JUSTICE QUARTERLY VOLUME 22 NUMBER 2 (JUNE 2005)

Testing the Racial Profiling Hypothesis for Seemingly Disparate Traffic Stops on the New Jersey Turnpike

James E. Lange, Mark B. Johnson, and Robert B. Voas

This paper describes two studies designed to produce benchmark values with which to compare police stop data in an effort to assess racial profiling. Racial profiling is often measured by comparing the racial and ethnic distribution from police stop rates to race and ethnicity data derived from regional census counts. However, benchmarks may be more appropriate that are based on (1) the population of drivers or (2) the population of traffic violators. This research surveyed drivers on the New Jersey Turnpike and produced benchmark distributions reflecting these two populations. Benchmark values then were compared to police stops collected from State Troopers patrolling the Turnpike. The results revealed that the racial make-up of speeders differed from that of nonspeeding drivers and closely approximated the racial composition of police stops. Specifically, the proportion of speeding drivers who were identified as Black mirrored the proportion of Black drivers stopped by police. This finding may explain the differences found between police stop rates and regional census data that are often interpreted as evidence of racial profiling. Interpretation and limitations of the results are discussed.

Keywords racial profiling; traffic stops; New Jersey Turnpike

James E. Lange received his doctoral degree in social psychology from the University of Maryland in 1996. He currently is the Coordinator of Alcohol and Other Drug Initiatives at San Diego State University, California. Mark B. Johnson received his doctoral degree in social psychology from the University of Maryland in 2002. He currently is a Research Scientist at the Pacific Institute for Research and Evaluation where he studies college drinking, drinking and driving, and traffic safety. Robert B. Voas is a Senior Research Scientist at Pacific Institute for Research and Evaluation in Calverton, Maryland. He has conducted research on alcohol and traffic safety for more than 30 years. He holds a doctorate in Psychology from the University of California at Los Angeles. Correspondence to: Mark B. Johnson, Pacific Institute for Research and Evaluation, 11710 Beltsville Drive, Suite 300, Calverton, MD 20705-3102, USA. E-mail: mjohnson@pire.org

ISSN 0741-8825 print/1745-9109 online/05/020193-31

DOI: 10.1080/07418820500088952

Introduction

Over the last 10 years, significant events have brought the issue of racial profiling by law enforcement to nationwide public attention (e.g., see Harris, 2002, for a review; see also court cases of *Rodriguez v. California Highway Patrol*, 2000; *State v. Pedro Soto*, 1996; *Wilkins v. Maryland State Police et al.*, 1993). Although evidence from internal documentation (on police training protocols and operating procedures) has been used, along with corroborating statistical data to identify law-enforcement agencies that were guilty of discriminatory practice, methods of appropriately measuring racial profiling remain elusive.

The issues surrounding the assessment of discriminatory police behavior are complex. Ultimately, a police officer's decision to stop a vehicle may stem from observation of an illegal operation or, alternatively, from cues that are not diagnostic in predicting criminal activity. In many cases, the motivating factor for a law-enforcement decision may be subjective and not externally ascertainable. Researchers who examined queries from officers' in-car Mobile Data Terminals were unable to determine whether patterns of treating African American drivers with suspicion were due to observations of race or of behavior (Meehan, 1998).

Understanding the motivating factors in law-enforcement decisions is complicated further by strong empirical evidence that stereotypes can influence judgments subconsciously (e.g., Devine, 1989), that stereotypes can be activated automatically without conscious control (e.g., Bargh, Lombardi, & Higgins, 1988), and that people are not always aware of the cognitive processes that underlie their judgments (e.g., Bargh, 1994). These findings suggest that racial and ethnic factors may affect police officers' behavior without their conscious awareness.

Racial profiling by law enforcement has been measured by comparing the racial distribution of traffic stops (within a jurisdiction or precinct) to a benchmark based on the racial distribution of individuals residing in the region. Deviations between these two distributions indicate that some racial or ethnic groups are being stopped at a rate disproportionate to their representation in the local population; this can be interpreted as evidence of racial profiling.

It is not clear, however, whether regional or local population estimates are the appropriate benchmark for comparison. In the *State v. Kennedy* (1991), evidence of discrepancy between population counts and stop rates was sufficient to raise legitimate questions as to whether police in New Jersey were selectively enforcing traffic laws as a function of driver race. However, in the case ruling, it was stated that a more appropriate benchmark for assessing racial profiling would be to estimate the racial composition of those who exceed the speed limit and compare that to the composition of individuals stopped and cited for that offense. Similarly, an interim report produced by the New Jersey State Police Review team (assigned to investigate allegations of racial profiling by the State Police) defined profiling as including reliance on race and ethnicity in "selecting vehicles to be stopped from among the universe of vehicles being operated in violation of the law …” (Verniero & Zoubek, 1999, p. 5). Only if the racial distribution of traffic violators reflects proportionately the racial distribution of local residences would the benchmark that typically is used to assess racial profiling be appropriate.

Methods of Measuring Racial Profiling

There have been attempts, however, to generate appropriate benchmarks with which to compare police stop data. For example, the data used in *State v. Pedro Soto*, as well as in *Wilkins v. Maryland State Police*, did not rely on population census data as a benchmark against which to measure racial profiling. Rather, the data were based on work by John Lamberth that assessed population estimates of highway drivers as a function of race. Lamberth's first study (see the *State v. Pedro Soto*) involved stationing observers by the side of the road to count the number of cars and the race of the occupants in randomly selected 3-hour blocks of time over a 2-week period. Between June 11 and June 24, 1993, at four sites between Exits 1 and 3 of the Turnpike, the observers counted roughly 43,000 cars, of which 13.5 percent had one or more African American occupants. This was consistent with the population figures for the 11 states from which most of the vehicles observed were registered.

