Abstract

The term “racial profiling” describes race-based selection of citizens for interdiction by police and other legal actors. Several studies have examined whether police disproportionately stop minority citizens both in cars and on foot, and, once stopped, whether police are more likely to search or arrest them. Whether these contacts are racially motivated has been the focus of research, litigation, political mobilization, and internal scrutiny by police departments. This article reviews definitions of practices that are commonly described as racial profiling, contrasts these narrow views with the more complex legal standards that have evolved in case law, and assesses whether recent data collection efforts can generate reliable information about the extent and nature of racially disproportionate police contacts with citizens. Data analysis procedures are identified to respond to both legal and normative questions about whether racial disparities in police stops and searches rise to the level of “profiling” and cross the threshold of a violation of constitutional guarantees. The article concludes with a brief discussion of mechanisms for regulating and monitoring police-citizen contacts to address concerns of police and citizens on the racial dimensions of policing.
In the late 1990s, the deaths or shootings of African-American citizens during encounters with police focused national attention on the practice of “racial profiling” by law enforcement officials. The fact that police pay more attention to non-whites was hardly news to most non-white Americans. Close surveillance by police has always been a part of everyday life for American minorities, especially African-Americans (Kennedy, 1997). More broadly, the use of race as an indicia of criminality also is hardly new.1 A recent Gallup poll showed that majorities of Americans of both races believe racial profiling is a problem.2 The reality of intensive police surveillance of minority citizens—especially persons of African descent—is no secret either to law enforcement officials, rank-and-file officers, or to minority citizens. What was surprising was revelations of the extent to which this intensive surveillance and interdiction of minorities was tolerated and condoned by police officials, and ingrained in police practice and organizational norms (see, for example, Veneiro, 1999; Goldberg, 1999).

Responding to both litigation and popular concern, several recent studies confirmed that police disproportionately stop minority citizens, and, once stopped, are more likely to search or arrest them (Cole, 1999; Harris, 1999, 2002). Racially disproportionate stops and searches by police take place in a wide range of everyday citizen transactions and movements: highway and local traffic stops, upon departure and arrival at both domestic and international air terminals, on intercity trains and buses, in shopping malls and private stores, at the borders with Canada and Mexico, on street corners in neighborhoods known for crime or gang activity, and in routine pedestrian travels (see, for example, Langan, Greenfeld, Smith, Durose, & Levin, 2001).

Whether these contacts are racially motivated is a hotly contested question that has been the focus of public and private litigation, political mobilization,

---

1 Examples include: immigration exclusion and other discrimination against Chinese immigrants in the 19th century (Yick Wo v. Hopkins, 118 U.S. 356, 1886); the racialized debate on the Harrison Act criminalizing heroin and cocaine in 1913 (Musto, 1973); the internment of the Japanese during World War II (Korematsu v. United States, 323 U.S. 214, 1944); border interdictions to halt illegal immigration (U.S. v. Martinez-Fuerte, 428 U.S. 543, 1976); the racial components of drug courier profiling (U.S. v. Harvey, 16F.3d 109, 115, 6th Cir. 1994); and the so-called “Carol Stuart” stops in Boston (Massachusetts Attorney General, 1990).

2 Associated Press, Poll Finds Most in U.S. Believe Police Practice “Racial Profiling,” Chicago Tribune, Dec. 11, 1999, p. 15 (describing results of a Gallup poll in which a majority of Americans believed that racial profiling is widespread and three fourths of black men said they have been stopped by police because of their race).

3 Veneiro admitted that New Jersey State Police officers engaged in racial profiling, but also that profiling is part of the culture of the State Police.
and self-scrutiny by several police departments (Garrett, 2001). The courts have been active, though inconsistent, in defining the boundaries of constitutionally acceptable actions by police in targeting minorities for stops and searches. Whether these practices constitute racial profiling is both an analytic and legal question.

How to address that question is the focus of this article. The first section reviews definitions of practices that are commonly described as racial profiling, and contrasts these narrow views with the more complex legal standards that have evolved in case law. Next comes a brief review of recent efforts to promote data collection to measure police-citizen interactions and police stops specifically. These data collection efforts are then assessed in terms of their ability to generate sufficient information to establish both the fact pattern of racialized policing and the base rates for the assumptions that animate police beliefs, norms, and actions. The fourth section reviews data analytic strategies that are best suited to respond to both legal and normative questions about whether racial disparities in police stops and searches rise to the level of profiling and cross the threshold of a violation of constitutional guarantees. Finally, the article concludes with a brief discussion of possibilities for regulation and monitoring of police-citizen contacts to address concerns of police and citizens on the racial dimensions of policing.

Profiling Controversies

What Is Racial Profiling?

The term “racial profiling” has emerged as a broad descriptive category that encompasses many different practices by police and other legal actors. Skolnick and Caplovitz (2001) note that the term first appeared in the U.S. press in 1987, but was infrequently mentioned through 1993. Beginning in 1994, its use in the popular literature rose exponentially, with nearly 1,000 mentions in 2000. In most current definitions, police engage in racial profiling when they select persons of a specific race for attention because they assume that such persons are more prone to commit a targeted crime or crime generally than a white or majority person (see, for example, International Association of Chiefs of Police [IACP], 1999; U.S. General Accounting Office, 2000). Skolnick and Caplovitz (p. 417) define racial profiling as “police targeting people solely or primarily because of their skin color.” Gross and Barnes (2001) and Gross and Livingston (2002) say that the term racial profiling has no set meaning, but they go on to say that racial profiling “occurs when a police officer stops, arrests, questions, searches or otherwise investigates a person because the officer believes that members of that person’s
racial or ethnic group are more likely than the population at large to commit the sort of crime that the officer is investigating” (p. 4).

