Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness

Jeffrey GROGGER and Greg RIDGEWAY

The key problem in testing for racial profiling in traffic stops is estimating the risk set, or "benchmark," against which to compare the race distribution of stopped drivers. To date, the two most common approaches have been to use residential population data or to conduct traffic surveys in which observers tally the race distribution of drivers at a certain location. It is widely recognized that residential population data provide poor estimates of the population at risk of a traffic stop; at the same time, traffic surveys have limitations and are more costly to carry out than the alternative that we propose herein. In this article we propose a test for racial profiling that does not require explicit, external estimates of the risk set. Rather, our approach makes use of what we call the "veil of darkness" hypothesis, which asserts that police are less likely to know the race of a motorist before making a stop after dark than they are during daylight. If we assume that racial differences in traffic patterns, driving behavior, and exposure to law enforcement do not vary between daylight and darkness, then we can test for racial profiling by comparing the race distribution of stops made during daylight to the race distribution of stops made after dark. We propose a means of weakening this assumption by restricting the sample to stops made during the evening hours and controlling for clock time while estimating daylight/darkness contrasts in the race distribution of stopped drivers. We provide conditions under which our estimates are robust to a substantial nonreporting problem present in our data and in many other studies of racial profiling. We propose an approach to assess the sensitivity of our results to departures from our maintained assumptions. Finally, we apply our method to data from Oakland, California and find that in this example the data yield little evidence of racial profiling in traffic stops.

KEY WORDS: Benchmarking; Racial profiling.

1. INTRODUCTION

Racial profiling is a significant social problem. Some 42% of African-Americans say that police have stopped them just because of their race, 59% of the U.S. public believes that the practice is widespread, and 81% disapprove of it (Gallup 1999).

Public concern over racial profiling has resulted in massive, costly data collection. At least 26 states have passed legislation to deal with racial profiling and require all agencies to collect race data for all traffic stops (Northeastern University 2005). Another 110 agencies in states without mandatory data collection have implemented their own data collection programs. Some collect such data voluntarily, whereas others, such as the Cincinnati and Los Angeles Police Departments, collect data on an ongoing basis as a result of legal settlements.

Despite all of the data collection, there remains considerable uncertainty as to how those data should be used to test for racial profiling. Many researchers suggest that a difference between the racial distribution of persons stopped by police and the racial distribution of the population at risk of being stopped would constitute evidence of the existence of racial profiling (San Jose Police Department 2002; Kadane and Terrin 1997; Smith and Alpert 2002; MacDonald 2001; Dominitz 2003; General Accounting Office 2000; Zingraff et al. 2000). This implicit definition reveals the key empirical problem in testing for racial profiling: measuring the risk set, or the "benchmark," against which to compare the racial distribution of traffic stops.

Measuring the risk set explicitly poses a number of problems. First, the race distribution of drivers within a jurisdiction may differ from the race distribution of the residential population, because car ownership and travel patterns may vary by race. They also may differ because part of the driving population originates outside of the jurisdiction. Furthermore, the race distribution of the at-risk population may differ even from that of the driving population if drivers of different races differ in their driving behavior, that is, if they commit traffic offenses at different rates. Finally, the at-risk population may vary due to differences in exposure to police, even when controlling for driving behavior.

The benchmarking problem has generally been dealt with in one of three ways: Analysts have used benchmarks based on residential populations or driver's license records, despite their limitations; have conducted traffic surveys, using observers to tally the race distribution of drivers or traffic violators at a certain location; or have ignored data on stops altogether, looking for racial disparities in other measures of police behavior. We discuss these approaches in more detail (see also Fridell 2004).

Our main goal in this article is to propose an alternative approach to testing for racial profiling in traffic stops that does not require explicit external estimates of the race distribution of the population at risk of being stopped. An important advantage of our approach is that it is inexpensive to implement, even on the ongoing basis often required by court settlements, because the benchmark that we propose can be constructed from traffic stop data themselves. We present the assumptions under which our approach yields a valid test, discuss how some of those assumptions may be relaxed, and provide some calculations to assess the sensitivity of the test to violations of those assumptions.

Our approach is based on a simple assumption: During the night, police have greater difficulty observing the race of a suspect before they actually make a stop. We refer to this as the "veil of darkness" hypothesis. The implication of the veil of darkness hypothesis is that the race distribution of drivers stopped during the day should differ from the race distribution of drivers stopped at night if officers engage in racial profiling. Thus if travel patterns, driving behavior, and exposure to police are similar between night and day, then we can test for racial profiling by comparing the race distribution of drivers stopped during the day to the race distribution of drivers stopped during the night.

Jeffrey Grogger is Irving Harris Professor in Urban Policy, Harris School, University of Chicago, Chicago, IL 60637 (E-mail: *jgrogger@uchicago.edu*). Greg Ridgeway is Statistician at RAND, Santa Monica, CA 90407-2138 (E-mail: *gregr@rand.org*). The authors thank Ronald Davis, the Oakland Racial Profiling Task Force, and an anonymous referee for their invaluable input.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Applications and Case Studies
DOI 10.1198/016214506000000168

The assumption that travel patterns are similar in the day and the night may be restrictive, because the time of employment is known to vary by race (Hamermesh 1996). To deal with this issue, we make use of natural variation in hours of daylight over the year. In the winter, it is dark by early evening, whereas in the summer it stays light much later. Limiting much of our analysis to stops occurring during the intertwilight period (i.e. between roughly 5 and 9 PM), we can test for differences in the race distribution of traffic stops between night and day, while controlling implicitly for racial variation in travel patterns by time of day. As we argue, limiting the sample period and using time-of-day controls may also equalize differences in risk arising due to differences in driving behavior and police exposure. Neighborhood controls may equalize any differences that remain.

In the next section we provide more detail on previous analyses of racial profiling. In Section 3 we discuss our data, and in Section 4 we formalize and extend our analytical approach. One important extension deals with a serious nonreporting problem that is common in the literature. We present the assumptions under which our approach yields valid qualitative tests. In Section 5 we present our main results based on traffic stop data from Oakland, California. We follow our main results with a sensitivity analysis that helps quantify the extent by which some of our assumptions would have to fail to reverse our qualitative conclusions. We conclude with a discussion of limitations and potential extensions of our approach in Section 6.

2. PREVIOUS RESEARCH ON RACIAL PROFILING

Our aim is to determine whether Oakland patrol officers engage in racial profiling when selecting particular vehicles to stop. Our notion of racial profiling derives from the definition used in the California Peace Officer Standards & Training (POST) program on racial profiling: "The 14th Amendment is also violated when law enforcement officers use a person's race as a factor in forming suspicion of an individual, unless race was provided as a specific descriptor of a specific person in a specific crime" (Peace Officer Standards & Training Program 2002, p. 2). California's definition of racial profiling is similar to that of the U.S. Justice Department, which intervenes in many racial profiling cases (Ramirez, McDevitt, and Farrell 2000).

