of arrestees are compared, there is some evidence that different thresholds were used. These results assume that of the arrestees who tested below 0.08, the shares of those arrestees who were impaired due to drugs (or a combination of alcohol and drugs) did not differ among racial groups, otherwise the results would be biased. When comparing black and non-minority DUI arrestees, the results did not provide any evidence that a lower probability-of-impairment threshold was used to arrest black motorists as compared to the threshold used to arrest non-minority motorists. In fact, when only male arrestees are considered, the results provide evidence that the impairment threshold used to arrest non-minority, male motorists was statistically lower than the threshold used to arrest black,

[178] The same is true if a trooper is identified as having searched stopped minority motorists at a significantly higher rate than similarly situated stopped non-minority motorists. This describes the type of result from the within trooper analysis (i.e., Approach I in Chapter 6). When this approach is used among similarly situated troopers, then the results can be compared among troopers to determine which troopers have the largest relative search rate disparities. Although unobserved factors that affect search decisions may influence the estimated disparity, those unobserved factors should be more similar among similarly situated troopers as compared to non-similarly situated troopers.

[179] It turns out that the results from detachment-level are not substantially used to interpret the trooper-level results because when all searches are considered, no particular trooper has a statistically significant result. When high-discretion searches are considered, Trooper 840 has a statistically significant result showing that he may have used race as a factor to search minority motorists on a high-discretion basis at a higher rate than similarly situated non-minority motorists; however, his result is essentially the same result as the detachment-level analysis since he made almost three-fourths of the high-discretion searches.

male motorists. However, when comparing Asian and non-minority DUI arrestees, the results show that Asians had a lower share of arrestees who tested at or above 0.08; however, when only arrestees who tested below 0.12 are considered, the difference is not statistically significant, due in part to the small sample size. As more of these troopers' DUI arrest data are collected and analyzed, those results will provide important evidence as to whether these results are due to chance or are consistent with the troopers using a relatively lower probability-of-impairment threshold to arrest Asian motorists for a DUI.

7.1.2 Trooper-Level Results

The study found that the use of race significantly varied among troopers for stop, search, and DUI arrest decisions. To summarize each trooper's estimated relative use of race for stop, search, and DUI arrest decisions from Chapters 4 and 6, I calculate the following three hypothetical values for each trooper. A positive value indicates the use of race adversely affected minority motorists, while a negative value indicates the use of race adversely affected non-minority motorists. For the trooper-level stop analysis, I calculate the number of each trooper's minority stops that would need to be replaced by non-minority stops in order to equalize the minority shares of the trooper's stops and his peers' similarly situated stops. The minority share of his peers' similarly situated stops is the weighted minority share of his peers' stops that was estimated in Chapter 4. The calculated number will be positive (negative) if the minority share of a trooper's stops was greater (less) than the minority share of his peers' similarly situated stops. In this calculation, the number of stops made by each trooper remains constant since minority stops are being replaced by non-minority stops, or vice versa. If a trooper changed his use of race as a factor in stop decisions in order to equalize the minority shares of his stops and his peers' similarly situated stops, he would still likely be able to make the same number of stops due to the large number of both minority and non-minority motorists driving on this stretch of Interstate 5 who could be stopped due to violating a traffic law.[180]

[180] An alternative calculation would be to calculate the reduction in the number of each trooper's minority stops needed to equalize the minority shares of the trooper's stops and his peers' similarly situated stops.

For the trooper-level search rate analysis, which includes DUI arrests, I calculate the reduction in the number of each trooper's minority searches needed to equalize the search rates of each trooper's stopped minority motorists and similarly situated stopped non-minority motorists. The search rate of the similarly situated stopped non-minority motorists is the weighted search rate of non-minority motorists estimated in Chapter 6 using Approach I. The calculated reduction will be positive (negative) if the search rate of a trooper's stopped minorities was greater (less) than the search rate of his similarly situated stopped non-minorities. Unless a trooper's search rate of stopped minority motorists equaled his search rate of similarly situated non-minority motorists, then this calculation assumes the trooper changes the number of searches in order to equalize the search rates of his stopped minority motorists and his similarly situated stopped non-minority motorists. Hence, if a trooper's search rate of stopped minorities was greater than his search rate of similarly situated stopped non-minorities, then the rates would be equalized by reducing the number of minorities that he searched (as opposed to increasing the number of non-minorities searched). This methodology was chosen due to anecdotal evidence from minority motorists reporting that they are sometimes searched without probable cause or reasonable suspicion (Meeks, 2000).[181]

For the trooper-level DUI arrest hit rate analysis, if a trooper's hit rate of marginal minority arrestees was less than his hit rate of marginal non-minority arrestees, then I calculate the reduction in the number of that trooper's minority DUI arrestees who tested below 0.08 that would be needed to equalize the hit rates of his marginal minority and non-minority arrestees. If the opposite was true (i.e., a trooper's hit rate of marginal minority arrestees was greater than his hit rate of marginal non-minority arrestees), then I calculate the reduction in the number of that trooper's non-minority DUI arrestees who tested below 0.08 that would be needed to equalize the hit rates of his marginal minority and non-minority arrestees. (In order to have consistent signs across the stop, search rate, and DUI arrest results, I use the negative of this last result.) The calculation is based on the results in Chapter 6 where a marginal arrestee is defined as an arrestee who tested

[181] An alternative calculation would be similar to one used in the stop calculation above, that is, I would calculate the number of each trooper's minority searches that would need to be replaced by non-minority searches in order to equalize the trooper's search rates of stopped minority motorists and similarly situated stopped non-minority motorists.