Other criteria—such as age, gender, location, or behavior—are irrelevant. Inherent in these narrow definitions is the claim that police officers make an a priori probabilistic assessment that members of a specific race are more likely than “similarly situated” members of another race to be engaged in criminal activity, and therefore select them for stops. Citing the difficulties in defining racial profiling, PERF recently adopted the term “racially-biased policing” (Fridell et al., 2001).

Perhaps the broadest definition was published in the U.S. Department of Justice Resource Guide on Racial Profiling Data Collection Systems, which defines racial profiling as “any police-initiated action that relies on race, ethnicity or national origin rather than the behavior of an individual who has been identified as being, or having been, engaged in criminal activity” (Ramirez, McDevitt, & Farrell, 2000, p. 3). This definition suggests that but for race, certain police actions would not occur. However, this definition conflates race-based suspect descriptions with selective (discretionary) enforcement.4

Legal Standards and Case Law

The courts have focused on the question of discriminatory intent as central to the legal standard to uphold a claim of racial profiling.5 The courts have considered litigation on profiling claiming both Fourth Amendment search and seizure violations and Fourteenth Amendment equal protection claims.

---

4 Whether racial profiling includes stops resulting from race-based suspect descriptions is controversial. Gross and Barnes (2001), Rudovsky (2001), Cole (1999), Kennedy (1997), Davis (1997) and several others carefully distinguish such stops from stops based on discretionary judgements about suspicion. Case law also distinguishes racial profiling from police stops motivated by race-based suspect descriptions. In Brown v. Oneonta (221 F.3d 769, 2nd Cir. 1999), for example, the Court of Appeals in the 2nd Circuit found no evidence of an equal protection violation when police stopped over 200 African-American males (and one woman) in the City of Oneonta based on a suspect description provided by a victim of attempted rape. But Banks (2001) claims that race-based suspect descriptions are no different than other stops motivated by racial classifications. He cites four weaknesses in victim-supplied suspect descriptions that reduce the targeting of suspects to the same probabilistic (and therefore flawed) determination as in a broader race-based profile: (1) the fallacy of appearances of a category of people who resemble the suspect, (2) the unreliability of victim descriptions, (3) the overly broad application of the suspect description, and (4) the excessive weighting of race in a “multiple factors” suspect description. Banks claims that by virtue of the similarity of race-based suspect descriptions and race-based profiles, both offend the principles of colorblindness in the Equal Protection Clause in that they disproportionately burden racial minorities, and both should be subject to strict scrutiny within the Fourteenth Amendment.
A racially motivated stop may not be discriminatory under Fourth Amendment law if additional factors—including the suspect’s demeanor as well as contextual factors—constitute “reasonable suspicion” that the suspect may be involved in a crime. Cases such as Brignoni-Ponce (422 U.S. 873, 1975) and Martinez-Fuerte (428 U.S. 543, 1976), involving stops of people of Mexican ancestry at or near the U.S. border, suggest that race can be taken into account together with other factors to constitute reasonable suspicion in deciding which cars to stop on suspicion of immigration violations. In U.S. v. Whren (116 S.Ct. 1769, 1996), the Court said that “selective enforcement” was permissible under Fourth Amendment law, and that any racial animus in the stop was irrelevant if the stop was based on reasonable suspicion. In the arc of these cases, the courts generally have not been open to Fourth Amendment challenges based on statistical claims of racial disproportionality.6

Recently the courts have expanded the reasonableness standard to include location of the individual as well as the individual’s behavior as an indicia of suspicion. In Illinois v. Wardlow (120 S.Ct. 673, 2000), the Supreme Court noted that although an individual’s presence in a “high crime area” does not meet the standard for a particularized suspicion of criminal activity, a location’s characteristics are relevant in determining whether an individual’s behavior is sufficiently suspicious to warrant further investigation. Since “high crime areas” often are areas with concentrations of minority citizens (Massey & Denton, 1993), this logic places minority neighborhoods at risk for elevating the suspiciousness of its residents. And, in State v. Dean, police officers questioned a suspect in a parked car because “he was a Mexican male in a predominantly white neighborhood”—in other words, because he was out of place (Kennedy, 1997).

5 The U.S. Supreme Court has held that strong statistical evidence of racial discrimination alone does not support an Equal Protection claim. In McClesky v. Kemp (481 U.S. 279, 1987), the Court didn’t contest the research by David Baldus showing that people who killed white victims were far more likely to receive death sentences than people who killed black victims. Baldus controlled for dozens of non-racial variables that might explain sentencing outcomes, and found a strong pattern of racial bias (Baldus, Woodworth, & Pulaski, 1990). But the court demanded evidence of discriminatory intent by the prosecutors in McClesky’s case to show that prosecutors acted with “discriminatory purpose” (McClesky, p. 282). The Court specifically declined to equate the statistical probability of discriminatory outcome with actual discrimination that would influence capital sentencing.

6 But in New Jersey v. Soto (734 A.2d 350), a criminal court accepted statistical evidence that black motorists on the New Jersey Turnpike were stopped at higher rates than were whites. But the important distinction here is that Soto and other cases involved suppression of evidence in criminal cases based on search and seizure (Fourth Amendment) violations, not equal protection (Fourteenth Amendment) discrimination claims.
The courts have given conflicting signals on whether statistical evidence of racial disproportionality is sufficient to show discriminatory practices that would invite Fourteenth Amendment protection (and therefore supportive of a racial profiling claim). To satisfy an equal protection claim, the Court in *Whren* said that the selective enforcement had to be both racially motivated and racially discriminatory to satisfy a Fourteenth Amendment claim. In *U.S. v. Armstrong* (517 U.S. 456, 1996), the Supreme Court established a “similarly situated person” (or but for) test to prove that a citizen of another race committed the same crime but that the law was not enforced—in other words, the burden of proof on plaintiffs in these cases is to show both discriminatory intent and discriminatory purpose.