This notion of racial profiling should be viewed as distinct from a practice that can be termed "neighborhood profiling," in which police commanders deploy patrol officers to minority neighborhoods in greater proportion than warranted on the basis of legitimate law enforcement objectives. Although a few studies have analyzed the spatial distribution of police patrols, the extent of neighborhood profiling per se appears to have received little if any study (Klinger 1997; Alpert and Dunham 1998). Most studies of racial profiling, like ours, seek to determine whether patrol officers are more likely to stop minority drivers than white drivers from the at-risk population.

To estimate the race distribution of the at-risk population, several studies have used secondary data. A number have used census-based estimates of the race distribution of residential populations (e.g., Steward 2004; Weiss and Grumet-Morris 2005). This approach has serious limitations that have been recognized by both researchers and the courts (San Jose Police Department 2002; Dominitz 2003; Smith and Alpert 2002; Chavez v. Illinois State Police). As mentioned earlier, out-of-area drivers and differences in car ownership and travel patterns may result in differences between the residential population and the at-risk population. Furthermore, if there are racial differences in driving behavior, then the racial distribution of the at-risk population may differ from the racial distribution of the driving population, because the U.S. Supreme Court has upheld the legality of traffic stops made pursuant even to trivial violations of the law (Harris 1999). Finally, differences in police exposure can cause differences between residential and at-risk populations. Police argue that they deploy patrols in neighborhoods in proportion to calls for service. Because in many communities a disproportionate number of calls for service come from minority neighborhoods, minority neighborhoods have a greater law enforcement presence. As a result, police may observe minority drivers more frequently (McMahon, Garner, Davis, and Kraus 2002; San Jose Police Department 2002).

Given the limitations of census data, several analysts have used other sources of secondary data. Zingraff et al. (2000) used the race distribution of licensed drivers rather than the residential population to estimate the race distribution of drivers at risk of being stopped. Although this approach accounts for racial differences in the rate at which the population holds driver's licenses, it does not account for out-of-jurisdiction drivers or for potential racial differences in travel patterns, driving behavior, or exposure to police. Alpert, Smith, and Dunham (2003) used data on the location of traffic accidents and the race of the not-at-fault drivers to estimate the race distribution of the at-risk population. Although this approach may measure the race distribution of drivers on the road, it does not account for potential racial differences in driving behavior. Other analysts have studied the race distribution of drivers flagged by photographic stoplight enforcement (Montgomery County Police Department 2002) and by aerial patrols (McConnell and Scheidegger 2001). Again, although these methods may provide reasonable estimates of the race distribution of the driving population, one can question whether they capture race differences in other aspects of stop risk, such as driving behavior and police exposure.

An alternative to using secondary data to estimate the race distribution of the at-risk population is to collect primary data through traffic surveys. Such surveys use observers to tally the race distribution of drivers and in some cases the race distribution of drivers committing certain traffic offenses. For example, Lamberth (1994) used observers to estimate the race distribution of all drivers and of drivers exceeding the speed limit by at least 5 mph on a stretch of the New Jersey Turnpike where motorists had lodged allegations of racial profiling against police.

The advantage of traffic surveys is that they provide plausibly valid estimates of the race distribution of drivers at a specific set of locations. However, traffic surveys have disadvantages as well. The first is their expense. By one estimate, carrying out such a survey requires 800 person-hours of labor (Pritchard 2001). Another problem is that the surveys' validity may suffer in multiethnic environments, where the ethnicity of a driver may be difficult to discern with precision during an observation period that may last only a few seconds. Finally, traffic surveys generally measure only a limited set of traffic offenses, which may influence estimates of racial differences in driving behavior. For example, Lamberth (1994) reported that virtually all drivers, regardless of race, exceeded the speed limit by at least 5 mph. However, in a separate traffic study conducted on the same stretch of the New Jersey Turnpike that Lamberth studied, Lange, Blackman, and Johnson (2001) found that black drivers were more likely than non-blacks to exceed speeds of 80 mph. Thus the extent to which traffic surveys capture racial differences in driving behavior depends on the specific traffic offenses tallied by the survey.

A final vein of research has ignored traffic stop data altogether, focusing on other measures of police behavior, such as the rate at which stopped drivers are searched or the rate at which searches yield contraband, referred to as the "hit rate." For example, Ridgeway (2006) used a propensity score technique to assess differences in stop duration, citation rates, and search rates. A practical virtue of focusing on poststop outcomes is that the risk sets are readily measured; the population at risk of being searched consists of drivers who are stopped, and the population at risk of being found with contraband consists of drivers who are searched. Beyond mere practicality, the emphasis on hit rates stems from an economic model of police behavior. Knowles, Persico, and Todd (2001) showed that in an environment in which police seek to maximize arrests, the equality of hit rates by race implies that police do not intentionally discriminate. However, the model implicitly assumes that police place no weight on the rate at which innocent motorists are detained. In contrast, much of American criminal law (starting with the Fourth Amendment) stresses the protection of the rights of the innocent. Because the rate at which innocents are wrongly detained is a function of the stop rate (Dominitz 2003), analyses that exclude stop rates omit this important consideration.

Our aim in this article is to assess whether there is race bias in traffic stops. In the next section we discuss the stop data to which we apply the approach that we spell out in Section 4.

3. OAKLAND'S TRAFFIC STOP DATA

The genesis for the data that we analyze were complaints by motorists and advocates that the Oakland Police Department (OPD) had engaged in racial profiling, discriminating in particular against black drivers (Oakland Police Department 2004). An early analysis of the OPD's stop data using the census benchmark method indicated that 56% of drivers stopped by the OPD were black, whereas blacks composed only 35% of the city's residential population. Although OPD started collecting stop data voluntarily, it later entered a settlement agreement with the U.S. Justice Department requiring that they collect such data on an ongoing basis (Allen et al. v. City of Oakland et al. 2003, sec. VI.B). Similar to the consent decrees involving other police departments, the Oakland litigation required regular monitoring of the stop data so as to detect trends in potentially discriminatory police behavior.

Under the terms of the agreement, Oakland police must record information on every stop that they initiate anywhere within the city limits of Oakland. Note that this implicitly excludes freeway stops, because freeways fall under the jurisdiction of the California Highway Patrol. Police officers must complete a report including items such as the reason for the stop, the time and location of the stop, and the race/ethnicity of the person stopped. These data are then entered into an electronic database, which the OPD made available for our analysis. Here we focus on motor vehicle stops.

The data that we analyzed included all reported vehicle stops carried out between June 15 and December 30, 2003, amounting to a total of 7,607 stops. Officers most frequently stop vehicles for nondangerous moving violations (48%) and dangerous moving violations (27%), although the danger distinction is subjective. Mechanical and registration violations were the reason for most of the remaining stops (20%), but some drivers were also stopped for criminal investigations (5%).

Vehicle stops are concentrated in the city's downtown (28%) and an area known as the Flatlands (25%). The Flatlands, in which 80% of the residents are black, is Oakland's high-crime area. The area contributes disproportionately to Oakland's homicide rate, which at 28 homicides per 100,000 residents in 2003 was more than 4 times the national average and greater than the homicide rates of Los Angeles and Chicago. Only 5% of the OPD's stops occur in the low crime, affluent Oakland hills, a predominantly white and Asian community.