below 0.12. As in the search rate calculation above, unless a trooper's marginal minority hit rate equaled his marginal non-minority hit rate, then this calculation assumes the trooper changes the number of arrestees in order to equalize the hit rates of his marginal minority and non-minority arrestees. Hence, if a trooper's marginal minority hit rate was less than his marginal non-minority hit rate, then the hit rates would be equalized by reducing the number of minority arrestees who tested below 0.08 (as opposed to increasing the number of non-minority arrestees who tested below 0.08). This methodology was chosen based on assuming that a trooper used an appropriate impairment threshold for the racial group that had the higher hit rate, but used too low of an impairment threshold for the racial group that had the lower hit rate (where the appropriate impairment threshold level is determined by a cost-benefit calculation from society's perspective).[182]

The results of these three calculations are shown in Figure 7.1, and the troopers are sorted based on how many of their results are positive. If a trooper's use of race in a particular traffic stop decision was statistically significant in Chapters 4 and 6 (from Approach I), then that result is marked with an asterisk in the figure. The results represent the change in each trooper's racial composition of stops, searches, and marginal DUI arrests that would be needed to result in a race-neutral outcome, as defined by the calculations above. These calculations assume that each trooper's use-of-race estimates are not biased by unobserved factors. Moreover, these calculations use the use-of-race point estimates, and thus, ignore the uncertainty surrounding these estimates. Although these are significant limitations, they do not significantly affect the figure's usefulness. The purpose of the figure is to summarize each trooper's use-of-race estimates and determine if there is pattern across the different types of traffic enforcement decisions that are analyzed.

The figure shows three interesting points. First, when the results provide statistically significant evidence that a trooper potentially used race as a factor in one type of traffic enforcement decision that adversely affected minority motorists, the results

[182] Alternative calculations could be based on different assumptions such as assuming the opposite (i.e., that a trooper used an appropriate impairment threshold for the racial group that had the lower hit rate, but used too high of an impairment threshold for the racial group that had the higher hit rate) or assuming that the

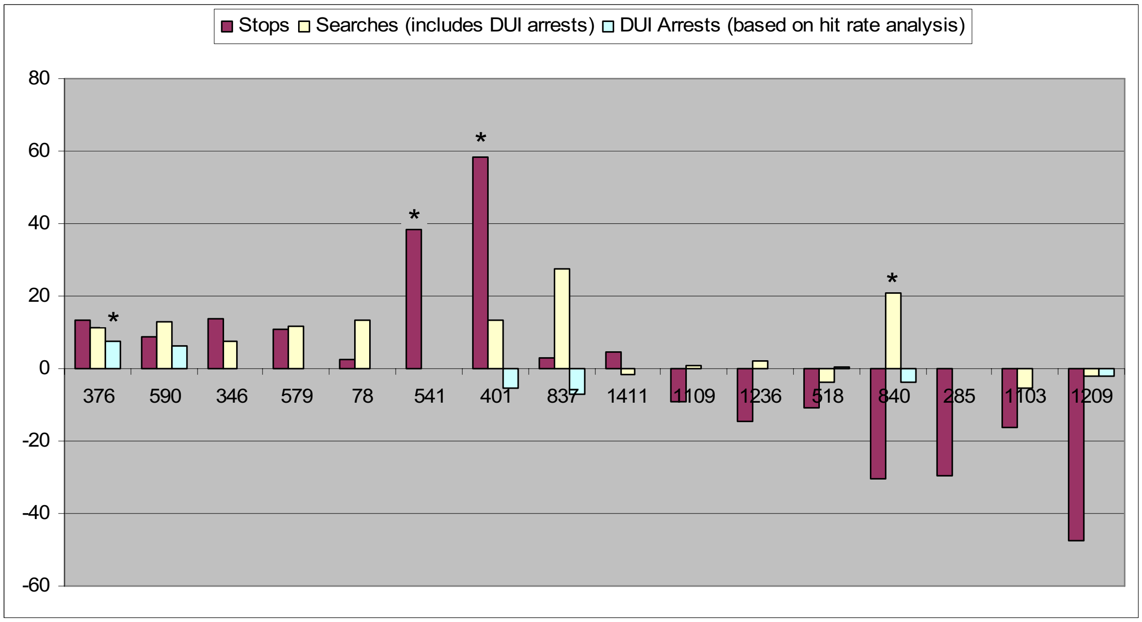
from the other types of traffic enforcement decisions did not provide statistically significant evidence that the trooper potentially used race as a factor that adversely affected minority motorists. However, the results from the other types of traffic enforcement decisions sometimes provide evidence that the trooper potentially used race as a factor that adversely affected minority motorists. Second, although the focus of this study was not to determine whether a trooper potentially used race as a factor to adversely affect non-minority motorists (or perhaps, used race as a factor in order to be more lenient on minority motorists), the results provide some evidence (albeit statistically inconclusive) that this may have been the case in particular traffic enforcement decisions for a few troopers.[183] Third, considering the evaluation period is two years and three months, a trooper's impact on minority motorists includes a somewhat small number of motorists; however, these troopers only represent two percent of the WSP's traffic enforcement officers.

---

appropriate hit rate was in between the hit rates of the two racial groups (e.g., the average of the two hit rates).

[183] If the stop, search, and marginal DUI arrest results are summed across troopers, the sums are negative five stops, 108 searches, and negative four marginal DUI arrests. Each sum will be discussed in turn. First, the stop sum was expected to be near zero because troopers were compared to each other. The result was not precisely zero because each trooper was compared to the other 15 troopers. If each trooper was compared to an average (i.e., an average that included the other 15 troopers as well as himself), then the sum would have been zero. Because this is a peer-to-peer comparison, this sum does not test whether these troopers on averaged used race as a factor in their stop decisions. Only a test that uses an external benchmark (e.g., the radar-based stop test in Chapter 3) addresses that question. Second, the search sum across troopers was not expected to be precisely zero since Approach I in Chapter 6 is not based on comparing the racial compositions of subject trooper's searches with his peers' searches, but instead, is based on comparing a subject trooper's search rate of minority motorists to his search rate of similarly situated non-minority motorists. Based on the aggregate search-rate results in Chapter 5, the search sum across troopers was expected to be well above zero. Third, the DUI arrest sum across troopers was not expected to be precisely zero since the approach compares each trooper's share of marginal minority and non-minority arrestees who tested at or above 0.08. However, based on the aggregate DUI arrest hit rate results in Chapter 5, the marginal DUI arrest sum across troopers was expected to be near zero.