In sum, the courts agree in both Fourth Amendment claims (e.g., *Soto*) and Fourteenth Amendment claims (*Whren*) that police cannot stop, search, or arrest citizens solely because of their race. The courts have often invoked “reasonable suspicion” in adjudicating racial profiling claims under both Fourth and Fourteenth Amendment doctrines, but offer no consistent standard for defining it. The necessity of additional factors that constitute reasonableness broadens the definition of profiling beyond race, and introduces into the definition factors such as suspect behavior and location. Accordingly, it is constitutionally permitted to stop or search people because of their race if race is one of several motivating factors. But the addition of non–race-based factors in fact opens the door to stops made based on a substitution of the correlates of race for the actual racial categorization. In effect, it is cheating statistically to cite other non-racial factors as a stop rationale when in fact the second factor adds no new information to the probabilistic determination. Ruling out race as the sole justifying factor in a police action opens profiling claims to “strict scrutiny” about the purposes and state interests in race-based actions.

How Should We Measure Racial Disparity in Policing?

**The Numerator**

Measurement systems for highway stops vary considerably (Ramirez et al., 2000). Typically, driver and passenger characteristics, vehicle information (in-
cluding license plate number and state of registration), the rationale for the stop, interior and exterior condition of the vehicle, and the outcomes are recorded. Subjective assessments of driver and passenger comportment are included in some systems (e.g., North Carolina). In some locales, police record whether the driver consented to the search and the legal basis for the search. There is surprising inconsistency in police recording of the location and conditions of the stop.

Methods of data collection vary. Information sometimes is transmitted by radio to central locations where it is entered, or, alternately, entered into a patrol car’s mobile data terminal (MDT) unit and then transmitted. Some departments have experimented with laptop computers, others with personal digital assistants (PDAs) (“Palm Pilots” and their progeny). In fact, designing equipment and software for police data collection is now a big business. For example, Mobile Commerce & Computing designs and sells wireless handheld computers and software with a system called TrafficStop to local police departments.

The reliability of these data collection systems is difficult to measure. I mention three points of vulnerability, recognizing that there are many others. First, discretion still resides with officers, and stops may be selectively recorded. In *Chavez v. Illinois State Police* (251 F.3d 612, 7th Cir. 2001), the court found that the database of field reports was too incomplete to be a reliable data source to accurately measure the stop activity of the state police. Even with the best of intentions and most rigorous training, police may not record stops that are preempted by an arrest whose recorded information subsumes what might be reported in a stop that has a different outcome.

---

9 See, for example, John Drake, Police Get a Hand From New Palm PCs, *The Washington Times*, August 29, 2000 at C1 (“Officer Johnson recently downloaded 76 traffic stops into a computer at headquarters in 2 1/2 minutes, cutting his paperwork to a fraction of its former amount. ‘It makes it a lot easier than filling out the paperwork, handing it to someone, who gives it to someone else, who gives it to a data entry person,’ Officer Johnson said.”); Michael Raphael and Joe Donahue, State Plans Database to Deter Profiling: Troopers to Get Laptops in $14 million Project, *The Star Ledger* (Newark, N.J.) June 23, 1999 (describing how New Jersey will provide laptops to all officers in the state, streamlining recordkeeping, and requiring data on all traffic stops). See also Greg Botonis and Holly Edwards, Deputies Note Races of Those They Contact Baca Follows LAPD Lead, L.A. Daily News, Jan. 11, 2001, at AV1 (“At first, we thought this would be a lot of extra work, but so far it hasn’t proved to be that bad,” said Deputy Dave Miklos); Armando Acuna, More Police Agencies Keeping Racial Data Driving, *Los Angeles Times*, Sept. 30, 1999, A3 (noting that San Diego now requires officers to enter information on stops on hand-held computers).

One way to address and monitor selectivity is to use video cameras. Harris (1997) studied stops on Interstate 95 near Volusia, Florida, and racial classifications were based on videotaped records by sheriffs’ deputies. Problems in the taping (inconsistent time periods when stops were taped, deputies taping over earlier taped stops) made the sampling parameters of the collection of taped stops uncertain. New Jersey required the installation of mobile video cameras or audio recorders in State Police Cars to document all stops. Other places invite the use of “bodycams” or other video recording devices.

Second, falsification of records, especially with respect to race, has been verified in some instances, and the extent of this practice is unknowable without a video record or independent observation. Sometimes falsification is deliberate, as in the case of Troopers McKenna and Hogan in New Jersey. (It was later shown to be far more widespread; see Kocieniewski, 2002), and other times it may be unintentional, as in Chavez, when Illinois State Troopers classified Chavez as white.

Third, racial classifications often are both challenging and problematic. San Diego’s data collection form includes 18 racial categories, but San Jose’s has 8. San Diego’s includes a designation for Middle Eastern ethnicity, San Jose’s does not. It certainly would be possible to validate the relative merits of these two designs for racial classification, but it is quite another question as to which is better and more accurately applied by police during the course of everyday contacts with citizens. How well they work in situ compared to in a controlled setting is still another question.

Beyond that, racial classification skills are highly variable. Even highly disaggregated racial classification schemes may be faulty when there is sufficient overlap in racial “appearance.” The potential disjunction between physical appearance and racial self-identification and classification is evident for mixed race or multi-racial persons, and the number of persons of mixed race is growing (Wesley, 1999). The eventual dissolving or transformation of current racial categories is inevitable, complicating the meaning of race-based equal protection challenges.

The potential for moments of racial confusion, perhaps intensified by the emotional arousal of a police-citizen encounter, can defeat current efforts to ac-

---


curately record racial classification (see, for example, Banks, 2001).\textsuperscript{13} Even without these problems, cross-racial identification skills vary and are subject to a variety of stereotyping processes (Thompson, 1999).\textsuperscript{14}

**Measures of Traffic Enforcement.** From stop data, a wide range of performance indicators are computed. Of course, racial disaggregation is central to these computations. In addition to this racially disaggregated “epidemiology” of stops, reports of police activity include: (1) percent of cars stopped that are searched with consent, (2) percent of cars searched without consent, (3) traffic citations issued, (4) arrests made of driver or passengers, (5) contraband seized (amount and type), (6) property seized, and (7) use of force (if any).