Despite the terms of the court settlement, there is evidence of a substantial nonreporting problem in the data. An audit of the stop reports led the OPD's Independent Monitoring Team to estimate that as many as 70% of all motor vehicle stops were not reported in the early phases of this study (Burges, Evans, Gruber, and Lopez 2004, p. 41). Court-ordered oversight and increased sanctions for noncompliance raised the number of completed stop forms, especially in October and November.

Such sizeable nonreporting problems seem fairly common in the literature. Kadane and Terrin (1997) noted that either race data were missing or no report was available for about 69% of the drivers stopped during the course of data collection for Lamberth's (1994) New Jersey Turnpike study. The General Accounting Office (2000) reported that the driver's race was missing from about 50% of the stops carried out during a racial profiling study in Philadelphia; Smith and Alpert (2002) reported that data were missing for 36% of the stops made in the course of a Richmond, Virginia study; and Steward (2004) reported that 34% of Texas law enforcement agencies failed to collect stop data mandated by recent state legislation.

Clearly, nonreporting problems are an issue that must be considered in testing for racial profiling. In the next section we provide conditions under which the veil of darkness approach yields valid tests despite the presence of substantial nonreporting. These conditions are weaker than might be expected; for example, we do not need to assume that the rate of nonreporting is independent of race. After we present our main analyses, we return to the nonreporting issue by assessing the extent to which the assumptions that we do require would have to be violated to overturn our qualitative conclusions.

4. METHODS

We begin by discussing an idealized approach that provides not only a test for racial profiling, but also a quantitative measure of its extent. The idealized approach is infeasible because it requires knowledge of visibility of race, which is a function not only of daylight and darkness, but also of such unobservable factors as daytime glare, nighttime street lighting, and the angle from which the police view oncoming traffic. Although the idealized test is infeasible, it demonstrates the important features of our approach.

The idealized approach also serves to highlight an important feature of the feasible test, which is based on observable daylight and darkness rather than on unobservable visibility. Because darkness serves as a proxy for visibility, our feasible veil of darkness test does not provide a quantitative measure of the extent of racial profiling. This is because the magnitude of our test statistic is a function both of the difference in the race distribution of stopped drivers between daylight and darkness and of the relationship between darkness and visibility. Nevertheless, we show that the feasible veil of darkness test is a consistent test for the presence of racial profiling.

We initially impose the restrictive assumption that relative risk is constant; that is, the race distribution of drivers at risk of being stopped is the same during daylight and darkness. We then show how limiting the sample to the intertwilight period and controlling flexibly for time of day through a regression model accounts for potential differences in relative risk arising due to differences in travel times. We argue further that the approach provides implicit controls for potential differences in relative risk that may arise due to differences in driving behavior and police exposure. Finally, we note that the nonreporting problem cannot be dealt with explicitly using the regression model. To deal with nonreporting, we first state the necessary conditions for our approach to yield a valid test, then provide a sensitivity analysis to assess the extent to which those conditions would have to fail to reverse our qualitative conclusions.

4.1 An Idealized Test for Racial Profiling

We begin with an idealized and restrictive form of the test. Let S be a binary random variable indicating whether officers stop a vehicle. Let the binary random variables B and $\bar{B}$ denote the event that a person is black and non-black and at risk of being stopped. To be at risk, the person must be driving a vehicle, be exposed to police, and be committing a traffic offense that would lead police to stop the vehicle if observed. Herein we often use the terms "black driver" and "non-black driver" as shorthand to refer to drivers in the at-risk population who are black and non-black.

Ideally, we would test whether the visibility of race influences officers' decisions to stop particular vehicles. Visibility refers to whether the officer can see the driver's race before making a stop. Although visibility may vary continuously as a function of daylight and other conditions, for simplicity we let V denote the event that race is visible and let $\bar{V}$ denote the event that race is invisible. The idealized test would be based on K_{ideal} in (1),

$$\frac{P(S|V,B)}{P(S|V,\bar{B})} = K_{\text{ideal}}\frac{P(S|\bar{V},B)}{P(S|\bar{V},\bar{B})}. \tag{1}$$

The left side of (1) is the relative risk of a black driver being stopped when race is visible, and the ratio on the right side of (1) is the relative risk of a black driver being stopped when race is not visible. In the absence of racial profiling, K_{ideal} would equal 1, so that the relative risk of being stopped would not depend on whether race was visible. In the presence of racial profiling, K_{ideal} provides a natural quantitative measure of its extent.

Of course, none of the quantities in (1) would be estimable even if V were observed. However, applying Bayes' rule and rearranging yields

$$K_{\text{ideal}} = \frac{P(B|S,V)P(\bar{B}|S,\bar{V})}{P(\bar{B}|S,V)P(B|S,\bar{V})} \times \frac{P(B|\bar{V})P(\bar{B}|V)}{P(\bar{B}|\bar{V})P(B|V)}. \tag{2}$$

The first term on the right side of (2) is an odds ratio measuring the association between visibility and the race of stopped drivers. If visibility were observed, then this term could be estimated from traffic stop data. The second term is the relative risk ratio, that is, the ratio of the relative risk of a black driver being stopped when race is not visible to the relative risk of a black driver being stopped when race is visible. If the relative risk were independent of visibility, then this second term would equal 1. If in addition visibility were observable, then an estimate of the extent of racial profiling, and a test of the null hypothesis of no racial profiling, could be based on the first term in (2).

4.2 The Feasible Veil of Darkness Test

Because no direct measures of visibility are available, we substitute daylight/darkness as a proxy measure for V. Let $d = 1$ represent a stop occurring in darkness and let $d = 0$ represent a stop occurring in daylight. Then, substituting $d = 0$ for V and $d = 1$ for $\bar{V}$ in (2) yields

$$K' = \frac{P(B|S,d=0)P(\bar{B}|S,d=1)}{P(\bar{B}|S,d=0)P(B|S,d=1)} \times \frac{P(B|d=1)P(\bar{B}|d=0)}{P(\bar{B}|d=1)P(B|d=0)}. \tag{3}$$

Equation (3) is analogous to (2) but is based on observable daylight/darkness rather than on unobservable visibility. The first term in (3) is an odds ratio, the odds of being black and stopped during daylight to the odds of being black and stopped during darkness. The second term is the relative risk ratio, defined in terms of daylight and darkness rather than of visibility. Assuming momentarily that the relative risk is constant (i.e., independent of daylight and darkness) yields the veil of darkness parameter K_{vod}, on which we base our test,

$$K_{\text{vod}} = \frac{P(B|S,d=0)P(\bar{B}|S,d=1)}{P(\bar{B}|S,d=0)P(B|S,d=1)}. \tag{4}$$

Proposition 1 shows that although K_{vod} does not in general equal K_{ideal}, it will exceed 1 if there is racial profiling.

Proposition 1: The veil of darkness test. If the following assumptions hold:

1. $K_{\text{ideal}} > 1$
   (there is a racial bias against black drivers);
2. $P(V|d=0) > P(V|d=1)$
   (darkness has a race blinding effect);
3. $\dfrac{P(\bar{B}|d=0)P(B|d=1)}{P(B|d=0)P(\bar{B}|d=1)} = 1$
   (the relative risk is constant: the racial mix of the at-risk population does not change between daylight and darkness),

then $1 < K_{\text{vod}} \le K_{\text{ideal}}$.