Figure 7.1: Reduction in the Number of Minority Law Enforcement Actions that Would Result in a Race-Neutral Outcome[184]



*Significant at the 0.05 Bonferroni-equivalent level (one-sided)[185]

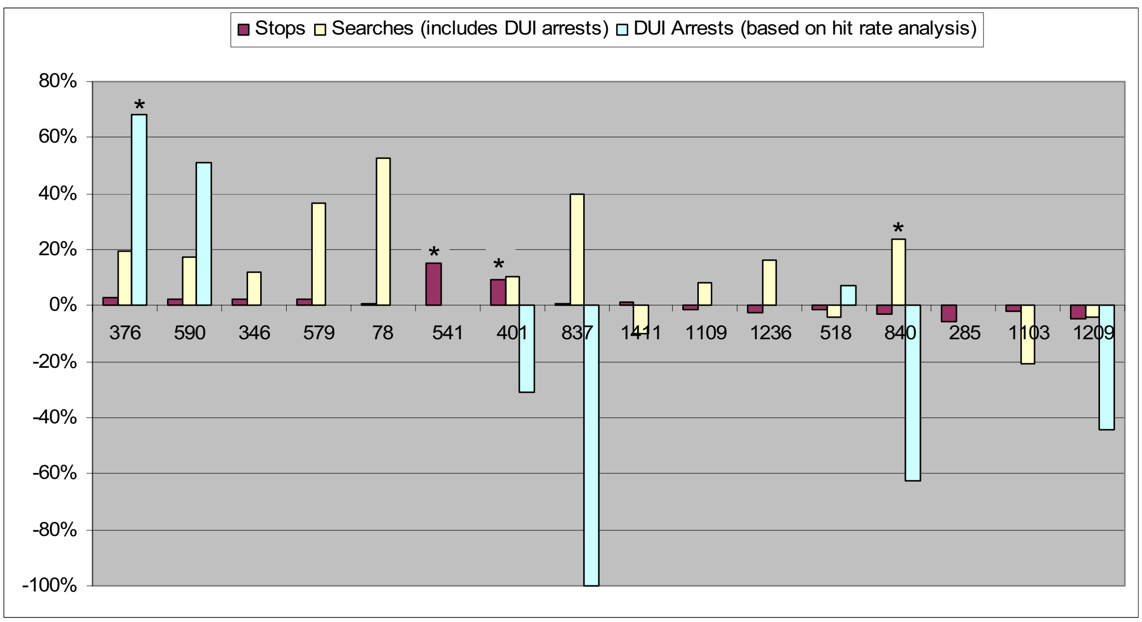
The magnitudes of a trooper's results in Figure 7.1 are likely to be larger for troopers with a relatively higher number stops, searches, and marginal DUI arrests, regardless of the motorist's race. In order to account for these differences among troopers, if a trooper's number of stops, searches, or marginal DUI arrests is positive in Figure 7.1, then that number is divided by the trooper's corresponding number of minority law enforcement actions: the number of minority stops, minority searches, or marginal minority DUI arrestees who tested below 0.08. If a trooper's number of stops, searches, or marginal DUI arrests is negative in Figure 7.1, then that number is divided

[184] As was done in Chapter 6, this figure does not include the search results for Troopers 285 and 541 due to their low number of searches. Similarly, this figure does not include the marginal DUI arrests results if a trooper did not have at least 10 minority and 10 non-minority marginal DUI arrests.

[185] The statistical significance is based on the statistical significance of the results in Chapters 4 and 6 (using Approach I). The statistically significant search rate for Trooper 840 is his high-discretion search rate. As discussed in Chapter 6 regarding Trooper 837, the search rate analysis showed that he had a statistically significant DUI arrest rate; however, his DUI arrest hit rate analysis did not support this. The DUI arrest hit rate test is less affected by unobserved factors that may be associated with minority arrestees that affect DUI arrest decisions; therefore, the result from his search rate analysis is not shown as being statistically significant. For the DUI arrest hit rate analysis, Trooper 376's statistically significant hit rate is based on the hit rate disparity between his non-minority and Asian arrestees (but the hit rate disparity between his non-minority and all minority arrestees was also statistically significant when arrestees who tested below 0.15 were considered).

by the trooper's corresponding number of non-minority law enforcement actions: the number of non-minority stops, non-minority searches, or marginal non-minority DUI arrestees who tested below 0.08. Hence, the results represent the percent reduction in each trooper's minority (or non-minority) stops, searches, and marginal DUI arrestees who tested below 0.08 that would be needed to result in a race-neutral outcome (as defined by the calculations used to generate Figure 7.1). (For illustrative purposes, I report non-minority percent reductions as negative percents.) While these results are informative when a trooper had a large number of stops, searches, or marginal DUI arrests, they are less informative when those numbers are small, which was the case for many troopers who had a small number of searches and marginal DUI arrests.[186]

Figure 7.2: Reduction in the Percent of Minority or Non-Minority Law Enforcement Actions that Would Result in a Race-Neutral Outcome



*Significant at the 0.05 Bonferroni-equivalent level (one-sided)[187]

While the results from the three types of traffic enforcement decisions should be summarized for each trooper in order to determine if a trooper consistently used race as a factor that adversely affected a racial group, a trooper's result within each type of traffic

---

[186] For example, Trooper 837 arrested 23 minority motorists for a DUI and each of them tested at or above 0.08. Therefore, in order to have had his non-minority hit rate also be 100 percent, then he would have had to not arrest the seven non-minority motorists who tested below 0.08. Hence, as shown in the figure, the required percent reduction in the number of non-minority arrestees who tested below 0.08 is 100 percent.
[187] The same footnote from Figure 7.1 applies here.