The simplicity of the data collection protocols belies the complexities of their interpretation. The selection of drivers for highway or street stops may reflect a speeding violation, or may reflect the officers’ judgement that a car has a burned-out license plate light, cracked windshield, broken tail light, or obscured rear-view mirror, or that the driver is not wearing a seatbelt. The latitude accorded to police officers is incredibly broad: a California Highway Patrol Officer bragged recently that “the Vehicle Code gives me fifteen hundred reasons to pull you over.”\textsuperscript{15} In other words, stops can occur with no legal requirement for a police officer to justify the stop. Under *Whren*, such stops are legal, since these putative vehicle code violations satisfy the Supreme Court’s requirement for reasonable suspicion. Understanding the rationale for a stop and when it is real or pretextual exceeds the information generated in any of the protocols that have been developed, and the selection process opens the discussion of what Samuel Walker has called “the denominator problem” (Walker, 2001). That discussion takes place in a later section of this article.

The validity of stop or search counts also is weakened by vagueness in defining “search.” The process of a stop may be extremely intrusive, and an officer may undertake search-like actions, yet because no actual search takes place, these events are neither recorded nor constitutionally regulated. Yet

\textsuperscript{13} In *Choi v. Gaston*, 220 F.3d 1010, 1015 (2000), police were given a description of a Vietnamese man and arrested a Korean man even though “none of the figures given matched Choi’s height, weight, or age, nor did Choi’s clothing match” the suspect’s. The court noted the “rash generalization” based on seeing another Asian man in the vicinity.

\textsuperscript{14} Thompson notes that “If, however, as social science data suggest, police behavior may result from various types of cues and negative perceptions that such cues trigger, then the best place to begin changing police officers’ behavior is by changing their experience with communities and individuals of color. Exposing officers to these communities in less confrontational ways may begin to broaden officers’ perception of individuals within these neighborhoods” (p. 1,009).

they can be intrusive and humiliating. In *Wilkins v. Maryland State Police*\(^{16}\) for example, the stop of Reverend Wilkins and his family lasted nearly 45 minutes. When Reverend Wilkins refused to consent to a search, the family was detained while a K-9 dog was summoned. The family had to wait outside the car in the rain at approximately 6:25 a.m. while the dog searched the car. None of this was recorded, and this particular type of search is not regulated by Fourth Amendment case law. Ordering the family out of a car, requesting consent to search, and an external sniffing of a car by a dog are “subsearches” that do not require reasonable suspicion, nor do they violate citizens’ expectations of privacy. After Wilkins, procedures were designed to avoid these issues. But Gross and Barnes (2001) report that only one in 70 drivers who were stopped in Maryland after the consent decree was signed were subjected to a search that would trigger the reporting guidelines that had been developed. Whatever happened in the unreported stop events escaped the new data collection system.

Once a full search is undertaken, its meaning also may exceed what the data can tell us, and the measure of search rates may have validity and reliability problems. A trooper or police officer needs one of two justifications to conduct a search: consent or probable cause. A search also may be conducted “incident to an arrest” (LaFave, 1996). Officers are required to file reports when searches are requested in Maryland, for example, and the consent rate of 96% is very high. But there is no way to verify that the high rates of drivers consenting to search are accurate, or whether the search data field is selectively completed only when and if the driver consents. Searches where consent is not given presumably are probable cause searches. Yet, most data systems are far from uniform with respect to recording the elements that informed the officers’ probable cause determination. Typical bases might include: drugs in plain view, odor of marijuana, a K-9 dog alert, or other factors; information about probable cause factors has been widely disseminated under Operation Pipeline and various local adaptations (the U.S. Department of Justice and the Drug Enforcement Administration have produced much information in this area; see also Harris, 2001). In Maryland, consent searches more often yielded contraband than did probable cause searches (Gross and Barnes, 2001).

Statistics and measurement are, of course, inextricably linked, and these reliability concerns about police stop and search data can cumulatively produce reporting biases in a direction that makes stops look both productive and racially more neutral than they may, in fact, be. In the Maryland litigation, one Maryland state police commander

---

Street Stops. Data collection in street stops differs from procedures in highway stops in small but important ways. For example, the New York City Police Department records information about stops of citizens on its UF-250 form. It was originally designed as an investigation aid, and later became a basic management tool. In the United Kingdom, the data collection system grew out of the Police and Criminal Evidence Act (the PACE Act). The Act itself empowered U.K. police officers to stop and search any person or vehicle when the officer has “reasonable grounds” for suspecting that stolen or prohibited articles would be recovered. PACE also permits officers to conduct full searches of suspects, searches that go well beyond the parameters of U.S. case law that limit searches to outer clothing.

The U.K. system distinguishes between high discretion or proactive searches and low discretion searches. The former include searches based on “suspicious activity” that may rise to the level of what U.S. case law defines as “reasonable suspicion.” New York’s Order Maintenance Policing strategy (Bratton and Knobler, 1998) fits this description. The latter are searches motivated by information given to police from “other sources,” such as victim descriptions of suspects or criminal acts observed by police officers. In New York, the Attorney General’s Report noted that only about 30% of all stops were “low discretion” stops in which police officers stated that the person stopped matched a suspect description or that they had observed a crime. The comparable indicator in the U.K. in its pilot study was 25% (Ramirez et al., 2000).

In the U.K., stops resulting in searches are not recorded (except in some pilot areas), but stops in New York, are (in theory) recorded for all citizens stops by police are (in theory) recorded. The New York State Attorney General’s Report (1999) documented lapses in reporting, lapses that were confirmed by the City in discovery motions in litigation. Policy in New York mandates specific situations requiring a stop form (the UF-250) be filled out, and forms are discretionary in other types of stops (Office of the Attorney General of the State of New York, 1999, citing the New York Police Department Patrol Guide). Both systems use paper forms.