For the proof see the Appendix.

Proposition 1 reveals two important properties of our test. First, it shows implicitly that the feasible test, unlike the idealized test, does not provide an estimate of the quantitative extent of racial profiling. As shown in the Appendix, we would have to know that $P(V|d=0)=1$ and $P(V|d=1)=0$ to quantify the extent of racial profiling as defined by K_{ideal}. The intuition is simple: Whereas a qualitative test requires only a restriction on the *sign* of the difference between $P(V|d=0)$ and $P(V|d=1)$, a quantitative measure requires a restriction on the actual *magnitudes*.

At the same time, Proposition 1 provides conditions under which K_{vod} can be used to test the null hypothesis of no racial profiling. Although such a qualitative test may be less informative than a quantitative measure, it is nevertheless an object of considerable importance. Many interest groups and law enforcement agencies have adopted a "zero-tolerance" position on racial profiling, suggesting that they would seek or take remedial action for any value of $K_{\text{ideal}} > 1$ (Williams 2000; U.S. Department of Transportation 2000; American Civil Liberties Union 2003; Dworkowitz 2004; Schwab 2004). Language from the consent decree between the Los Angeles Police Department and the U.S. Justice Department underscores the importance of testing for the null of no racial profiling. According to this decree, "LAPD officers may not use race, color, ethnicity, or national origin (*to any extent or degree*) in conducting stops or detentions..." [emphasis ours] (Los Angeles Police Department 2000).

The assumptions underlying Proposition 1 merit some discussion. Assumption 1 obviously requires that racial profiling be present. Assumption 2 requires that visibility be lower during darkness than during daylight. This does not require complete race-blindness in darkness nor complete race-visibility during daylight, however. The test would be most powerful, and we would have $K_{\text{vod}} = K_{\text{ideal}}$, if d and V were perfectly correlated, but in general this will not be the case.

Some evidence from the literature supports the sign restriction required by assumption 2. For example, Lamberth (2003) described a traffic survey in which the driver's race could be identified in 95% of the vehicles but for which nighttime observations required auxiliary lighting. Greenwald (2001) canceled plans for evening surveys after his observer could identify the race of only 6% of the drivers viewed around dusk. In general, $P(V|d)$ is unknown, but provided that visibility is lower after dark, assumption 2 should hold.

Assumption 3 requires that relative risks be constant. Put differently, it requires that the race distribution of the at-risk population not change between daylight and dark. Because this assumption is not likely to hold in general, we relax it in the next section by controlling for clock time and limiting the sample to stops carried out during the intertwilight period.

4.3 Generalizing the Test

For a number of reasons, the assumption of constant relative risk is restrictive. One reason for this is that temporal travel patterns may vary by race due to differences in hours of work. If so, then the race distribution of the at-risk population may vary by time of day. Racial differences in police exposure or driving behavior could also cause the relative risks to vary. The test also needs to address the nonreporting problem discussed in Section 3.

To relax the assumption that the relative risks are constant, we introduce clock time t into the analysis. We generalize the simple test from Section 4.2 by basing our test for racial profiling on a test of $K(t)$ in the relation

$$\frac{P(S|B,t,d=0)}{P(S|\bar{B},t,d=0)} = K(t)\frac{P(S|B,t,d=1)}{P(S|\bar{B},t,d=1)}. \tag{5}$$

In the absence of racial profiling, we should find that $K(t)=1$ for all t. In the presence of racial profiling, we should find $K(t) > 1$, that is, that blacks are at greater relative risk of being stopped during the daylight than during the dark, when (by hypothesis) racial profiling is more difficult.

We proceed as before by applying Bayes' rule to each of the four probability terms in (5), then solving for the logarithm of $K(t)$ to obtain

$$\begin{aligned}\log K(t) &= \log\frac{P(S|B,t,d=0)}{P(S|\bar{B},t,d=0)}\frac{P(S|\bar{B},t,d=1)}{P(S|B,t,d=1)}\\ &= \log\frac{P(B|S,t,d=0)}{P(\bar{B}|S,t,d=0)}\frac{P(\bar{B}|S,t,d=1)}{P(B|S,t,d=1)}\\ &\quad\times\frac{P(\bar{B}|t,d=0)}{P(B|t,d=0)}\frac{P(B|t,d=1)}{P(\bar{B}|t,d=1)}.\end{aligned} \tag{6}$$

To analyze nonreporting, let R be a binary random variable indicating whether the officer reported the stop. We introduce nonreporting in the expression for $\log K(t)$ by means of the probability relation

$$P(B|S,t,d) = \frac{P(B|R,S,t,d)P(R|S,t,d)}{P(R|B,S,t,d)}. \tag{7}$$

Substituting (7) into (6), collecting similar terms, and making use of the fact that $P(\bar{B}|R,S,t,d) = 1 - P(B|R,S,t,d)$, we obtain

$$\begin{aligned}&\log K(t)\\ &\quad= \log\frac{P(B|R,S,t,d=0)}{1-P(B|R,S,t,d=0)} - \log\frac{P(B|R,S,t,d=1)}{1-P(B|R,S,t,d=1)}\\ &\qquad+\log\frac{P(\bar{B}|t,d=0)}{P(B|t,d=0)}\frac{P(B|t,d=1)}{P(\bar{B}|t,d=1)}\\ &\qquad+\log\frac{P(R|\bar{B},S,t,d=0)}{P(R|\bar{B},S,t,d=1)}\frac{P(R|B,S,t,d=1)}{P(R|B,S,t,d=0)}.\end{aligned} \tag{8}$$

Equation (8) is the key to the analysis that follows. The probabilities in the first line condition only on reported stops, exactly the data that we observe. We can estimate this line from the observed data using logistic regression in which the dependent variable is a race indicator (black/non-black) with d (the darkness indicator) and t (clock time) as covariates. The logistic regression model estimates the regression $f(d,t)$ from the observed data as

$$\log\frac{P(B|R,S,t,d)}{1-P(B|R,S,t,d)} = f(t,d). \tag{9}$$

The second line of (8) is then simply $f(t,0) - f(t,1)$. If the effect of darkness is additive, then this difference is simply the coefficient on the darkness variable times -1.

The third line of (8) measures how the mix of black and white drivers in the at-risk population changes depending on darkness and clock time. If the race distribution of the at-risk population is independent of darkness, then, conditional on clock time, this term vanishes. This is weaker than the assumption of constant relative risk in Proposition 1. Here we discuss the circumstances that may satisfy this weaker condition.

First note that to condition on clock time while estimating daylight/darkness contrasts in the race distribution of stopped drivers, we must limit the sample to stops made at times when it is daylight during certain times of year and dark at other times. In Oakland, the latest occurrence of the end of civil twilight, which we use to define "dark," falls on June 22 at 9:06 PM, and the earliest occurrence falls on December 5 at 5:19 PM. For the remainder of the analysis, we limit the sample to stops occurring between 5:19 and 9:06 PM, which we refer to as the intertwilight period. Restricting the sample in this way allows us to construct contrasts by dark and daylight while controlling for clock time.