enforcement decision is important by itself because a trooper may have only used race as a factor within that particular type of traffic enforcement decision. The statistically significant results include Trooper 541's and Trooper 401's stops, Trooper 840's high-discretion searches, and Trooper 376's DUI arrests (for marginal Asian arrestees). Within these results, the magnitude of the impact is important to consider. For example, in order for the minority share of Trooper 541's stops to have equaled the minority share of his peers' similarly situated stops, then 38 (or 15 percent) of his minority stops would have had to have been non-minority stops. In order for the hit rates of Trooper 376's marginal minority and non-minority arrestees to have been equal, he would have had to arrest only 7.5 fewer minority motorists who tested below 0.08; however, this represents 68 percent of his minority arrestees who tested below 0.08.

When the results are significant for a trooper, it must be emphasized that these results identify *potential* problem troopers with respect to their use of race in these decisions. As with the results from an EI system, the trooper-level results from this study are just a starting point to begin an inquiry, because alone, they are not conclusive. This is because there may be extenuating circumstances that may explain a trooper's results (e.g., a trooper temporarily had an assignment that differed from his peers or other unobserved factors exist that explain the differences).

7.1.3 Limitations

The limitations of the study are primarily discussed within each chapter; however, this section discusses other more general limitations. The results only apply to the 16 troopers who were analyzed, not the entire WSP, which had between 700 and 800 traffic enforcement troopers during this study's evaluation period. However, the methods could be used to analyze the traffic enforcement decisions of the other troopers.

The data for a traffic stop observation are generated by the trooper who made the stop. On a form, he records the stop's characteristics, including the motorist's race. Because the data are generated by a trooper, a trooper who uses race as a factor in his traffic enforcement decisions to adversely affect minority motorists may try to conceal his use of race by recording a minority motorist as being non-minority. As with other studies where the officer records the stop characteristics, concerns have been raised that

WSP troopers may have misclassified the motorists' races (Lovrich et al., 2005). To address this concern, Lovrich used two auditors to review the driver's license photos of 812 randomly selected motorists who were stopped during daylight hours and who were recorded as being white by a WSP trooper. Both auditors were at least fairly certain of whether the motorist was white or non-white in 84 percent of the cases, and within these cases, there were few cases (3.7 percent) when both auditors thought the motorist was non-white. These findings show that determining whether a motorist is white or non-white was sometimes difficult for the auditors, and thus, was probably also difficult for WSP troopers. While these findings show a large degree of racial classification consistency between the troopers and auditors, a 3.7 percent difference could be substantial since racial misclassification not only reduces the number of minorities, but also increases the number of non-minorities. To determine whether racial misclassification depends on the motorist's race, the driver's license photos of motorists who were recorded as being non-white would need to be examined. One limitation that Lovrich noted was that the original sample size was reduced due to missing data, but they did not have evidence that the missing data was non-random. Moreover, their analysis was limited to analyzing daylight stops due to the trooper having better visibility to identify the race of the motorist; however, the minority share of stops significantly increases during nighttime hours, so a review of those photos would help answer whether troopers were misclassifying race during those hours. Lastly, while the random sample of stops included stops from within each WSP district, which presumably included some stops from the 16 troopers analyzed in this study, Lovrich's test of racial misclassification only tested the WSP as a whole, not the particular troopers analyzed in this study.

The trooper-level analyses relied on peer comparisons of similarly situated troopers. The analyses identified particular troopers who potentially used race as a factor in their stop, search, and DUI arrest decisions that adversely affected minority motorists. For each of these troopers, an EI system intervention would first try to determine whether extenuating circumstances or unobserved factors might explain a trooper's results. These troopers may have been on special assignment that would result in more contacts with minority motorists even when race was not used as a factor by the trooper. For example, the trooper may have been assisting another law enforcement agency in a gang crackdown whose members included a higher share of minorities as compared to the average share traveling on Interstate 5 at risk of being stopped. This special assignment may explain the trooper's higher minority stop and search rates as compared to his peers.[188] As explanations come to light, if possible, they should be captured in variables that are incorporated into the trooper-level analyses. This is important because if those variables are able to change the status of a potential problem trooper to a non-problem trooper, then incorporating those same variables may have the opposite effect and change the status of a non-problem trooper to a potential problem trooper. The impact would depend on whether those variables significantly varied among troopers and their influence on generating additional stops and searches involving minority motorists. Hence, the troopers who were not identified as potential problem troopers may become potential problem troopers when these additional variables are included in the analyses. Therefore, in general, an EI system should continually evolve in order to improve its accuracy in identifying potential problem troopers.

7.2 Policy Recommendations

The policy recommendations are organized into the following three sections: racial profiling law, EI systems, followed by specific recommendations for the Washington State Patrol. The recommendations should be of interest to law enforcement agencies, legislators, analysts, and community groups interested in racial profiling or EI systems. When developing laws and policies regarding the use of race as a factor in traffic law enforcement decisions (and the non-traffic law enforcement decisions that occur during a traffic stop), legislation and police department racial profiling policies should address the three factors that existing law uses to decide whether and to what degree race may be used as a factor to establish probable cause or reasonable suspicion.

- How important the state's interest is, that is, the importance of the objective that law enforcement is attempting to meet
- How much influence race has in the probable cause decision, ranging from one of many factors to the sole factor

[188] In this case, the decision as to whether the gang crackdown was justified should be evaluated at the department level, not the trooper level.