To address more directly the benchmark outlined in *State v. Kennedy*, Lamberth conducted a second study to examine speeding violations as a function of driver race along the New Jersey Turnpike (*State v. Pedro Soto*) and along I-95 in Maryland (see *Wilkins v. Maryland State Police*). Lamberth measured speeding violations by instructing research assistants to drive along a predetermined thoroughfare (e.g., the New Jersey Turnpike or a section of I-95) and record separately the race of drivers who either passed the research vehicle or were passed by the research vehicle. The research vehicle was driven 5 mph over the legal speed limit. Consequently, those who passed the researchers could be categorized as speeding violators. Research assistants also recorded the race of each passing or passed vehicle.

Data from Lamberth's New Jersey Turnpike survey revealed that approximately 98 percent of the vehicles exceeded the 60 mph speed limit on the Turnpike. Further, they found that African American drivers made up 14.8 percent of the speeding violators, which was not significantly higher than the portion (13.5 percent) of African American drivers using the New Jersey Turnpike. Lamberth argued that these data validate the assumption that African American drivers, as a group, do not drive any differently than White drivers.

Other research found results consistent with those of Lamberth. For example, survey data indicate that African American drivers reported less risky driving than White drivers (Wright, Tomaskovic-Devey, & Zingraff, 2000), and were more likely to perceive that their driving behavior did not warrant being stopped (Lundman & Kaufman, 2003). If these results were true, the evidence of disproportionate police stops of African American drivers would more strongly suggest biased police behavior.

However, results from additional studies suggest the contrary. Focus groups of San Diego officers indicated that roughly one quarter of the stops were pretext stops and had nothing to do with traffic violations. It was argued, therefore, that one quarter of police stops would be expected to match the demography of criminal suspects. Accordingly, the San Diego Police adjusted their census-based benchmark so that only 75 percent of the estimates came from population demography, and 25 percent was based on the demography of criminal suspects. Because African Americans were more heavily represented among criminal suspects than in the general population, basing the racial benchmark, in part, on criminal suspects increased the relative proportion of minorities. When this new benchmark was compared against police stop data, the evidence of disproportionate stop rates disappeared (Cordner, Williams, & Velasco, 2002).

In Cincinnati, researchers (Eck, Liu, & Bostaph, 2003) examined contact data that were collected during police stops to test the hypothesis of biased policing. Their analysis indicated that the spatial pattern of police stops in Cincinnati appeared to be correlated with driving patterns, crime patterns, drug calls, and overall demand for police services. This suggested that the disproportionate stops of African American drivers may be explained by workload factors rather than biased policing.

Their data also indicated that the duration of stops for African American drivers was significantly longer than for White drivers; however, this difference was accounted for, in part, by evidence that stopped vehicles driven by African Americans contained more occupants than those driven by Whites. Finally, although vehicles of African American drivers were searched at disproportionately higher rates, the probability of finding contraband was constant across race. Consequently, the researchers indicated that if African American drivers were not stopped arbitrarily, the rates of finding contraband would be expected to be lower.

The Present Research

Following a consent decree between the State of New Jersey Attorney General's Office and the United States Department of Justice (USDOJ) regarding monitoring the State Police for discriminatory behavior, a contract was let to a nonprofit research organization to conduct a survey of drivers on the Turnpike. The purpose for this survey was to establish a benchmark with which to compare police stop rates; it was argued that estimating the racial distribution of Turnpike drivers would produce a better standard of comparison than estimates generated based on local census figures. This survey clearly mirrored the focus of Lamberth's first study (described previously) but extended that research in two important ways. First, the survey was conducted along the entire Turnpike, not just the southern section. Second, as will be described, our survey relied on self-reports of race and ethnicity, thus eliminating measurement error due to unreliable observations.

Shortly after the population survey of Turnpike drivers (hereinafter referred to as the "Tollbooth Survey"; described in the previous paragraph) was completed and the final report submitted, the State of New Jersey again contracted the researchers to conduct a survey of speeders along the Turnpike (hereinafter referred to as the "Turnpike Speed Survey") to produce an estimated racial distribution of traffic-speed violators. Using this distribution as a benchmark would more closely accommodate the definition of an appropriate benchmark as established in the *State v. Kennedy* and as defined in the New Jersey State Police Review Team's interim report. The purpose of this Speed Survey clearly mirrors that of Lamberth's second study; however, ours differed in that we used a different definition of "speeder," surveyed a broader sample, and used a substantially different methodology.

This research involved data from two studies and included a total of eight data sources. The data sources that comprise the Tollbooth Study and the Turnpike Speed Study are detailed in Table 1.

Tollbooth Survey

Overview

This study involved two strategies: one to assess the race and ethnicity of driving participants, and the other to create reliable and generalizable

Table 1 Overview of data collections for New Jersey Turnpike studies

	Dataset	Data-collection specifications
Tollbooth Survey	Interview survey of drivers at Turnpike exits	May 2000; 4,656 drivers at 48 sites
	Total count of vehicles exiting toll-way at each exit	May 2000; 201,313 cases at 48 sites
	Police stops during Tollbooth Survey	May 2000; 7,559 cases
Turnpike Speed Survey	Radar-triggered camera	March 31–June 30, 2001, 14 locations; 15,046 usable cases
	Randomly triggered camera	March 31–June 30, 2001, 14 locations; 11,288 cases
	Population counts of nonspeeders	March 31–June 30, 2001; 430,554 cases
	Population counts of speeders	March 31–June 30, 2001; 17,010 cases
	Police stops during speed survey	March–June 2001; 29,486 cases

population estimates for different sections of the Turnpike. Because the survey took place on the Turnpike at tollbooths where drivers must exit, we had face-to-face contact with participants. Our research assistants, positioned in tollbooths, collected self-reported racial and ethnic information from participants, which benefited the study by reducing measurement error (relative to observational approaches), particularly for Hispanic drivers. We also obtained total vehicle counts on the Turnpike by monitoring the total flow of traffic as it exited at the tollbooths. These counts were used to weight our sample cases appropriately.