Figure 1 represents this idea visually. The horizontal axis indicates the clock time and the vertical axis indicates hours since dark. Throughout the analysis, we omit stops carried out during the roughly 30-minute period between sunset and the end of civil twilight, because that period is difficult to classify as either daylight or dark. The solid points indicate stops of black drivers, and the open circles represent stops of non-black drivers. At any time between 5:19 and 9:06 PM, some stops are made when it is dark (gray shading) and some are made when it is light (no shading). The diagonal bands are a result of the natural variation in daylight hours over the course of the study period. In particular, the large diagonal gap is a result of the shift from Pacific Daylight Time to Pacific Standard Time at the end of October. This shift is especially useful for our comparison because it creates extremes in visibility for fixed clock times.

Within the intertwilight period, we can construct contrasts by daylight and darkness in the fraction of stopped drivers who are black, controlling flexibly for time of day. For example, the vertical lines mark a period around 6:30 PM, within which we can assess whether darkness influences the race of drivers stopped.

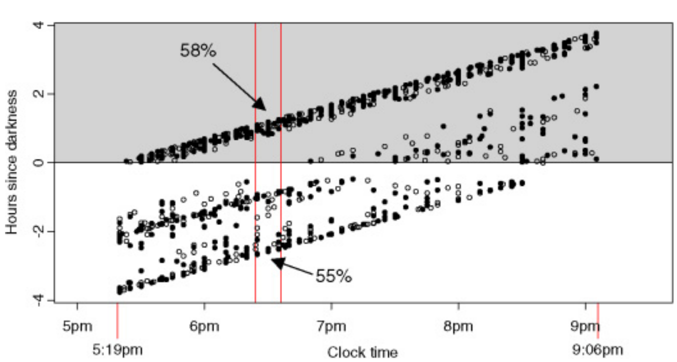

Figure 1. Plot of Stops by Clock Time and Darkness. The solid points indicate black drivers, and the open circles represent non-black drivers. The shaded region indicates those stops occurring after the end of civil twilight. The large diagonal gap is a result of the shift from Pacific Daylight Time to Pacific Standard Time. The figure excludes stops occurring between sunset and the end of civil twilight. The vertical lines near 6:30 PM mark the example region discussed in the text.

During daylight hours, 55% of the stops involved black drivers; after dark, this figure increased to 58%. The full regression analysis will combine such comparisons across the intertwilight period. Note that, although we could potentially include stops carried out during the morning intertwilight period as well as during the evening intertwilight period depicted, we exclude the morning stops simply because they are rare.

Conditioning on clock time makes the assumption that the relative risk is constant between daylight and dark more plausible; see the third line of (8). Recall that the random variable B denotes the event that a black motorist is driving, committing a traffic offense, and observed by police. If travel patterns vary between the races due to variation in commuting times, and commuting times are determined by work hours, it may be reasonable to assume that the drivers who are on the road at 6:30 PM are the same regardless of whether it is daylight or dark. If so, then travel patterns are independent of daylight, conditional on time of day. As for the driving behavior of individuals, differences may arise due to composition effects; drivers on the road at 8 PM may differ on average from those on the road at 6 PM, because the former include a higher proportion of drivers en route to entertainment venues, whereas the latter include a higher proportion of those on their way home from work. Such differences represent time effects rather than daylight effects, so controlling for clock time should equalize them. In a similar vein, in Oakland it is the clock, rather than darkness, that dictates police shifts and allocations. Thus the distribution of police at 6:30 PM should be the same whether or not 6:30 PM occurs after dark. To further control for possible differences in police exposure arising due to differences in patrol intensity by location, we include neighborhood controls in one of the models that we report on later. More generally, the sensitivity test that we carry out in Section 5.3 will help to assess the extent to which our key assumption—that the relative risks are independent of daylight conditional on time of day—would have to be violated to reverse our conclusions.

The fourth line of (8) reveals the condition that reporting rates must satisfy for the regression to yield a valid test. The two ratios in this term measure how much reporting rates change between daylight versus darkness by race, given clock time. If reporting rates vary by race but race-specific reporting rates do not vary between day and night (conditional on clock time), then these two terms vanish. It is important to note that equal reporting rates by race are not needed. Compared with the New Jersey traffic study, where equal reporting rates by race would have been necessary to identify the extent of racial profiling (Kadane and Terrin 1997), our requirement is weaker. Note, however, that if there is a substantial number of officers who are not reporting stops and engaging in racial profiling, then the reporting rate for black drivers during the day is likely to be smaller than the reporting rate for black drivers at night. Newer data collection procedures and audits, such as those implemented by Canter (2004), may increase reporting rates to the point that the probabilities in the nonreporting term are near 1. After presenting our main results in the next section, we return to the nonreporting issue, asking to what extent racial reporting ratios would have to differ between day and night for the conclusions from our main analysis to be reversed.

4.4 Factors Affecting the Veil of Darkness

Anything that reduces the difference between $P(V|d=0)$ and $P(V|d=1)$ may reduce the power of the veil of darkness test. Most obviously, this includes street lighting. Bright street lighting would increase visibility during darkness, reducing the difference between $P(V|d=0)$ and $P(V|d=1)$ and shrinking K_{vod} toward 1. However, it would not affect the sign consistency of $\log K_{\text{vod}}$ unless it completely eliminated the difference between $P(V|d=0)$ and $P(V|d=1)$.

A related problem involves what might called "car profiling." Officers may focus on the characteristics of a vehicle to infer the race of the driver in the vehicle. If car characteristics are correlated with the race of the driver and are visible during darkness, then car profiling has essentially the same effect as bright street lighting, reducing the difference between $P(V|d=0)$ and $P(V|d=1)$. As before, this does not bias the test, but does reduce the test's power to reject the null of no racial profiling.

Both of these problems can be mitigated by additional data collection. For example, city engineering departments may have data on street lighting; such information could be merged with traffic stop data and incorporated into the regression model. Similarly, data on car characteristics could be collected as part of the traffic stop protocol and included in the regression. In future analyses, a seemingly small amount of additional data collection could raise the power of the veil of darkness test.

5. RESULTS

5.1 Comparing Stops During Daylight and Dark

The simple approach described in Section 4.2 can be implemented with the full sample of data. In the full sample, we define daylight as extending from sunrise to sunset and define dark as extending from the end of civil twilight in the evening until the beginning of civil twilight the following morning.

Column 1 of Table 1 displays statistics and sample sizes from our full sample. Of the 7,607 stops at our disposal, we omitted 329 that were made pursuant to a criminal investigation, where the use of race as an identifying factor is explicitly allowed. Another 549 observations were lacking race or time information, 155 were missing the reason for the stop, and another 72 were missing for other reasons unknown to us. Deleting these stops leaves 6,563 usable observations.

The first column of Table 1 presents the fraction of blacks among drivers stopped in the full sample. Among drivers stopped during daylight, 49% were black; among drivers stopped when it was dark, 65% were black. Under the restrictive conditions discussed in Section 4.1, we can test for racial profiling by comparing these two numbers. If anything, this comparison suggests "reverse" racial profiling, because it shows that non-black drivers are disproportionately stopped during daylight when visibility is high. Whether this reflects police behavior or the effect of an important omitted variable, such as racial differences in travel patterns, cannot be said.