- How specific the suspect's description is, that is, whether race is part of a specific suspect's physical description, part of a criminal profile, or part of a general statistic concerning offending rates

By addressing these factors, officers as well as the public will be better informed about when the use of race is appropriate versus inappropriate in a traffic enforcement decision (and the non-traffic law enforcement decisions that occur during a traffic stop). In order to reduce racial tension between racial groups and law enforcement, the law should also consider how to compensate innocent individuals who are adversely affected when race is legally used as a factor in law enforcement decisions. A precedent for this type of compensation exists in wrongful conviction indemnification laws (Bernhard, 1999).

Most studies have estimated the use of race at the department level, but there is a growing interest to estimate each officer's use of race and incorporate the results into an EI system designed to identify problem officers and improve officer performance (Walker, 2001; 2003a). A police department's key policy decision is whether to incorporate traffic enforcement data into an EI system in order to estimate each officer's use of race, and ultimately, to identify potential problem officers. This decision will depend on being able to accurately estimate each officer's use of race as well as having officers support its incorporation. Assuming these two criteria are met, then the benefits of its incorporation would have to be weighed against the additional data collection and analysis costs.

As with other department-level studies, this study concludes that the results from a department-level study do not identify whether particular officers are using race. For example, this study's detachment-level analyses provide some evidence (albeit inconclusive) that these troopers may have inappropriately used race as a factor that adversely affected minority motorists. However, without an officer-level analysis, police leadership is unable to identify whether particular troopers are inappropriately using race.

The benefits of incorporating racial profiling data into an EI system is that police leadership will be better able to identify officers who inappropriately use race as a factor in their traffic enforcement decisions. The leadership will then be able to individualize the corrective action and be better able to correct problematic behavior before it becomes too serious. If problems are corrected early on, the police department will have a better relationship with the public and also potentially avoid costly civil rights lawsuits.

Whether a racial profiling study is conducted at the department or officer level, the dominant concern among social scientists engaged in this research centers on developing methods to accurately estimate the use of race in both stop and post-stop decisions. In general, this study found that police departments should use propensity score weighting and multivariate regression techniques in order to identify potential problem officers regarding their use of race since an officer's peer group may change over time and non-racial motorist characteristics may differentially affect an officer's and his peers' traffic enforcement decisions. For these same reasons, these techniques should also be used to identify potential problem officers for other EI system performance indicators.

Regarding specific methods, this study found that if radar-measured speeding stops are going to be used to estimate each racial group's share of motorists at risk of being stopped, then it is important to allow for the possibility that radar may be used at different intensities over the evaluation period and that troopers may use non-racial motorist characteristics during non-radar-based stops to decide which motorists to stop. Second, DUI arrests should be analyzed separately from other incident-to-arrest searches since the decisions leading up to a DUI arrest involve a higher degree of discretion as compared to a typical incident-to-arrest search. Additionally, as compared to a typical search, a DUI arrest is substantially more burdensome on the motorist. The DUI arrest hit rate method shows that it is important to use the marginal arrestees (e.g., arrestees who tested below 0.12 percent) to estimate the use of race in DUI arrest decisions.

For an EI system to be effective, it is important to obtain police union support because the front-line officers and their direct supervisors will ultimately be responsible for its implementation. Officer support will increase if the EI system accurately identifies problem officers and does not falsely identify non-problem officers.

The data collection and analysis costs of an officer-level study will likely be higher than a department-level study. For an officer-level study, the traffic stop data should be supplemented with administrative data that indicates the officer's assignment, location, and time shift so his peers can be easily identified. This is especially important within police departments whose officers do not make as many stops as a state trooper, making it more necessary to compare a particular officer to a set of peers that may change over time. For an officer-level study, the data analysis costs will be higher because multiple peer-group analyses will have to be conducted. These additional costs should be weighed against the benefits of an officer-level study.

If an officer-level racial profiling analysis is incorporated into an EI system, then other performance measures could be incorporated with relatively little additional cost as compared to the additional benefit. The cost of analyzing these additional measures should not significantly increase the overall costs since the officer-level data have already been collected for the officer-level racial profiling analysis. The benefit of analyzing these additional measures is so police leadership will better understand their department's operations and will be better able to identify officers who warrant additional scrutiny. Some of these measures include each officer's average number of stops per shift and each officer's citation, search, and DUI arrest rates.

The following two policy recommendations are for the Washington State Patrol. They are followed by data collection recommendations. First, the WSP should consider incorporating officer-level use-of-race analyses into its Time and Activity Reporting System (TARS). As stated above, this decision will depend on being able to accurately estimate each trooper's use of race as well as having troopers support its incorporation. Assuming these two criteria are met, then the benefits of its incorporation would have to be weighed against the additional data collection and analysis costs. The methods used in this study estimate each trooper's relative use of race, but the accuracy of the estimates is subject to the limitations discussed, which primarily involve unobserved factors that affect a trooper's traffic enforcement decisions that may also be associated with a racial group. When troopers are identified as potential problem troopers with respect to their use of race, the first step of inquiry should try to determine whether there are unobserved factors that explain their use-of-race estimates. If no explanations surface, then these troopers could be referred to interventions such as counseling, mentoring, and training to avoid potential problems from escalating.

If the WSP decides to incorporate racial profiling analyses into an EI system, it is important to obtain the WSP Troopers Association support because traffic enforcement troopers and their direct supervisors will ultimately be responsible for its implementation. Additionally, if trooper support is obtained, the benefits discussed above should be weighed against the additional data collection and analysis costs that will be incurred when racial profiling data is incorporated into an EI system. Because trooper-level traffic stop data is already captured in TARS, the additional data collection costs should be minimal.

The second policy recommendation for the WSP involves its racial profiling policy. The current policy restricts troopers from using race as the sole factor in a traffic enforcement decision. The WSP should consider addressing how the state's interest, the influence of race, and the suspect's specificity affect a trooper's permitted use of race. This is especially important given that the WSP's mission is so broad—from terrorism prevention to drug interdiction to routine traffic enforcement—and the state's interest significantly varies among the WSP's mission areas. By incorporating these factors into the racial profiling policy, WSP troopers as well as the public will be better informed about when the use of race is appropriate versus inappropriate in a traffic enforcement decision (and the non-traffic law enforcement decisions that occur during a traffic stop).