Participants

We contacted 4,656 drivers at the tollbooths, and 4,039 (86.8 percent) agreed to participate. For those who did not agree, interviewers estimated the driver's race/ethnicity and age based upon observation. Because drivers who refused often spoke with the interviewer or the toll collector, vocal characteristics could be used for the estimation. These nonparticipants (with estimated demographics) were included in the analysis. Our decision to include nonparticipants was made a priori and was based on the assumption that refusals may differ significantly from participants. We felt that excluding refusals might introduce more bias than would be introduced by using interviewers' estimates. We have no data regarding the accuracy of interviewers' observations, and although the refusal rate was relatively small (less than 14 percent), this is one limitation of the research.

Procedure

Sampling structure The data-collection structure first was stratified as a function of weekend versus weekday under the assumption that the demographics of drivers might differ according to day of week. Second, stratification occurred according to the geographic segment of the Turnpike. Three segments—South (up to the 56-mile post), Central (up to the 97-mile post), and North (up to the end of the Turnpike at 118 miles)—were used. These segment boundaries reflected the different jurisdiction boundaries of New Jersey State Highway Patrol units assigned to the Turnpike. Finally, the data were stratified according to time of day. Four 6-hour time blocks were chosen: 4 a.m. – 9:59 a.m.; 10 a.m. – 3:59 p.m.; 4 p.m. – 9:59 p.m.; 10 p.m. – 3:59 a.m. The choice of timeframes for these strata helped to ensure that rush-hour periods would be contained within a single shift. Because the time-of-day strata spanned across days, weekend days were defined as beginning Saturday at 4 a.m. and ending Monday at 3:59 a.m. All other times were considered as a weekday.

Within each level of the 24 levels of stratification (2 weekend/day × 3 geographic segments × 4 time block), two exit tollbooths were randomly sampled (without replacement). Random selection of the exits with levels of stratification was accomplished through a random number generator in the SPSS statistical package. Between-strata exits could be selected more than once.

Field procedure At the selected Turnpike exits, surveyors were stationed at two tollbooths, positioned behind the toll collector. When there were more than two booths in operation at a given exit, the booths were selected at random. There are no "exact change" lanes along the Turnpike, and the "speed-pass" lanes had not yet been fully implemented.

A timing mechanism within surveyors' handheld PCs prompted them when to recruit vehicles for participation. After each successful interview, the interview program would go into a timeout mode for 2.5 minutes. At the end of that period, the program would instruct the interviewer to contact the third vehicle to arrive at the tollbooth, so drivers were selected before they arrived at the booth, without regard to their ethnic or racial identification. State vehicles and vehicles with more than four wheels were excluded from the sample. This produced an unbiased sample of drivers at a rate of approximately 10 per hour (if traffic flow permitted).

Drivers of selected vehicles would hand the toll collector their Turnpike ticket, and then the surveyor would lean out the booth window and offer the driver $5 to answer a few brief questions. The driver was assured that the survey was voluntary and confidential. A $5 bill was placed inside an envelope that had a contact telephone number on it so drivers could learn more about the survey if they wished.

Participants first were asked to indicate at which exit they entered the Turnpike and their age. They were then asked to indicate their ethnic/racial group from choices of White, Black, Hispanic/Latino, Asian, American Indian, or Other (repeated verbally). Participants were told that they could choose as many as applied. Surveyors recorded participants' sex from their own observations. Next, surveyors asked follow-up questions. Participants not indicating a Hispanic or Latino identification were asked whether they also considered themselves to be Latino or Hispanic, and participants indicating they were Latino or Hispanic were asked whether they considered themselves to be White, Black, Asian, American Indian, or Other. Finally, participants who reported an Asian identification were asked to indicate whether their country of ancestry was from the Far East (e.g., Japan, China, or Korea), from the Subcontinent (e.g., India, Pakistan, or Bangladesh), or Any Other.

At the end of each interview, the interviewers recorded whether the vehicle had commercial markings and the license plate number, state of registration, exit number, and time. For those who did not participate, age and race/ethnicity were estimated and recorded by the interviewers. The computer program did not permit cases to have missing gender or race/ethnicity; however, it did provide an "Other" category for those who responded to the race/ethnicity question with a national origin or religious affiliation. It also provided a "comment" area where interviewers were instructed to record unusual circumstances or what was said when they checked "Other" as a response.

Population counts The sampling structure dictated that the analysis consider (a) the proper weighting of cases to provide appropriate population parameters and (b) the effect on the estimation of variance to compute the confidence intervals of the population estimates. Population counts of Class 1 vehicles paying tolls at exits during the survey times were obtained directly from the Turnpike Authority. With both the population size and the sample size known, the case weight was calculated as the inverse probability of vehicle selection.

Police stop data Data of police vehicle stops were obtained from the New Jersey State Police through the state's Open Public Records Act. As dictated by the Consent Decree between the State of New Jersey and the USDOJ, law enforcement personnel are mandated to document the demographic characteristics of individuals subjected to police stop and search. At the onset of each vehicle stop, officers are required to contact the state law-enforcement communication center and provide a description of the stopped vehicle and its occupants, including the number of occupants, their apparent race/ethnicity, and their apparent gender. The information provided to the communication center is compiled into a database of police stops.