The second column of Table 1 presents the percentage of blacks among drivers stopped in the intertwilight sample. Among drivers stopped during daylight, 52% were black; among drivers stopped when it was dark, 57% were black. Restricting the sample to the intertwilight period reduces the contrast between day and night. The intertwilight sample provides little evidence of racial profiling.

5.2 Regression Results

We first consider a simple model that assumes that racial profiling is constant over time. This model takes the form

$$\log \frac{P(B|d,t)}{1-P(B|d,t)} = \beta_0 + \beta_1 d + \gamma_1^T ns_6(t), \tag{10}$$

where $ns_6(t)$ denotes a natural spline basis in clock time with 6 degrees of freedom, γ_1 is a column vector of six parameters, and the superscript "T" denotes transposition. The natural spline allows the model considerable flexibility in adjusting for clock time while enforcing some smoothness to preserve degrees of freedom. For this model, the racial profiling effect is a constant, $\log K(t) = -\beta_1$.

Table 2 presents the estimates of $\log K$ from the intertwilight sample. The estimate in the first row makes no adjustment for clock time and essentially uses only the numbers presented in second column of Table 1 [i.e., $-.19 \approx \log(.52/.48 \times .43/.57)$]. The estimate in the second row adjusts for clock time. The estimate is negative, which constitutes evidence against racial profiling and is consistent with officers stopping black drivers slightly less frequently during daylight than during darkness. Estimation of $\log K$ is imprecise, because the coefficient is smaller in absolute value than its standard error. Adding time-of-day controls has little effect on the evidence of racial profiling.

We also estimate a model that allows for the extent of racial profiling to vary with clock time. This model takes the form

$$\log \frac{P(B|t,d)}{1-P(B|t,d)} = \beta_0 + \beta_1 d + \gamma_1^T ns_6(t) + \gamma_2^T d \times ns_6(t). \tag{11}$$

For this model, $\log K(t) = -\beta_1 - \gamma_2^T ns_6(t)$. Figure 2 plots the estimate by clock time. The shaded area indicates ± 2 pointwise standard errors. Like the previous simpler model, this model yields little evidence of racial profiling; $\log K(t)$ first peaks just before 7 PM but is still well within sampling variability of the horizontal line at 0. It trends upward again after 8:00 PM, but the paucity of stops at that time during daylight causes large standard error estimates.

Table 1. Percent Black Among Stopped Drivers, by Daylight

	Full sample	Intertwilight sample
Total	55% ($n = 6{,}563$)	55% ($n = 1{,}130$)
Daylight ($d = 0$)	49% ($n = 4{,}041$)	52% ($n = 392$)
Dark ($d = 1$)	65% ($n = 2{,}522$)	57% ($n = 738$)

Table 2. Regression Estimates of the Racial Profiling Effect

Adjustments	log K	Standard error
None	−.19	.13
Clock time	−.11	.14
Clock time and neighborhood	−.12	.14

NOTE: In addition to the indicator variable for darkness, the clock time-adjusted models include a natural spline in clock time with 6 degrees of freedom. The third model also includes a set of patrol-area indicators.

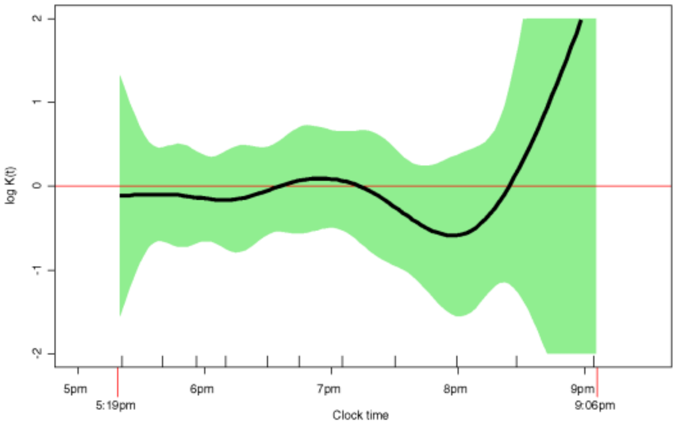

Figure 2. Estimate of log K(t). The curve is the best estimate of K(t), with the shaded area indicating ±2 pointwise standard errors. The horizontal line indicates the K(t) that we would expect under no racial profiling. The inward tickmarks along the x-axis indicate the deciles of the observed stop times.

Finally, we estimated a version of (10) to which we added an indicator variable for each patrol area in the city. These indicator variables provide additional controls for differential exposure to law enforcement between blacks and non-blacks arising from differences in patrol intensity across neighborhoods. The OPD has divided the city into 35 community policing beats that we aggregated into 6 regions. The third row of Table 2 reports the resulting estimate of $\log K(t)$. Controlling for neighborhood with an additive model still yields no evidence of racial profiling at the citywide level. We can refine this one step further with the data available to us by including a darkness×neighborhood interaction term. This will allow us to estimate a racial profiling effect for each neighborhood, as shown in Table 3. We continue to estimate an additive effect of clock time that does not vary by neighborhood.

With the exception of the Hills and West Oakland, the standard errors exceed the estimate of $\log K$. The Hills lie along the eastern border of the city and are predominantly white; West Oakland lies just south of the city's downtown core and is more than 80% non-white. In both areas, $\log K$ is negative, again implying that, if anything, officers are less likely to stop black drivers during daylight.

5.3 Sensitivity Analysis

Although the foregoing results above suggest that there is no racial profiling in traffic stops, those results hinge on assumptions concerning risk ratios and reporting rates. In the second row of Table 2, we estimated $-\beta_1$ to be $-.11$. Under the assumptions maintained earlier,—namely, that differences in reporting ratios do not vary between day and night, and likewise that risk ratios do not vary between day and night in a manner independent of clock time—$\log K = -\beta_1$, because the last two lines of (8) equal 0. But if our assumptions are violated, then the nuisance terms in the last two lines of (8) may be different from 0, in which case $\log K$ would differ from $-\beta_1$. If the sum of those nuisance terms differed from 0 to such an extent that the lower end of the confidence interval exceeded 0, then we would question our conclusion regarding the absence of racial profiling. Although we cannot estimate the nuisance terms directly, in this section we illustrate the magnitude that these terms would have to achieve to overturn our main conclusion.

Table 3. Regression Estimates of the Racial Profiling Effect by Neighborhood

Neighborhood	*log K (neighborhood)*	*Standard error*
Downtown	.12	.26
East Oakland	−.04	.29
Midtown Oakland	−.10	.32
West Oakland	−.51	.29
North Oakland	.67	.72
Hills	−1.07	.90

The lower bound for a 95% confidence interval for $-\beta_1$ is $-.38$. This implies that if the sum of the nuisance terms exceeded .38, then this would shift the estimate for $\log K$ sufficiently for the data to suggest the presence of racial profiling. We focus first on the risk ratio term [the third line in (8)], assuming for the moment that the reporting ratio term [the fourth line in (8)] equals 0.