The data collection recommendations focus on how searches are coded within the traffic stop dataset and the type of information that is collected about an arrestee in the BrAC test dataset. The intent of the recommendations is to reduce the number of unobserved confounding factors (or characteristics), that is, contextual and motorist characteristics that are associated with a racial group that also affect the probability of a stopped motorist being searched or arrested for a DUI. In the traffic stop dataset, a search that occurred due to a DUI arrest should be added as a legal basis search category since the events leading to a DUI arrest involve more officer discretion as compared to a typical incident-to-arrest search. Moreover, DUI arrests represent the majority of incident-to-arrest searches. Second, probable cause to search should also be added as a legal basis search category, including sub-categories that define the basis used to establish probable cause. This additional search category will distinguish between searches that are based on probable cause to arrest versus probable cause to search. These two types of probable cause may involve different types of evidence and degrees of discretion. For example, probable cause to arrest is established if a trooper stops a motorist for whom there is an outstanding arrest warrant. If the arrest occurs, the resulting search would be coded as an incident-to-arrest search, which is the current practice. On the other hand, probable cause to search may involve less definitive evidence such as the trooper detecting an odor somewhat resembling the smell of drugs. This type of search would be coded as probable cause to search even if drugs were found and the motorist was arrested. Additionally, within the probable cause to search category, sub-categories that define the basis used to establish probable cause could be included, such as smelling drugs, contraband in plain view, or exigent circumstances. The primary goal of both of these recommendations is to record potential confounding factors, which mostly identify the level of discretion that the trooper had in the search decision.

The DUI arrest data collection recommendations also strive to record potential confounding factors. The only impairment measure in the BrAC test dataset is the breath test result. If an arrestee's alcohol concentration is estimated using something other than his breath (e.g., blood alcohol concentration test), then that measure should also be included in the BrAC test dataset. The most important additional impairment measure to include is whether and to what degree the arrestee was impaired due to drugs. Other additional measures include the results from the SFST, the portable breath test conducted in the field, and other indicators that the trooper used to establish probable cause to arrest the motorist. This information will not only help determine whether the probability-of-impairment threshold used to arrest motorists differed among racial groups, but can also be used to measure the share of all arrestees who tested at or above 0.08 who were arrested solely due to being impaired from alcohol, which could be used to show trends or to compare troopers to each other. Moreover, the information could be used to determine which combination of indicators from the field best predicts a motorist's level of impairment (as measured by an alcohol concentration test). Lastly, so the BrAC test dataset can be easily merged into the traffic stop dataset, each DUI arrestee within the BrAC test dataset should have a unique identifier that can be linked to his information within the traffic stop dataset.

The key policy questions police departments face are whether to analyze traffic enforcement data in order to estimate the use of race at the officer level (in addition to department level), and if so, how to incorporate the data into an EI system to accurately identify potential problem officers. These decisions should focus on creating a system that meets Perez's (1994) principles of an effective reform system, which include integrity, legitimacy, and learning, and the above policy recommendations are a step in that direction.

Bibliography

Alpert, Geoffrey (2004), "Miami-Dade Police Department Racial Profiling Study," The Alpert Group, www.miamidade.gov/irp/report_racial_profiling.asp, accessed July 20, 2006.

Alpert, Geoffrey P., Elizabeth Becker, Alan P. Meister, Michael R. Smith, and Bruce A. Strombom (2005), *Proposed Pedestrian and Motor Vehicle Stop Data Analyses Methodology Report*, Los Angeles, CA: Analysis Group, Inc.

Alpert, Geoffrey P. and Roger G. Dunham (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): 43-65.

Alschuler, Albert W. (2002), "Racial Profiling and the Constitution," *University of Chicago Legal Forum*.

Amnesty International USA (2004), *Threat and Humiliation: Racial Profiling, Domestic Security, and Human Rights in the United States*, New York: Amnesty International USA.

Anwar, Shamena and Hanming Fang (2006), "An Alternative Test Of Racial Prejudice In Motor Vehicle Searches: Theory And Evidence," *The American Economic Review*, 96(1): 127-151.

Arrow, Kenneth J. (1973), "The Theory of Discrimination," in Orley Ashenfelter and Albert Rees, eds. *Discrimination in Labor Markets*. Princeton, NJ: Princeton University Press, 3-33.

Batiste, John R. (2005), Washington State Patrol, July 26 press release.

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

Becker, Gary S. (1968), "Crime and Punishment: An Economic Approach," *The Journal of Political Economy*, 76(2): 169-217.

Bernhard, Adele (1999), "When Justice Fails: Indemnification for Unjust Conviction," *The University of Chicago Law School Roundtable*, 6(73).

Bjerk, David (2006), "Racial Profiling, Statistical Discrimination, and the Effect of a Colorblind Policy on the Crime Rate," June 28, working paper, http://socserv.socsci.mcmaster.ca/bjerk/profilingJPET.pdf, accessed July 29, 2006.

Blincoe, L., A. Seay, E. Zaloshnja, T. Miller, E. Romano, S. Luchter, and R. Spicer (2000), *The Economic Impact of Motor Vehicle Crashes, 2000*, Washington, D.C.: National Highway Traffic Safety Administration.

Carlson, Darren K. (2004), "Racial Profiling Seen as Pervasive, Unjust," The Gallup Organization, July 20.

Cepeda, M. Soledad, Ray Boston, John T. Farrar, and Brian L. Strom (2003), "Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders" *American Journal of Epidemiology*, 158(3): 280-287.

Compton, Wilson M., Bridget F. Grant, James D. Colliver, Meyer D. Glantz, and Frederick S. Stinson (2004), "Prevalence of Marijuana Use Disorders in the United States: 1991-1992 and 2001-2002," *JAMA*, 291(17): 1114-1121.