We obtained records of police activity along the New Jersey Turnpike during May 2000 (the period during which the Tollbooth survey was conducted). For each vehicle stop, the race of the driver (White, Black, Hispanic, Asian, and American Indian), the time of the stop, and the location along the Turnpike were recorded. This dataset did not distinguish between moving and nonmoving violations; stop data did not include expanded categories (for type of violation) until July 21, 2000. Only traffic stops along the Turnpike were included. This dataset included 7,559 cases of which 7,296 contained sufficient information to identify the geographic segment in which the stop occurred.

Analytic procedure

Race/ethnicity categorization As the interview protocol permitted multiple responses for the race/ethnicity items, participants were categorized using the following scheme. Those indicating Hispanic/Latino identification were classified as Hispanic/Latino, and then further subclassified by other racial identifications such as White, Black, Asian, American Indian, or a mix of multiple identifications. These subclassifications remained separate from non-Hispanic/

Latino participants. Participants not identifying themselves as Hispanic/Latino were classified by their racial identification only, which again could be a mix of any of the responses other than Hispanic/Latino.

The Other category was used primarily for the race/ethnicity question when drivers gave a national origin (such as Italian) or religious affiliation. Therefore, the Other responses were identified only by their co-occurring major categorization. For instance, if the response was White and Catholic, the interviewer recorded White and Other. For this study, the driver was only included in the White category. Respondents indicating no additional major race/ethnic categorization were classified simply as Other.

Appendix A shows the racial and ethnic composition (using broad racial/ethnic categories) between participants (self-report) and refusals (interviewers' estimates). Although the proportion of White and Hispanic drivers was reasonably consistent between self-report and interviewer estimates, the data show that refusals were less likely to be categorized as Black and more likely to be categorized as Other. However, because refusals comprised a relatively small proportion of the dataset, the racial composition for the total dataset (which was used in analyses) was quite comparable to that from those participants who reported their racial and ethnic identification.

Weighting and estimation Within each level of strata, a weight was computed by dividing the population count for that stratum by the sample count for that stratum. The weight was further expanded to account for the probability that the particular primary sampling unit (PSU) was sampled within the strata as well as the number of strata in the design. Obviously, because the weekend/weekday strata included unequal days, the weight was calculated using the number of days within those strata. The resulting weighting system allows for parameters that are generalizable to all drivers, across the entire Turnpike, at all times of day and night during the survey. The margins of error were calculated using the sampling error estimated through a Taylor series estimation procedure that included the sampling design (i.e., the strata and clusters), thus correcting for their effects on the estimate. The statistics package SUDAAN (version 8.0) was used for this procedure.

Turnpike Speed Survey

The purpose of the Turnpike Speed Survey was to determine whether the racial composition of drivers varied as a function of speed. For the Turnpike Speed Survey, we measured the speeds and captured high-resolution photographs of a sample of vehicles on the Turnpike. Trained coders examined each photograph to determine the race or ethnicity of the driver, and vehicle speeds were used to determine whether the driver was a speeder or nonspeeder. Total counts of vehicles were used to weight the cases.

Sample

A total of 38,745 photographs were taken: 21,536 were nonspeeders, and 17,209 were speeders. Of these, 15,046 and 11,288 cases of nonspeeders and speeders, respectively, produced usable and reliable racial/ethnicity judgments. Speeders were explicitly oversampled in the research. Demographic features of the drivers in these photographs are the focus of the analysis, and sample characteristics are detailed in the results.

Equipment

The digital photographs were captured by a TC-2000 camera system, integrated with an AutoPatrol PR-100 radar system, provided by Transcore, Inc. The equipment, other than two large strobe lights, was mounted inside an unmarked van, parked behind pre-existing guide rails along the Turnpike. The camera and radar sensor pointed out of the van's back window toward oncoming traffic. The two strobe lights were mounted on tripods behind the van and directed toward oncoming traffic. Transcore's employees operated the equipment.

Procedure

Sampling structure Four hierarchically organized strata were included in the design: (1) weekend versus weekday, (2) sampling location, and (3) sampling time, and (4) random versus speed-triggered sampling. Each of the 14 camera locations was indicated by the milepost along the Turnpike and the camera's angle of view (northbound traffic or southbound traffic). Approximately 48 hours of data images were collected at each of the 14 locations along the Turnpike between March 31 and June 30, 2001. Images were taken at each location on a weekend and weekday. No sampling occurred on holidays.

Two sampling systems—one random and one speed-triggered—operated concurrently within each geo-temporal stratum. The purpose of using this dual-sampling method was (1) to produce a random sample of nonspeeding vehicles (those traveling slower than 15 mph over the limit); and (2) to oversample speeders to ensure that enough data were captured for sufficient statistical power. We did not know the actual proportion of speeders in the population, and a simple random sample of vehicles may have included too few speeders to generate reasonable margins of error. Because cases eventually would be weighted to reflect population counts, oversampling did not threaten or bias the integrity of the study's results.

For random sampling, the camera operators were instructed to press a button 25–50 times per hour that would set the camera to capture an image of the next vehicle detected, regardless of speed. They pressed the button without looking at the traffic flow and, generally, did not know which vehicle would ultimately trigger the camera: essentially, this was a random sample of vehicles traveling on the Turnpike. Additionally, for speed-triggered sampling, the camera equipment was set to capture images of all vehicles traveling at or faster than 79 mph[1] in areas with a 65 mph speed limit, and 70 mph in areas with a 55 mph speed limit. The speed-triggered system was designed to capture close to a census of all speeders.[2]

Our decision to define speeders as those traveling 15 mph or more above the posted speed limit was based on discussion with representatives of the New Jersey State Police. A conclusion was reached that 15 mph above the limit reflected a speed at which most State Troopers would initiate a traffic stop. Although most of the analyses presented in the paper treated speeding dichotomously (speeder vs. nonspeeder), we do present an analysis where vehicle speed is treated as a continuous variable (i.e., all speed-values are represented).