The well-documented protocols developed in New York and in the United Kingdom differ in small but important ways. First, the use of self-reported ethnicity

---

in the U.K. program avoids problems noted above of complex racial categorizations, but raises new concerns that weaken the resulting data. Police use 17 self-identification categories, inviting reliability problems. Both the categories and the procedures were not without their problems (Miller, Bland and Quinton, in this volume), including out-of-range codes, refusals by some citizens to self-identify their race, and discomfort for some officers. One in three refused to report details about their race, and officers were unable to classify 8.5%. Race, ethnicity, and country of origin were confounded in some of the responses, and other responses were clever and sarcastic evasions.

Even when such “low discretion” stops are motivated by victim-supplied suspect descriptions, they are vulnerable to problems of racial classification and other perceptual distortions mentioned earlier. Inability to accurately translate racial descriptions may produce inefficiencies in citizen stops, or may also supply a pretext for a broader range of stops (see Thompson, 1999, on problems in racial identification of suspects by both victims and police officers; Office of the Attorney General of the State of New York, 1999). These problems—in both the U.S. and the U.K.—may contribute to racial imbalances.

Second, police in the UK record details about the rationale for the stop, specifying one of several grounds in a precoded checklist. Until recently, police in New York used a narrative to describe the stop rationale, and the New York State Attorney General’s report noted that about one in six stops was not justified constitutionally, and another 15% had too little information to be classified as constitutional or unconstitutional. The NYPD recently switched to a precoded form with a proscribed list of circumstances that officers use to classify rationales for the stop. This reporting format makes it extremely difficult for police officers to self-report unconstitutional stops. At least part of this procedural change was motivated by the difficulties in data entry inherent in the older UF-250 forms.19

In both of these systems, police record details about the suspect, the circumstances (including location) and outcomes of the stop, and details of any search (frisk) that was conducted. Officers in the U.K. also provide details about any contraband that is seized, and obtain the suspect’s telephone number, in addition to his or her residential location. Data are recorded on paper and later entered into databases for analysis.

Measures of Street Stop Activity. Similar to traffic stop data, the data generated from the data collection systems described above can produce multiple indicators

---

19 According to statements by Daniel S. Connolly representing the City of New York in National Congress of Puerto Rican Rights v. City of New York, the UF-250 form used until recently to record stop and frisks in New York was “extraordinarily ill designed for data entry” since the card had “an awful lot of narrative.”
of police activity, including both strategic dimensions of policing as well as management and supervision applications. Again, as in traffic stops, racial disaggregation is central to these computations. In addition to this racially disaggregated “epidemiology” of stops, reports of police activity are then computed, including: (1) percent of people stopped who are frisked and/or searched, (2) arrests made, (3) contraband seized (amount and type), (4) property seized, and (5) use of force (if any).

These data are used to compute “hit rates” as measured by arrests made, as well as other indicators of stop outcomes (such as persons frisked). Hit rates can be disaggregated by the alleged penal code violation that motivated the stop, or by the rationale for the stop. Hit rates also are sensitive to department policies that may alter the circumstances and criteria that officers are trained to use for detaining and/or searching a suspect. Recent policy changes in the U.K. emphasizing the quality of arrests (and concentrating on low discretion situations) lowered the volume of arrests resulting from stops (Ramirez et al., 2000, citing Fitzgerald, 1988, in a report issued by the Metropolitan Police). In addition to the policy context, Fitzgerald explains the drop in arrests in part on an “observation” effect produced by the ongoing McPherson investigation.

The Denominator: Supply of Individuals for Stops and Searches

Whether police practices rise to a threshold of profiling or selective enforcement requires tests of the extent of racial disparities, under what circumstances, and the role of race relative to the rate of the targeted behavior(s) in the area under the patrol jurisdiction of the department being examined. Different definitions and legal standards of profiling require different standards and tests. The central question in any of these considerations is whether the stops and searches were made because of race, net of other variables. This “but for” requires precise estimates of the supply of individuals who are engaged in similar behaviors. Walker (2001) calls this the “denominator” problem. Smith and Alpert (2002) call it the “baseline.” Fagan and Davies (2000) call it the “base rate.” Miller (2000) calls it “populations available for stops and searches.”

**Baselines and Benchmarks in Vehicle Stops.** Investigators have approached the baseline question in different ways. Smith and Alpert (2002) usefully distinguish baselines of race-specific rates of populations engaged in the targeted behaviors from benchmark information about the population characteristics generally. Certainly the former is critical for an assessment of selective enforcement. The latter may be helpful in providing background information, but cannot be dispositive in establishing a selective enforcement claim. The “but for” standard in case law requires such baseline data.

In Chávez, for example, public defenders mounted a case control study where a Hispanic and white driver traveled in separate vehicles at the same speed along
Interstate 80. They also used analyses of a database of drivers and computed population estimates of Hispanic drivers using a surname match. The court rejected the validity of the database of field (stop) reports because of extensive missing data and problems in ascertaining its comprehensiveness. The court also rejected the use of census data to describe the supply of persons available for stops, distinguishing it from the (unknown) population of drivers, and the use of survey data whose samples were not relevant to the local situation nor an accurate estimate of Illinois travelers or speeders disaggregated by race.

Traffic studies have relied on methods designed by John Lamberth (1994) to generate estimates of both the demographic characteristics of the overall driver population within a specific area, and the comparative rate of speeding and other violations by race. Lamberth’s studies in New Jersey and Maryland (1994; n.d.) were used in litigation to establish a pattern of racial disparity in vehicular stops. They also used driving “drones” to estimate comparative speeding rates by race. Zingraff et al. (2000) used similar methods, and complemented his study with surveys of the driving habits of North Carolina residents. These ambitious efforts were limited in several ways by the technology used. First, observations of vehicular offenses other than speeding are very hard to do using either observations or videotape. Second, racial classification of drivers is imperfect and complex, and extremely difficult for drivers of Hispanic descent. Third, some cars have windows obscured by tinting. It is common to use multiple raters to verify race, but often raters do not agree. Finally, samples tend to be limited to specific stretches of highway. Whether the same probabilities exist on other stretches of highway is unknown. The addition of multiple sampling units will increase costs and complexities of estimating the baselines, but may help reduce errors in baseline estimates.