Davis, Robert C., Christopher W. Ortiz, Nicole J. Henderson, Joel Miller, and Michelle K. Massie (2002), *Turning Necessity Into Virtue: Pittsburgh's Experience with a Federal Consent Decree*, New York: Vera Institute of Justice.

Decker, Scott, and Jeff Rojek (2002), *Saint Louis Metropolitan Police Department Traffic Stop Patterns*. Report submitted by the University of Missouri, St. Louis to the St. Louis Police Department, January. This was an internal report that I did not obtain, but it was referred to in Fridell (2004).

Deheji, Rajeev H. and Sadek Wahba (2002), "Propensity Score-Matching Methods for Non-Experimental Causal Studies," *The Review of Economics and Statistics*, 84(1): 151-161.

Department of Justice, United States (2003) "Fact Sheet: Racial Profiling" March 17, http://www.tsa.gov/interweb/assetlibrary/DOJ_racial_profiling.pdf, accessed February 27, 2006.

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

Farrell, Amy, Jana Rumminger, and Jack McDevitt (2005), *New Challenges in Confronting Racial Profiling in the 21st Century: Learning from Research & Practice*, Boston, MA: Institute on Race and Justice, Northeastern University.

Federal Bureau of Investigation (FBI) (2003), *Age-Specific Arrest Rates and Race-Specific Arrest Rates for Selected Offenses, 1993-2001*, Washington D.C.: U.S. Department of Justice.

Friday, Steve (2002), “Data Collection: Ohio State Highway Patrol Model.” Paper presented at the conference entitled “Bias-Based Policing: Where Are We Now?” sponsored by the Ohio Association of Chiefs of Police, Columbus, Ohio, September 25.

Fridell, Lorie A. (2004), *By The Numbers: A Guide To Analyzing Race Data From Vehicle Stops*, Washington D.C.: Police Executive Research Forum.

Fridell, Lorie, Robert Lunney, Drew Diamond, and Bruce Kubu (2001), *Racially Biased Policing: A Principled Response*, Washington D.C.: Police Executive Research Forum.

Grogger, Jeffrey and Greg Ridgeway (in press), “Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness,” *Journal of the American Statistical Association*.

Gross, Samuel R. and Katherine Y. Barnes (2002), “Racial Profiling and Drug Interdiction on the Highway,” *Michigan Law Review*, 101(651), December.

Guest, Avery Mason (2005), “First Declaration of Avery Mason Guest in Support of Plaintiff’s Response to Defendant’s Motion for Summary Judgment,” *Lacy v. Villeneuve*, United States District Court, Western District of Washington at Seattle, Case No.: CV03-2442JLR, October 28.

Hirano, Keisuke and Guido W. Imbens (2001), “Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization,” *Health Services & Outcomes Research Methodology*, 2: 259-278.

Holmstrom, Bengt and Paul Milgrom (1991), “Multi-task Principle-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design,” *Journal of Law, Economics, and Organization*, Vol. 7 (special issue): 24-52.

Kennedy, Randall (1997), *Race, Crime, and the Law*, New York: Vintage Books.

Klitgaard, Robert and Paul C. Light ed. (2004), *High-Performance Government: Structure, Leadership, Incentives*, Santa Monica, CA: RAND Corporation.

Klitgaard, Robert, Ronald Maclean-Abaroa, and H. Lindsey Parris (2000), *Corrupt Cities: A Practical Guide to Cure and Prevention*, Oakland and Washington, D.C.: ICS Press and World Bank Institute.

Knowles, John, Nicola Persico and Petra Todd (2001), “Racial Bias in Motor Vehicle Searches: Theory and Evidence,” *Journal of Political Economy*, 109(1): 203-229.

Lamberth, John (1994), *Revised Statistical Analysis of the Incidence of Police Stops and Arrests of Black Drivers/Travelers on the New Jersey Turnpike between Exits or*

Interchanges 1 and 3 from the Years 1988 through 1991. Report of defendant's expert in *State v. Pedro Soto*, 734 A. 2d 350 (N.J. Super. Ct. Law Div. 1996).

Logan, Barry K. and Eugene W. Schwilke (1996), "Drug and Alcohol Use in Fatality Injured Drivers in Washington State," *Journal of Forensic Sciences*, 41(3): 505-510.

Loginsky, Pamela B. (2005), *Confessions, Search, Seizure, and Arrest: A Guide for Police Officers and Prosecutors*, Olympia, WA: Washington Association of Prosecuting Attorneys, May.

Lovrich, Nicholas P., Michael J. Gaffney, Clayton Mosher, Mitchell Pickerill, and Travis C. Pratt (2005), *Analysis of Traffic Stop Data Collected by the Washington State Patrol: Assessment of Racial and Ethnic Equity and Bias in Stops, Citations, and Searches Using Multivariate Quantitative and Multi-Method Qualitative Research Techniques*, Pullman: Washington State University, March.

Lovrich, Nicholas P., Michael J. Gaffney, Clayton Mosher, Mitchell Pickerill, and Michael R. Smith (2003), *Washington State Patrol Traffic Stop Data Analysis Project*, Pullman: Washington State University, June 1.

Ludwick, Jack (2003), "Americans See Racial Profiling as Widespread," The Gallup Organization, May 13.

MacDonald, Heather (2003), *Are Cops Racist?* Chicago, IL: Ivan R. Dee.

McCaffrey, Daniel F., Greg Ridgeway, and Andrew R. Morral (2004), "Propensity Score Estimation with Boosted Regression for Evaluating Adolescent Substance Abuse Treatment," *Psychological Methods*, 9(4): 403-425.