Population counts. The radar system counted and recorded the speed of each vehicle that passed. Therefore, population counts of speeders (15 mph over the limit) and nonspeeders could be computed for each level of stratum. Within each stratum, a separate weight was created for speeders and nonspeeders.

Analytic procedure

Data reduction Initially, two panels of three raters examined the still photographs. Raters were selected from a pool of applicants on the basis of a trial prescreening procedure. This prescreening involved potential raters watching videotapes from the Tollbooth Survey (without sound) and surmising the race of the driver. Applicants whose assessments more closely matched drivers' self-report measures were given the position. The pool of selected raters was racially and ethnically diverse.

Raters worked independently, with each panel assigned half of the images. The AutoScan Image Review Station software (provided by Transcore, Inc.) displayed the images in a large window on the screen. Raters could control the zoom of the image, allowing for enlargement of the driver's face, and select from on-screen rating categories the driver's characteristics and the vehicle type. The speed of the vehicle was masked, thus keeping the rater blind to speeder status.

1. Although speed triggering on the camera was set to 79 mph, only those traveling at or faster than 80 mph were used in our analyses. The 79 mph cases were collected only as a safeguard against the possibility that too few cases would be gathered using the predetermined 80 mph criterion. Using 80 mph as the criterion, we obtained sufficient data to allow for estimates we deemed adequately precise. Consequently, we dropped all speed-triggered cases traveling at 79 mph.

2. The speed-triggered system was designed to capture data from all speeders. By chance, some speeders (801 cases) were captured instead by the random-sampling system. These cases, however, were moved in the speeder sample for analysis. In addition, some speeding vehicles may not have been photographed by the speed-triggered system because of technical limitations of the camera system; for example, the camera system may not have captured both speeding vehicles if two were driving side-by-side.

Measures Raters classified each case on several variables including gender, age, race, and Hispanic ethnicity. Race allowed for four options: White, Black, Other, and Indeterminable. Hispanic ethnicity was judged separately from race. When two or more raters concurred that a driver was Hispanic, for purposes of analyses, the driver's race was reclassified from White, Black, or Other to Hispanic. Raters were instructed to rate non-Black Hispanic drivers as being of White race and Hispanic ethnicity. We anticipated that the raters would have great difficulty identifying Hispanic ethnicity for any racial group other than White.

The *vehicles* variable allowed for six options: truck, commercial, police, automobile, motorcycle, and RV. The four levels of the *age* variables were younger than 25, 25–45, older than 45, and indeterminable. In addition to the observational elements, strata and time were recorded, as well as the posted speed limit for the areas.

Coder training

We recruited from a broad pool of applicants with few predefined qualifications. Following initial recruitment efforts, we screened applicants based upon their ability to classify drivers properly. A sample of driver images was randomly selected from the videotaped records of our Tollbooth Survey. Because these cases had known self-identified racial and ethnic classifications, we used the concurrence of the raters with these self-reports as a major hiring criterion. To train the selected raters, we used additional images from the videotaped surveys. Through group discussion of images with known self-identification, we made the process of racial and ethnic classification as systematic and uniform as possible for the visual cues to be used.

Rater reliability

We treated cases with at least two identical ratings (out of three raters) as conclusive, and those without two identical ratings as unreliable and, therefore, unclassifiable. In addition, some cases were reliably identified as unusable (i.e., at least two of three raters agreed that sufficient information to make a code categorization could not be extracted from the photo). Often, this occurred because glare on the windshield from the sun obscured a view of the vehicle interior.[3] If a case was unreliable, or reliably judged to be unusable, it was removed from the working dataset.

Presumably, rater biases or errors, as well as unusable and unreliable cases, were equally distributed between speeders and nonspeeders. If so, rater-related biases and errors cannot produce a *false* conclusion that racial or ethnic differences in speed violation rates exist, assuming no such differences actually exist

3. As will be described later in the paper, unusability can be explained, in part, by the position of the camera relative to the lane of vehicle that was photographed.

(Type I errors). The most harmful effect this error could produce, given the design used, is a false conclusion that *no* racial or ethnic differences in speed violation rates exist, assuming such differences actually exist (Type II errors). In other words, rating errors may cause us to miss actual racial or ethnic differences that do exist but will not cause us to observe differences that do not exist.

Weighting and estimation

Whereas the speed-triggered sampling system was designed to capture a photograph of nearly every vehicle that was detected speeding, the random-sampling system captured only 25–30 photographs per hour—a small fraction of the total numbers of vehicles that drove by. Because the speeders that were captured by chance using the random-sampling system were moved into the speeder sample, the random-sampling system contained only nonspeeders.

Population counts of speeders (15 mph over the limit) and nonspeeders were computed for each level of stratum. After eliminating unusable or unreliable cases from the dataset, weights for the remaining cases were calculated by dividing the population count within each stratum by the sample count within each stratum. For cases sampled during the weekend, the weight was multiplied by 2. The weight for cases sampled during weekdays was multiplied by 5. Thus, a weighted aggregate of cases would project to the population counts during a week.

For example, within a given stratum, if we photographed 25 race-reliable speeders (i.e., 15 mph above the posted limit) and ascertained from the separate radar system that 50 speeding vehicles in total passed through that stratum, those 25 cases would be assigned an initial weight of 2. Similarly, we may have photographed 50 race-reliable nonspeeders during that same stratum and counted 500 nonspeeders in total. The initial weights for these nonspeeding cases would be 10. If the data collection happened to fall on a weekend, the initial weights would be multiplied by 2; if data collection occurred on a weekday, they would be multiplied by 5. When applied to our sample data, these weights will allow us to estimate counts of speeders and nonspeeders expected in the traffic population, even if the proportion of speeders and nonspeeders in our sample is biased.