We consider the circumstances under which

$$\frac{\frac{P(\bar{B}|t,d=0)}{P(B|t,d=0)}}{\frac{P(\bar{B}|t,d=1)}{P(B|t,d=1)}} = \exp(.38) = 1.46. \tag{12}$$

To assess this magnitude, assume that at 6:30 PM on days when 6:30 PM occurs during daylight, black and non-black drivers are at equal risk for being stopped, that is, $P(B|t, d = 0) = .50$. In this case an odds ratio of 1.46 implies that at 6:30 PM on dark days, black drivers compose 59% of the at-risk population. The proportion of black drivers would have to increase by 19% between the days on which it was light at 6:30 PM and days on which it was dark at 6:30 PM.

Focusing next on the reporting term, and assuming that the risk ratio term is 0, if the reporting term exceeds 1.46, then we likewise have evidence for racial profiling,

$$\frac{\frac{P(R|B,S,t,d=1)}{P(R|B,S,t,d=0)}}{\frac{P(R|\bar{B},S,t,d=1)}{P(R|\bar{B},S,t,d=0)}} = \exp(.38) = 1.46. \tag{13}$$

Assume that reporting rates for non-black drivers vary by t but not by d, so that the denominator of (13) is 1. For the reporting term to exceed 1.46, stops involving black drivers would have to be 46% more likely to be reported at night than during the day (e.g., 30% during daylight and 44% in darkness), requiring a substantial fraction of the nonreporting police force to be engaging in racial profiling. We can rearrange the left side of (13) to consider another black/non-black comparison. If stops involving black drivers were twice as likely to be reported during the day as stops involving non-black drivers, then officers would have to report black drivers nearly three times as often as non-black drivers at night to invalidate the "no racial profiling" conclusion.

The sensitivity analysis has considered deviating from the assumptions about the exposure term being 0 and the reporting term being 0, but has not considered both violations simultaneously. If the risk ratio in (12) were 1.21 and simultaneously the reporting ratio in (13) were 1.21, then we would begin to have evidence of racial profiling.

6. CONCLUSIONS

The key problem in testing for racial profiling in traffic stops is estimating the risk set against which to compare the race distribution of stopped drivers. Previous analyses have relied on external estimates of the risk set constructed from either secondary data or traffic surveys. The validity of estimates from secondary data has been questioned. The approach we have proposed here does not require external estimates of the risk set, but it does require certain assumptions. In the case of the Oakland data, our approach yields little evidence of racial profiling, and our sensitivity analysis suggests that the departures from our maintained assumptions would have to be substantial to overturn our conclusions.

A few points concerning limitations are in order. We have noted that our estimates are valid if, controlling for clock time, racial differences in risk sets do not vary between day and night. Implicitly, we have assumed that there is no seasonality in day–night risk differentials. In areas with substantial tourist inflows, this assumption may be violated. To mitigate this risk, one could focus the analysis on those stops that occurred near the switch to and from Daylight Saving Time, ensuring that all stops occurred in the same season.

The method also may be sensitive to violations associated with both driver's race and darkness, such as having a headlight out. Generally such violations represent only a small fraction of the stops. If they are cause for concern, then they may be removed from the analysis. Our analyses were insensitive to the inclusion or exclusion of such stops. A further caveat is that the results are limited to the intertwilight period. Our approach cannot speak directly to the question of racial profiling during other hours.

Because we make assumptions only about the qualitative relationship between darkness and visibility, we can compute only a qualitative test, rather than a quantitative measure of the extent of racial profiling. The test is consistent, but its power is reduced by anything that reduces the correlation between visibility and darkness. In the case of two important examples, street lighting and car characteristics, additional data collection could boost the power of the test to detect racial profiling.

Our approach is designed to assess the extent of racial profiling in traffic stops only. Other studies have noted racial disparities in poststop outcomes, such as stop duration and search rates (Ridgeway 2006). Data on a full set of poststop outcomes are needed to provide a comprehensive assessment of racial profiling. Finally, we stress that our empirical results apply only to Oakland and say nothing about the presence or absence of racial profiling in other jurisdictions.

APPENDIX: PROOF OF PROPOSITION 1

From (4), we have

$$K_{\text{vod}} = \frac{P(B|S, d=0)P(\bar{B}|S, d=1)}{P(\bar{B}|S, d=0)P(B|S, d=1)}$$
$$= \frac{P(S|B, d=0)P(B|d=0)}{P(S|B, d=1)P(B|d=1)} \times \frac{P(S|\bar{B}, d=1)P(\bar{B}|d=1)}{P(S|\bar{B}, d=0)P(\bar{B}|d=0)}. \quad \text{(A.1)}$$

Assumption 3 yields

$$K_{\text{vod}} = \frac{P(S|B, d=0)}{P(S|B, d=1)} \frac{P(S|\bar{B}, d=1)}{P(S|\bar{B}, d=0)}. \quad \text{(A.2)}$$

Note that darkness only influences the probability of stop through visibility, so that

$$\begin{aligned} &P(S|B, d=1) \\ &\quad = P(S|V, B, d=1)P(V|B, d=1) \\ &\qquad + P(S|\bar{V}, B, d=1)P(\bar{V}|B, d=1) \\ &\quad = P(S|V, B)P(V|d=1) + P(S|\bar{V}, B)P(\bar{V}|d=1). \end{aligned} \quad \text{(A.3)}$$

The second equality in (A.3) uses the fact that S is independent of d given V and B. Let $\alpha_0 = P(V|d=0)$ and $\alpha_1 = P(V|d=1)$. Substituting the relation in (A.3) into (A.2), we have

$$K_{\text{vod}} = \frac{P(S|V, B)\alpha_0 + P(S|\bar{V}, B)(1-\alpha_0)}{P(S|V, B)\alpha_1 + P(S|\bar{V}, B)(1-\alpha_1)} \times \frac{P(S|V, \bar{B})\alpha_1 + P(S|\bar{V}, \bar{B})(1-\alpha_1)}{P(S|V, \bar{B})\alpha_0 + P(S|\bar{V}, \bar{B})(1-\alpha_0)}. \quad \text{(A.4)}$$

Therefore, K_{vod} depends on a nonlinear function of the four stop probabilities and the two visibility probabilities. Note that if there is no veil of darkness, then $\alpha_1 = \alpha_0$ (darkness is uncorrelated with visibility) and $K_{\text{vod}} = 1$ regardless of the value of K_{ideal} and the extent of racial bias. On the other hand, if the veil of darkness is perfect, then $\alpha_1 = 0$ and $\alpha_0 = 1$ (darkness completely hides race and daylight completely reveals it) and $K_{\text{vod}} = K_{\text{ideal}}$. When $\alpha_1 < \alpha_0$, from assumption 2,

$$\begin{aligned} &\frac{\partial}{\partial \alpha_1} \log K_{\text{vod}} \\ &\quad = \left(1 - \frac{P(S|V, B)}{P(S|\bar{V}, B)} \frac{P(S|\bar{V}, \bar{B})}{P(S|V, \bar{B})}\right) \\ &\qquad \times P(S|V, \bar{B})P(S|\bar{V}, B) \\ &\qquad \times \Big(\big(P(S|V, \bar{B})\alpha_1 + P(S|\bar{V}, \bar{B})(1-\alpha_1)\big) \\ &\qquad\qquad \times \big(P(S|V, B)\alpha_1 + P(S|\bar{V}, B)(1-\alpha_1)\big)\Big)^{-1}. \end{aligned} \quad \text{(A.5)}$$

The first term in (A.5) is $1 - K_{\text{ideal}}$, which, by assumption 1, is negative. The second term is positive, implying that K_{vod} is strictly decreasing in α_1. At α_1's extremes, we know that K_{vod} can equal K_{ideal} and 1. Because K_{vod} is strictly decreasing in α_1, we have $1 < K_{\text{vod}} \le K_{\text{ideal}}$.