The San Jose Police Department (SJPD) used a benchmark of census characteristics and two baselines using crime indicators decomposed by race (Walker, 2001). They navigated around a strict baseline estimate by claiming that minority drivers were more often “exposed” to stops and arrests because they were more likely to encounter police officers in their everyday travels. According to the SJPD, neighborhoods with higher concentrations of minority citizens had more crime and more calls for service, requiring greater concentration of patrol resources. However, the fact of differential exposure to police surveillance does not explain differential stop rates, absent a valid count of race-specific vehicle code violations for speeding or broken equipment.

The SJPD also claimed that vehicles in poor neighborhoods were more likely to have violations because the population was less able to maintain their cars. Thus, the benchmark of population was used not as a reference to assess selective enforcement; instead, population correlates (crime, poverty, and bad cars) were used as substitutes for intensifying patrol and surveillance. A second analysis
used crime rates, disaggregated by police district and decomposed by race, to estimate relative risks of a traffic stop. Accordingly, the race-specific vehicular stop rates were proportionate to the race-specific crime rates. The SJPD offered no evidence to show a correlation between vehicle code violations and crime participation. Their theory was essentially an automotive version of Broken Windows (Wilson & Kelling, 1982).

While the Lamberth model has been accepted by the courts (and noted favorably in cases such as Chavez using other methods), imperfections in its design call for further testing and refinement before the results are replicable.20

**Baselines for Street Stops.** The New York State Attorney General’s study of the stop and frisk practices of the New York City Police Department used a baseline of four different race-specific crime rates in each precinct to establish baselines, and compared rates of race-specific stops that alleged each type of crime. The study estimated race-specific crime rates for each precinct in the year preceding the study period. The populations used for the computation of the crime rates included two (updated) census files, one that had the residential population and a second that included the daytime population estimated from the Census Bureau’s “journey file.” These daytime or working populations were especially important to include since many stops and frisks took place in central business districts. Weights were assigned based on the proportion of stops that took place in the day or night (see also Miller, 2000, for the U.K. estimates of differences between pedestrian and residential populations).

The U.K. Home Office computed race-specific populations in several cities and multiple neighborhoods within cities, adding a temporal frame for sampling to capture hour-to-hour variation in populations available for stop and search. Using a combination of human and video observations, race-specific population estimates were computed for each area. Miller (2000) compared these estimates to census data and found very different ethnic and racial makeup in each area from the observational data, confirming the importance of adjustments for pedestrian populations.

The use of data on actual offending as baselines offers an estimate of the supply of persons by race engaging in crime in a neighborhood or other space. But nearly 70% of the stops in the U.K. and New York studies are “high discretion” stops that are animated by suspicious behavior, not observed crime. Accordingly,

---

using crime rates as baselines leaves unresolved the question of the supply of persons whose behaviors are assessed as “suspicious,” and therefore eligible for a police stop and detention. While techniques for behavioral rating of videotaped interactions are widely used in social science, the very specific cognitive frames that police officers apply in judging behavior of persons may vary significantly from the perceptions of neutral coders. Just as the refinement of observational studies on streets and highways is essential to improving estimates of the supply of persons available for stops, so too does improvement in measuring racial disparities in pedestrian stops depend on parallel developments in observational techniques and classification of suspicious behavior.

Is It Profiling?

Do racial disparities in patterns of street or highway stops by police meet constitutional or social science thresholds to be termed profiling? This section discusses how to assess whether (a) patterns of racial disparities occur by chance or are systematic, (b) “but for” the suspect’s race a stop would have occurred, and (c) the observed patterns are robust and generalizable. There are several considerations in this assessment. One involves the selection of criteria for expressing a pattern of discrimination. A second is the selection of analytic models to estimate race-specific differences in enforcement. And a third is the normative judgments about whether a pattern of stops should meet a conceptually broad or constitutionally narrow definition of profiling to merit legal remedy.

Three Constructions of Discrimination

A valid test of profiling can be constructed around a “but for” standard of race. Whren and other cases permit the use of race in conjunction with other factors. But those other factors may well be correlated with race or even substitutes for race, creating redundancy and masking the role of race in the decision to stop a citizen. There are two ways to conduct a “but for” test to address this problem. One is a case-control model, similar to those used in housing and employment discrimination cases. This design meets the constitutional test of how a similarly situated person is treated.

A different approach to testing this construction of discrimination is whether race is a significant predictor of stops, after controlling for other factors. It requires an offset that estimates the supply of persons available for stops, and controls for other factors that might also affect police decisionmaking. The supply should be specifically constructed to estimate the behavior targeted—driving violations, crime, or other suspicious behavior for pedestrian stops. This use of offsets is important
to control for assumptions about the base rates, or what Smith and Alpert (in press) call the baseline. Sampling strategies should account for the possibility of differences by police administrative units, such as precincts or patrol areas.

But this type of test runs into difficulties in urban areas, for example, where racial residential segregation concentrates minority persons in specific neighborhoods or traffic zones. Surveillance strategies are confounded with crime rates. Here, racial discrimination may be less obvious, and comparative tests are compromised by low exposure of non-whites in minority areas, or the reverse. (It is simpler to test for racial incongruity stops of the sort that motivated (in part) the Terry case.21) The use of offsets that are proxies for the race-specific crime rates in those areas can establish a test that compares actual stop rates to the rate that would be predicted by the estimated “supply” or base rate in that area. This was the strategy used by the New York State Attorney General (1999) in the study of racial disparities in stop and frisk policies in New York City. In other words, this model tests whether the stop rates within a neighborhood differ from the rates that would be predicted from the baseline of targeted behaviors.