The margins of error were calculated using the sampling error estimated through a Taylor series estimation procedure. This procedure includes the stratification structure, thus correcting for their effects on the estimate. The statistics package SUDAAN (version 7.5.2) was used for this procedure.

Police stop data

We obtained records of police activity along the New Jersey Turnpike from March through June 2001 (the period during which the Speed Survey was conducted). For each vehicle stop, the race of driver (White, Black, Hispanic, Asian Indian, Other Asian, American Indian, and Unable to Observe), nature of the violation (moving vs. nonmoving violation), time of the stop, and location along the Turnpike were recorded. Non-Turnpike traffic stops were excluded. This dataset contained 30,570 cases, of which 29,486 cases contained sufficient information to identify the Turnpike segment in which the stop occurred. The analysis of police stop data used only moving violations, reducing the sample size of usable cases to 27,691. The dataset, which was provided by the New Jersey State Police, did not differentiate among types of moving violations (e.g., speeding).

Results

Data Reliability

Racial categorization

Our greatest concern over reliability of racial categorizations came from the Turnpike Speed Survey, where raters surmised driver race and ethnicity. We minimized the potential impact of interpretation unreliability by using a comparison group (i.e., nonspeeders) with which to compare speeders; we assumed that the influence of unreliability would be equal between both groups. However, there was a statistically reliable overrepresentation of unusable data in the speeder category, Wald (1) = 86.5, $p < .01$ (odds ratio for usability by speeders = .72). For nonspeeders, 69.9 percent of the cases were judged as reliable, 7.8 percent were unreliable, and 22.3 percent were unusable. For speeders, 65.6 percent of the cases were reliable, 9.2 percent were unreliable, and 25.2 percent were unusable. Overall, 68.0 percent of cases were reliable, 8.4 percent were unreliable, and 23.6 percent were unusable.

Although this would raise concerns about the validity of the remaining data if usability were also related to race/ethnicity, we have reason to believe it is not. Largely, the lane in which the vehicle was traveling and the position of the camera and strobe lights determined image usability. When the equipment was placed on the right shoulder of the Turnpike, vehicles in the far left lane (the fast lane) were farthest from the camera and strobe lights. Consequently, the images of the driver were often poorest for the fast lanes.

Beginning with sampling in the Central segment, we identified some areas on the Turnpike where the equipment could be safely placed in the median. Under these circumstances, images of drivers where the left lanes were closest to the camera and strobe lights could be obtained. An analysis of the Central segment was conducted to determine if equipment placement and lane position predicted usability. Controlling for vehicle lane position, equipment location, and the interaction of lane and equipment position completely eliminated the difference in usability between speeders and nonspeeders, Wald (1) = .23, $p = .63$ (odds ratio for usability by speeders = .98).

In the Southern segment, the lack of variability in equipment position meant that the analysis could not be conducted as it was for the Central segment data. However, lane position alone did somewhat attenuate the effect of the speeder variable on usability, indicating again that the technical aspects of the sampling unrelated to driver characteristics are most likely producing the effect (the odds ratio for usability by speeders changed from .51 to .61). There is no reason to believe that Black drivers and White drivers who traveled in the same lane would differ in windshield glare. Thus, we have no reason to be concerned that the elimination of unusable cases from the dataset has substantively biased the results.

Appendix B describes in greater detail the sample counts of race-reliable, unreliable, and unusual cases as a function of sampling location along the Turnpike. In addition, Appendix C provides sample counts of racial and ethnic categories based on (1) a 2-out-of-3 rating criterion, as well as (2) a unanimous 3-out-of-3 ratings criterion. Although approximately 32 percent of cases were unreliable regarding race or ethnicity using the 2-out-of-3 criterion, 60 percent of cases were unreliable using a unanimous criterion.

Age reliability

Raters were asked to code each driver's age using three categories: younger than 25, 25–45, and older than 45 years. We again used the concurrence of two or more raters to assign the variable levels to cases; however, we found very few cases where raters agreed upon the younger age categories. Because of this finding, we recategorized the coding system to be a two-level variable: 45 and younger and older than 45. Using this system, 96.6 percent of the useable cases had reliable age estimates. No cases were eliminated from the dataset based upon the lack of a valid age estimate (e.g., for weighting purposes); however, cases with missing age estimates were not used in any analysis that required the age variable.

Survey reliability

The Tollbooth and Speed Surveys used very different methods for assessing driver race: the former relied on self-report, whereas the latter was based on visual observation. Nevertheless, we were interested in comparing the population estimates from each survey. Convergent evidence between the two surveys would contribute to our confidence that our procedure for coding the still photos from the Speed Survey was adequate.

For this comparison, both speeders and nonspeeders from the Speed Survey were included to create population estimates of all drivers traveling on the Turnpike. Only vehicles rated as an *automobile*, whether commercial or private, were included in the comparative analysis to make the data more comparable between the studies. In addition, we recategorized drivers' race/ethnicity from the Tollbooth Survey to reflect purely visual categories. For example, in the Speeder Survey, it was difficult to determine accurately the ethnic origin of Black drivers. Therefore, it was reasonable to consider the Black category irrespective of Hispanic ethnicity to create an equivalent category between the two studies. Similarly, we aggregated White drivers, irrespective of Hispanic ethnicity.