McCormick, Erin and Jim Herron Zamora (2001), "Racial Bias in CHP Searches: Latinos, Blacks More Likely to Have Vehicles Examined After Being Pulled Over," *San Francisco Chronicle*, July 15, http://sfgate.com/cgi-bin/article.cgi?file=/chronicle/archive/2001/07/15/MN207903.DTL&type=printable, accessed August 29, 2006.

Meeks, Kenneth (2000), *Driving While Black*, New York: Broadway Books.

Moose, Charles A. (2002), *Montgomery County, Maryland Department of Police: Traffic Stop Data Collection and Analysis, Fourth Report*, December 20.

Moskowitz, Herbert and Dary Fiorentino (2000), *A Review of the Literature on the Effects of Low Doses of Alcohol on Driving-Related Skills*, Encino, CA: Herbert Moskowitz, Ph.D., Inc., http://www.nhtsa.dot.gov/people/injury/research/pub/Hs809028/Title.htm, accessed June 6, 2006.

Moskowitz, H., M. Burns, D. Fiorentino, A. Smiley, and P. Zador (2000), *Driver Characteristics and Impairment at Various BACs*, Los Angeles, CA: Southern California Research Institute, http://www.nhtsa.gov/people/injury/research/pub/impaired_driving/BAC/, accessed June 6, 2006.

National Highway Traffic Safety Administration (2005), *Traffic Safety Facts 2004: A Compilation of Motor Vehicle Crash Data for the Fatality Analysis Reporting System and the General Estimates System*, Washington, D.C.: U.S. Department of Transportation.

National Highway Traffic Safety Administration (2004), *Traffic Safety Facts 2003: A Compilation of Motor Vehicle Crash Data for the Fatality Analysis Reporting System and the General Estimates System*, Washington, D.C.: U.S. Department of Transportation.

Newport, Frank (1999), "Racial Profiling is Seen as Widespread, Particularly Among Young Black Men," The Gallup Organization, December 9.

Nowak, John E. and Ronald D. Rotunda (2000), *Constitutional Law*, 6th ed., St. Paul, MN: West Group.

Office of the Attorney General, New York (1999), *The New York City Police Department's 'Stop & Frisk' Practices: A Report to the People of the State of New York From The Office of the Attorney General*, December 1, http://www.oag.state.ny.us/press/reports/stop_frisk/stop_frisk.html, accessed November 18, 2004.

Pacific Institute for Research and Evaluation, "Impaired Driving in Washington," Calverton, Md: Pacific Institute for Research and Evaluation.

Perez, Douglas W. (1994), *Common Sense About Police Review*, Philadelphia: Temple University Press.

Ramchand, Rajeev, Rosalie Liccardo Pacula, and Martin Y. Iguchi (in press), "Racial Differences in Marijuana-Users' Risk of Arrest in the United States," *Drug and Alcohol Dependence*.

Richardson, Stella (2003), "In Landmark Racial Profiling Settlement, California Highway Patrol Agrees to Major Reforms: *Curtis V. Rodriguez, et al. v. California Highway Patrol, et al.*," American Civil Liberties Union – Northern California, February 27, http://www.aclunc.org/pressrel/030227-chp.html, accessed August 24, 2006.

Ridgeway, Greg (in press), "Assessing the Effect of Race Bias in Post-Traffic Stop Outcomes Using Propensity Scores," *Journal of Quantitative Criminology*.

Ridgeway, Greg (2006), "The GBM Package" http://cran.r-project.org/doc/packages/gbm.pdf, last updated January 21, 2006, accessed February 13, 2006.

Riley, K. Jack, Susan Turner, John MacDonald, Greg Ridgeway, Terry Schell, Jeremy Wilson, Travis L. Dixon, Terry Fain, Dionne Barnes-Proby, and Brent Fulton (2005), *Police-Community Relations in Cincinnati*, Santa Monica, CA: RAND Corporation.

Rosenbaum, Paul R. (1987), "Model-Based Direct Adjustment," *Journal of the American Statistical Association*, 82(398): 387-394.

Rosenbaum, P.R. and Rubin, D.B. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects, *Biometrika*, 70:41-55.

Smith, Michael R. (2005), "Depoliticizing Racial Profiling: Suggestions for the Limited Use and Management of Race in Police Decision-Making," *George Mason University Civil Rights Law Journal*, spring.

Solop, Frederic I. (2002), *Statistical Analysis of I-40 Stop Data and I-40 Violator Study Data from Cocoino County, Arizona*, August 19.

Stuster, Jack (2001), *Development of a Standardized Field Sobriety Test (SFST) Training Management System*, Santa Barbara, CA: Anacapa Sciences, Inc., http://www.nhtsa.dot.gov/people/injury/alcohol/SFST/tech_doc_page.htm, accessed August 9, 2006.

Walker, Samuel (2003a), "Internal Benchmarking for Traffic Stop Data: An Early Intervention System Approach," Discussion Paper, April, www.policeaccountability.org.

Walker, Samuel (2003b), *Early Intervention Systems for Law Enforcement Agencies: A Planning and Management Guide*, Washington, D.C.: U.S. Department of Justice.

Walker, Samuel (2001), "Searching for the Denominator: Problems with Police Traffic Stop Data and an Early Warning System Solution," *Justice Research and Policy*, 3(1): 1-33.

Walker, Samuel and Geoffrey P. Alpert (2000), "Police Accountability and Early Warning Systems: Developing Policies and Programs," *Justice Research and Policy*, 2(2): 59-72.

Walker, Samuel, Stacy O. Milligan, and Anna Berke (2005), *Supervision and Intervention within Early Intervention Systems: A Guide for Law Enforcement Chief Executives*, Washington D.C.: Police Executive Research Forum.

Washington State Patrol and Criminal Justice Training Commission (2001), "Report to the Legislature on Routine Traffic Stop Data," January.

Washington State Patrol (2005a), "Washington State Patrol Quarterly Report, April-June."

Washington State Patrol (2005b), "Washington State Patrol 2004 Annual Report."

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: The MIT Press.