[Received July 2004. Revised October 2005.]

REFERENCES

Alpert, G. P., and Dunham, R. G. (1998), *Policing Multi-Ethnic Neighborhoods*, Westport, CT: Greenwood Press.

Alpert, G. P., Smith, M. R., and Dunham, R. G. (2003), "Toward a Better Benchmark: Assessing the Utility of Not-at-Fault Traffic Crash Data in Racial Profiling Research," in *Confronting Racial Profiling in the 21st Century: Implications for Racial Justice*, Boston.

American Civil Liberties Union (2003), "Summary of Sacramento Data," press release, March 2003.

Burges, R., Evans, K., Gruber, C., and Lopez, C. (2004), "Second Quarterly Report of the Independent Monitor," available at *http://www.oaklandpolice.com/agree/2qtr.pdf*.

Canter, P. (2004), "Baltimore County Police Department's Experience With Traffic Stop Data Collection Forms," PERF/COPS Conference.

Dominitz, J. (2003), "How Do the Laws of Probability Constrain Legislative and Judicial Efforts to Stop Racial Profiling?" *American Law and Economics Review*, 5, 412–432.

Dworkowitz, A. (2004), "Round up the Black Guys?" *Hartford Advocate*, available at *www.hartfordadvocate.com/gbase/News/content.html?oid=oid:74187*.

Fridell, L. A. (2004), *By the Numbers: A Guide for Analyzing Race Data From Vehicle Stops*, Washington, DC: Police Executive Research Forum.

Gallup, G. Sr. (1999), *The Gallup Poll: Public Opinion 1999*, Wilmington, DE: Scholarly Resources, Inc.

General Accounting Office (2000), *Racial Profiling: Limited Data Available on Motorist Stops*, Washington, DC: Author.

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

Hamermesh, D. S. (1996), *Workdays, Workhours, and Work Schedules: Evidence for the United States and Germany*, Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Harris, D. A. (1999), "The Stories, the Statistics, and the Law: Why 'Driving While Black' Matters," *Minnesota Law Review*, 84, 265–326.

Kadane, J. B., and Terrin, N. (1997), "Missing Data in the Forensic Context," *Journal of the Royal Statistical Society*, Ser. A, 160, 351–357.

Klinger, D. A. (1997), "Negotiating Order in Patrol Work: An Ecological Theory of Police Response to Deviance," *Criminology*, 35, 277–306.

Knowles, J., Persico, N., and Todd, P. (2001), "Racial Bias in Motor Vehicle Searches: Theory and Evidence," *Journal of Political Economy*, 109, 203–229.

Lamberth, J. (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, Temple University, Dept. of Psychology.

——— (2003), "Racial Profiling Data Analysis Study: Final Report for the San Antonio Police Department," available at *http://www.sanantonio.gov/sapd/pdf/LamberthSanAntonioRpt_2003.pdf*.

Lange, J. E., Blackman, K. O., and Johnson, M. B. (2001), "Speed Violation Survey of the New Jersey Turnpike: Final Report," report to New Jersey Attorney General's Office, Public Services Research Institute.

Los Angeles Police Department (2000), "Settlement Agreement in United States vs. City of Los Angeles, Board of Police Commissioners, and the Los Angeles Police Department," available at *http://www.lapdonline.org/pdf_files/boi/final_consent_decree.pdf*.

MacDonald, H. (2001), "The Myth of Racial Profiling," *City Journal*, available at *http://www.city-journal.org/html/11_2_the_myth_.html*.

McConnell, E. H., and Scheidegger, A. R. (2001), "Race and Speeding Citations: Comparing Speeding Citations Issued by Air Traffic Officers With Those Issued by Ground Traffic Officers," paper presented at the annual meeting of the Academy of Criminal Justice Sciences, Washington, DC, April 4–8.

McMahon, J., Garner, J., Davis, R., and Kraus, A. (2002), *How to Correctly Collect and Analyze Racial Profiling Data: Your Reputation Depends on It!*, Washington, DC: U.S. Government Printing Office.

Montgomery County Department of Police (2002), "Traffic Stop Data Collection Analysis," 3rd report.

Northeastern University Data Collection Resource Center (2005), available at *http://www.racialprofilinganalysis.neu.edu/background/jurisdictions.php*.

Oakland Police Department (2004), *Promoting Cooperative Strategies to Reduce Racial Profiling: A Technical Guide*, Oakland, CA: Author.

Peace Officer Standards & Training Program (2002), *Racial Profiling: Issues and Impact, Preparation Guide*, Sacramento, CA: Author.

Pritchard, J. (2001), "Racial Profiling a Conundrum for Police," *Los Angeles Times*, January 21, A22.

Ramirez, D., McDevitt, J., and Farrell, A. (2000), *A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned*, Washington, DC: U.S. Department of Justice.

Ridgeway, G. (2006), "Assessing the Effect of Race Bias in Post–Traffic Stop Outcomes Using Propensity Scores," *Journal of Quantitative Criminology*, to appear.

San Jose Police Department (2002), *Vehicle Stop Demographic Study*, San Jose, CA: Author.

Schwab, J. (2004), "New Police Chief Talks Policy With Minority Officers' Group," *Boston Bay-State Banner On-Line*, available at *www.baystatebanner.com/archives/stories/2004/March%201104-2.htm*.

Smith, M. R., and Alpert, G. P. (2002), "Searching for Direction: Courts, Social Science, and the Adjudication of Racial Profiling Claims," *Justice Quarterly*, 19, 673–703.

Steward, D. (2004), *Racial Profiling: Texas Traffic Stops and Searches*, Austin, TX: Steward Research Group.

U.S. Department of Transportation (2000), "U.S. Secretary of Transportation Slater Calls for Zero Tolerance for Racial Profiling," news release, available at *http://www.dot.gov/affairs/2000/7500.htm*.

Weiss, A., and Grumet-Morris, A. (2005), *Illinois Traffic Stop Statistics Act Report for the Year 2004*, Evanston, IL: Northwestern University Center for Public Safety, available at *http://www.dot.state.il.us/trafficstop/2004summary.pdf*.

Williams, H. (2000), "Statement on National Police Practices and Civil Rights to U.S. Commission on Civil Rights," available at *http://www.policefoundation.org/docs/061600.html*.

Zingraff, M. T., Mason, M., Smith, W., Tomaskovic-Devey, D., Warrent, P., McMurray, H. L., and Fenlon, R. C. (2000), "Evaluating North Carolina State Highway Patrol Data: Citation, Warnings, and Searches in 1998," report submitted to North Carolina Department of Crime Control and Public Safety and North Carolina State Highway Patrol.