When using visually based racial categories, we found no statistically significant differences in the proportion of Black and White drivers between the Tollbooth Survey and the Turnpike Speed Survey. For the Tollbooth Survey, 74.0 percent ± 3.4 of Turnpike drivers were visually White; this estimate from the Turnpike Speed Survey was 74.8 percent ± 0.9. For the Tollbooth Survey, 14.2 percent ± 2.1 of Turnpike drivers were visually Black; this estimate from the Turnpike Speed Survey was 15.6 percent ± 0.7.This finding suggests that our strategy for categorizing the race of drivers from our Turnpike Speed Survey produced valid results. It also indicates that the sampling systems used in both surveys yielded substantively equivalent population parameters, irrespective of methodological idiosyncrasies within each survey.

Primary Analyses

Race, speeding, and police stop rates

Using the data obtained through the previously described methods and sources, we present a series of population estimates representing racial/ethnic differences among (1) exiting Turnpike drivers; (2) nonspeeders along the Turnpike; (3) speeders (defined as 15 mph over the limit) along the Turnpike; and (4) police traffic stops during the periods of the Tollbooth (moving and nonmoving violations) and Turnpike Speed Surveys (moving violations only). Estimates are presented separately for Southern, Central (65 mph) and Northern (55 mph) segments of the Turnpike. These data allowed us to examine whether the New Jersey State Police stopped members of racial and ethnic groups proportional to their overall representation on the Turnpike and whether the police stopped members of racial and ethnic groups proportional to their representation among speeding violators on the Turnpike.

Table 2 depicts the racial /ethnic composition of drivers in each geographic segment of the Turnpike from several data sources: the Tollbooth Survey;[4] nonspeeders from the Speed Survey; speeders from the Speed Survey; and

4. The Tollbooth Survey recorded the entry and exit locations along the Turnpike for each vehicle case (entry location data was obtained from the vehicle tollbooth ticket). We were able to identify the span of the Turnpike across which each vehicle traveled. The Tollbooth Survey estimates presented in Table 5 reflect the fact that the same vehicle may have traveled through more than one geographic segment, so these estimates are not based simply on the segment in which vehicles exited the Turnpike.

Table 2 Racial/ethnic distributions from the Tollbooth Survey, Turnpike Speed Survey, and police stops

	Tollbooth Study	NJ Turnpike police stop data	NJ Turnpike Speed Survey: Nonspeeders	NJ Turnpike Speed Survey: Speeders *(15 mph+)*	NJ Turnpike police stop data
	May 2000	May 2000	March–June 2001	March–June 2001	March–June 2001
Southern Segment					
White (%)	65.9 ± 3.8	56.6	71.2 ± 1.3	58.3 ± 3.0	51.7
Black (%)	15.1 ± 2.6	28.3	16.0 ± 1.1	26.0 ± 2.6	28.8
Hispanic (%)	10.7 ± 3.0	8.0	6.8 ± 0.8	7.0 ± 1.4	9.6
Other (%)	9.2 ± 2.5	7.1	6.0 ± 0.7	8.6 ± 1.7	9.9
Central Segment					
White (%)	63.9 ± 5.0	62.8	70.7 ± 1.3	60.2 ± 2.4	56.6
Black (%)	12.5 ± 2.4	21.5	15.9 ± 1.0	25.6 ± 2.1	23.0
Hispanic (%)	12.8 ± 2.3	9.5	4.5 ± 0.6	4.6 ± 1.1	11.9
Other (%)	10.8 ± 1.9	6.2	9.0 ± 0.8	9.6 ± 1.4	8.5
Northern Segment					
White (%)	58.7 ± 4.2	63.6	65.4 ± 2.8	68.4 ± 1.2	60.3
Black (%)	13.1 ± 2.6	17.0	18.0 ± 2.2	18.2 ± 1.0	16.4
Hispanic (%)	16.6 ± 2.7	13.1	7.2 ± 1.6	6.0 ± 0.6	14.7
Other (%)	11.7 ± 2.2	6.3	9.4 ± 1.6	7.4 ± 0.6	8.6

police stop rates collected during May 2000 and from March through June 2001. The table is arranged so that population estimates from the Tollbooth Study can be easily compared to police stops from the same period (May 2000). Furthermore, population estimates of nonspeeders and speeders (15+ mph over the limit) collected between March and June 2001 can be easily compared to police stop rates collected during the same period.

Examination of the data from the Tollbooth Survey reveals that in the Southern and Central segments of the New Jersey Turnpike, Black drivers were overrepresented among police stops. In the Southern segment, 15.1 percent (± 2.6) of drivers indicated that they were Black, yet 28.3 percent of police stops involved Black drivers. In the Central segment, 12.5 percent (± 2.4) of drivers indicated that they were Black, yet 21.5 percent of police stops involved Black drivers. In the Northern segment, however, the extent to which Black drivers were overrepresented among police stops was small relative to other segments.

The data from the Speed Survey, however, revealed differences in the representation of Black drivers among speeders and nonspeeders. In the Southern segment, 16.0 percent (± 1.1) of nonspeeders were Black (this approximates the proportion observed in the Tollbooth Study). However, among individuals driving 80+ mph in a 65 mph zone, 26.0 percent (± 2.6) were Black. A similar pattern was observed in the Central segment: Black drivers made up 15.9 percent (± 1.0) of nonspeeders but 25.6 percent (± 2.1) of speeders. In the Northern segment, however, there was no such overrepresentation (18.0 percent ± 2.2 vs. 18.2 percent ± 1.0).

The Tollbooth Study and Speed Survey used different data-collection methods to capture different populations of interest. These studies also differed in the nature of their comparison police stop data. The stop data that were compared with the Tollbooth Study contained both moving and nonmoving violations, whereas the stop data contrasted against our Speed Survey data contained only moving violations. This may be particularly relevant in light of suggestions that minority drivers are more likely to be stopped for minor nonmoving violations.