A third test might simply compare hit rates, or efficiencies. The New York Attorney General’s Report compared the number of stops that produced arrests by race, and found evidence of statistical discrimination. Gross and Barnes (2001) conducted a similar test for highway stops. These tests bypass the question of who is stopped (and for what reason), and look instead only at outcomes, and make no assumptions about racial animus. These tests do suggest that there are assumptions about efficiencies by race that reflect endogenous considerations by police of the base rate of the targeted behavior. While of interest to policy, these tests may not offer information of probative value in equal protection litigation.

When police scan high volumes of people looking for crimes that have not been reported or have not yet occurred, they select persons to stop using markers or indicia based on their understanding of the base rates of the targeted behavior. This probabilistic estimate is vulnerable to both false positives and false negatives, and the balancing of these errors will determine the threshold of regulation that should be applied to police practices. Ayres (in this volume) points out that these

---

21 I have been stopped twice in the past two years in New York City within two miles of my home, near the Bergen Houses, a housing project with mostly minority residents. I was driving my SUV (a 1995 Pathfinder) and wearing my baseball cap. Neither time was my car searched, in part because everything in a SUV is in plain view. One time the stop was for an alleged seatbelt violation, where the seatbelt was concealed beneath a heavy coat. No citation was issued. The second stop was to verify that the inspection sticker was valid, since it was curled and partially obscured on the interior of the windshield. No citation was issued.
tests are free of omitted variable biases. By examining the distribution of detriments (false positives or unproductive searches) compared to the public benefits produced (true positives), policymakers can estimate the net impacts of police practice. This type of benefits test also bypasses the question of prejudiced assumptions and intent, and is agnostic on the question of the supply of persons for stops (or, the “qualified pool” in Ayres’ language), but only produces a test of disparate impact.

**What Model?**

The selection of statistical models can have obvious impacts on the estimate of the extent and magnitude of racial disparities in policing. Consider two issues that might influence model selection. First, fixed effects ordinary least squares (OLS) regression models may be inappropriate to test for differences in stop rates by race, given nonlinearity both in the distributions of stops or searches, and also in the predictors. Second, individuals and events often are nested within areas that exert specific effects on police decisions to stop. For example, racial incongruity stops may reflect the characteristics of an area in which a person is traveling, and the likelihood of being stopped may be explained jointly by area and individual characteristics. An individual may be subject to stop and search in an area where his or her race is heavily underrepresented, but ignored in an area where that person is in the majority.

In these instances, mixed effects or hierarchical models may be more appropriate than linear models, with a loglinear poisson distribution corrected for overdispersion in areas where stops are less frequent (McCullaugh & Nelder, 1989; Finkelstein & Levin, 2001). Poisson regression is a useful and appropriate method to analyze factors that predict counts of events, and to determine the relationship of these counts to a set of explanatory or predictive variables (Kennedy, 1994; Greene, 1993). The model estimates the expected value of the number of events in relation to the causal factors and other explanatory variables of interest, with estimates of cross-level interactions to account for conditional relationships between individual and area predictors.

In the New York State Attorney General’s report, for example, the question was whether the race-specific count of events (stops) in an area (sub-borough) was predicted by factors that might influence these events (arrest rates, social disorganization variables, and physical disorder variables) using multiple offsets for the race-specific and offense-specific crime rate in that area. The baseline model tested the hypothesis that the race-specific stop count is proportional to the number of arrests in the area. The full model assessed whether factors beyond the arrest count predicted the stop count in the area. To estimate the magnitude of racial disparity, race- and crime-specific exponentiated coefficients were compared. This strategy addressed—in part—the question of effect size and its
bearing on the conclusion of statistical discrimination, and is preferable to relying exclusively on statistical significance for specific categories of race. It is especially useful when there are multiple contrasts when multiple racial groups are subject to police scrutiny and detention. When the ratios of exponentiated coefficients for African-Americans and whites for suspicion of carrying a weapon, for example, exceed two to one, one might reasonably and confidently claim that statistical discrimination exists. It is a much tougher call when the ratio shrinks to 1:1.

An alternate model might consider stops or searches as part of a system of dynamic equilibrium, in which police behaviors and the target behaviors of suspects are endogenous factors that mutually influence each other. Knowles, Perisco and Todd (2001) used an equilibrium model to estimate the expected returns to police officers from conducting searches. Their test for statistical discrimination relied on whether the expected returns were the same for all searches. Analyzing Maryland highway stop data, they found evidence not of discrimination, but of troopers acting to maximize the number of searches that turned up drugs.22 Knowles et al. distinguish between “prejudice” and “statistical discrimination.” Prejudice is a preference for searching one racial group, even when those searches are no more likely to be successful. Statistical discrimination occurs when unbiased police officers search one group more often because the chances of success differ by group. The higher search rate for minority drivers may be “statistical discrimination,” they say, but it more likely is the “unprejudiced” use of race to decide which cars to search. Their logic leads to the strange conclusion that the Maryland state police took race into account by race-neutral means.23 The analysis is incomplete, moreover, because they fail to account for the selection process of who is stopped and therefore available for searches. This omission makes their race-neutrality assumption inherently implausible.

Finally, recent studies of legal decisionmaking have used “identification” models to characterize the process of allocating individuals to different conditions of case processing (see, for example, Manski & Nagin, 1998). Identification models assume that sorting and allocation processes are non-random, and that they reflect an interaction of the characteristics of individuals in the “supply”

---

22 They also assume that motorists who might carry drugs are deterred by the race-specific probability of being searched. They offer no evidence from motorists that these assumptions are valid. This assumption would move the marginal rate closer to the average rate. While this may justify using an equilibrium model, it also sets up an unfalsifiable claim about the effects of deterrence on hit rates.

23 Their logic also implies that there are race-specific differences in which drivers possess drugs, an empirical claim that is inconsistent with several data sources.
with the characteristics of the decisionmakers themselves who select populations for different conditions or responses. This construction of decisionmaking brings a new dimension to profiling studies by integrating organizational and individual-officer factors within decision processes. Studies of police-citizen interactions have generally ignored individual officer effects, despite the variation that is likely to occur between officers. Walker's (2001) emphasis on early warning systems explicitly recognizes that officer characteristics are important, and that these characteristics explain variation in which officers engage in race-based or race-preferenced stop patterns. While of limited use in a litigation framework, the potential for management applications—including future training designs—is quite strong.

**Regulation by Information**

The crisis of race and policing has contributed to a crisis of trust and confidence of minority citizens in the law and legal actors. Even though nonwhites are both frequent victims of crime and beneficiaries of recent crime declines, minority citizens most often are directly the subjects of crime control policies designed to bring them relief (see, for example, Harcourt, 2001; Skolnick & Caplovitz, 2001). Recent surveys show a wide gap between minorities and whites in their levels of trust and confidence in the police (Sampson & Bartusch, 1998; Bureau of Justice Statistics, 1996). By fall 1999, 81% of all Americans in a Gallup Poll said they disapproved of “racial profiling” (Gallup Poll, 1999). Profiling confirms and aggravates feelings of racial grievance against the police among African-Americans and, perhaps to a lesser extent, other minorities. Procedural concerns about policing and police activities are expressed more often and more deeply by minority citizens in their everyday interactions with law and

---

24 In New York City, for example, civilian complaints to the police spiked sharply following implementation of “order-maintenance policing” tactics; more than three in four complaints were made by black or Latino citizens (Skolnick and Caplovitz, 2001, at note 14).

25 For example, more than one in three whites had little confidence in the police, compared to more than half the black respondents. Hispanics are found to fall somewhere between whites and blacks. Similarly, minorities express less confidence and trust in the courts.

26 The poll showed that a majority of Americans believed that racial profiling is widespread and three fourths of black men said they have been stopped by police because of their race. The poll defined profiling as the practice by some police officers of stopping “motorists of certain racial or ethnic groups because the officers believe that these groups are more likely than others to commit certain types of crimes” (Gallup Poll, 1999).
legal institutions (Tyler & Huo, 2002). Perceptions of racial profiling threaten to erode the legitimacy of the criminal, discouraging citizens from participation in the everyday interactions with police that contribute to crime control and public safety. Permitting the police to use race as an indicia of crime imposes what Kennedy terms a “racial tax” on minorities who are subjects of stop and frisk (1997, p. 159). Objections to that tax are the affective component of a second normative objection to racial policing.

Measurement of police actions with an eye toward assessing racial imbalances in policing has a declarative effect for citizens of all colors and racial groups. These effects can restore public perception that the police are in the business of producing justice, and that the production of justice has value that competes well with efficiency and crime control (Moore, 1997). Congress has considered statutes that would mandate and standardize data collection for highway stops and police-citizen interactions more generally.27 President Bill Clinton ordered federal law enforcement agencies to compile race data on individuals questioned, searched, or arrested (Holmes, 1999). Such expressions of normative rejection of profiling provide counterweight to institutionalized practices and efficiency-based justifications. The creation of data and statistics on police actions also creates mechanisms for accountability that can help offset the legitimacy costs of racial imbalances in policing.

As information builds, so too does the potential for the application of information in citizen-police collaborations to regulate and manage racial disparities in policing. Other domains of public policy have struggled with methods to use information for regulatory purposes. The growth of innovations in citizen participation in environmental regulation is one example. “Backyard environmentalism” (Karkainnen, 2001) illustrates how the demand for data—toxic release inventories, for example—can spur new forms of citizen-government interaction to optimize competing interests (such as safety versus efficiency). It is an adaptive form of regulation that recognizes competing interests and capitalizes on them. Implicit in these dynamics is that regulation is information-driven and exchange-driven, with legitimacy payoffs to both citizens and police from participation. Over the past decade, litigation has launched data collection enterprises that offer mechanisms to assist regulation and public policy choices. The spread of these systems is likely to continue through the natural processes of diffusion of innovation. Their applications in management are complemented by their utility for making public policy choices and enforcement of legal standards reject practices that reify racial biases into law enforcement.

References


San Jose (California) Police Department. (1999, December 17). Vehicle stop demographic study. San Jose, California: San Jose Police Department.

Shuster, B. (2000, February 1). Deputies to Test Tiny Videocams Safety: They will wear lens as part of pilot project. Traffic stop tapes will be used as evidence. New York Times, p. B1


Cases

Brignoni-Ponce, 422 U.S. 873 (1975)
Brown v. Oneonta, 251 F.3d 769 (2nd Cir. 1999)
Chavez v. Illinois State Police, 251 F.3d 612 (7th Cir. 2001)
Daniels v. City of New York, 2001 WL 62893 (S.D.N.Y.)
Korematsu v. United States, 323 U.S. 214 (1944)
Martinez-Fuerte, 428 U.S. 543 (1976)
NAACP v. Maryland Department of State Police, 72 F. Supp. 2d 560 (D.Md. 1999)
State v. Dean, 112 Ariz. 437, 543 P.2d 425 (1975)
Terry v. Ohio, 392 U.S. 1 (1968)
U.S. v. Brignoni-Ponce, 422 U.S. 873 (1975)
U.S. v. Harvey, 16 F.3d 109, 115 (6th Cir. 1994)
U.S. v. Weaver, 966 F.2d 391 (8th Cir. 1992)
Whitfield v. Board of County Commissioners of Eagle County, 837 F.Supp.338 (D. Colo. 1993)
Yick Wo v. Hopkins, 118 U.S. 356 (1886)