However, the racial composition of police stops from the Tollbooth Study stop data (May 2000) and from the Speed Survey data (March–June 2001) is quite comparable. The exception appears to be that Hispanic drivers in the Central and Northern segments represent a higher proportion of stops when nonmoving violations are considered. If the racial composition of police stops for moving violations during March 2000 involves fewer minorities than what is reported in Table 2, then the degree to which minorities are stopped disproportionately (according to the tollbooth data) would be even larger. The discrepancy between the Tollbooth and Speed Surveys would be exaggerated as well.

Speeding rates and police stop rates by time of night

Speeding rates may differ across time of night due to lower traffic density during the late-night and early-morning hours. If racial/ethnic distributions differ as a function of time of day as well, this confound may account for the overrepresentation of Black drivers among speeders. In Table 3, we display racial/ethnic percentages of nonspeeders, speeders, and police stops for Black and White drivers in each of six 4-hour time blocks. Separate tables are provided for the Southern/Central (65 mph zone) segment and the Northern (55 mph zone) segment.

The data show differences as a function of time of day. Black drivers make up relatively higher percentages of drivers—particularly speeders—during the late-night and early-morning hours, and are overrepresented. White drivers make up relatively high percentages of drivers during morning, afternoon, and evening hours. The distribution of Black and White speeders over time approximates closely the pattern of police stop rates. As illustrated in Table 3, the time periods where Black drivers are highly overrepresented among speeders are the same time periods where Black drivers are highly overrepresented in police stops. These data do not suggest that differences in time of day explain the overrepresentation of Black drivers among speeders.

Table 3 Black and White drivers from the Turnpike Speed Survey and from police stops by time of night

		Percentage nonspeeders	Percentage speeders	Percentage police stops
Southern/Central (65 mph)				
Midnight–4 a.m.	White	56.9	39.5	35.8
	Black	27.7	40.63	41.3
4 a.m.–8 a.m.	White	69.8	54.5	51.9
	Black	17.4	30.92	28.2
8 a.m.–12 p.m.	White	76.9	71.8	56.2
	Black	11.7	19.0	24.0
12 p.m.–4 p.m.	White	80.1	68.2	58.0
	Black	10.8	22.7	23.8
4 p.m.–8 p.m.	White	80.1	68.2	58.0
	Black	13.2	18.6	23.7
8 p.m.–Midnight	White	58.9	48.4	49.2
	Black	23.0	30.7	28.2
Northern (55 mph)				
Midnight–4 a.m.	White	54.5	59.2	49.3
	Black	27.4	22.7	24.0
4 a.m.–8 a.m.	White	64.3	69.8	58.8
	Black	14.2	17.8	16.6
8 a.m.–12 p.m.	White	68.7	73.4	60.9
	Black	19.1	17.2	16.0
12 p.m.–4 p.m.	White	70.8	71.6	62.4
	Black	13.6	17.8	15.3
4 p.m.–8 p.m.	White	67.3	73.6	65.2
	Black	18.0	17.7	15.0
8 p.m.–Midnight	White	62.8	65.1	56.8
	Black	23.0	23.6	17.0

Speeding and police stop rates

If police patrolling the New Jersey Turnpike do stop drivers in an unbiased and nondiscriminatory manner, then the racial/ethnic distribution of the police stop data should approximate the proportions of speeding violators. This pattern emerges when we compare the data from the speeders in the Speed Survey with the police stop data. In the Southern segment, Black drivers made up 26.0 percent (±2.6) of speeders (80+ mph in a 65 mph zone) and 28.8 percent of police stops for moving violations. In the Central segment, Black drivers made up 25.6 percent (±2.1) of speeders and 23.0 percent of police stops. In the Northern segment, Black drivers made up 18.2 percent (±1.0) of speeders but only 16.4 percent of police stops. These data suggest that police stops approximate what we might expect if police were stopping vehicles in an unbiased manner.

The data also suggest Hispanic drivers are overrepresented in police stops, particularly in the Central and Northern segments. However, we are less confident in the validity of our identification of Hispanic drivers because of the divergence between the Speeder and the Tollbooth Surveys.

Race and speed relationship

A series of analyses were conducted to better understand the nature of the relationship between race/ethnicity and speeding. Given that there appears to be a rather uniform race/ethnicity effect across all parts of the Turnpike where the speed limit is 65 mph, these analyses were conducted by Turnpike speed limit, not by Turnpike segment. These analyses revealed that the majority of drivers did not drive above the criterion used to define speeder. This is true for all racial/ethnic categories (see Table 4). Further, we found that the average (mean) speed for each racial/ethnic group of drivers is very similar. This implies that, in general, the average driver is very similar for each racial/ethnic group.

However, Tables 2 and 3 clearly show that at the extreme high ends of the speed distribution, Black drivers are overrepresented in 65 mph zones. These findings are confirmed by logistic regression analysis (see Table 5) that found that Black drivers were 64 percent, and drivers rated as Other are 18 percent more likely to speed (as defined by traveling 15 mph or more above the posted

Table 4 Proportion of drivers defined as speeder and average speed

Turnpike speed limit	Race/ethnicity	Percentage nonspeeder	Percentage speeder	Percentage average speed
65 mph	White	98.6	1.4	66.3
	Black	97.3	2.7	66.8
	Hispanic	98.2	1.8	66.3
	Other	98.1	1.9	66.2
55 mph	White	86.5	13.5	62.6
	Black	86.9	13.1	63.3
	Hispanic	89.0	11.0	61.3
	Other	89.5	10.5	62.5

speed limit) than White drivers in 65 mph zones. Although we included sex and age control variables in the analysis, it is important to remember that the *age* variable was poorly measured in this dataset. This limits its ability to serve as a covariate. Our Tollbooth Survey found that the average age of minority drivers was lower than for White drivers; thus, it is possible that age differences may account for observed differences between Black and White drivers. The interaction between sex and race was examined, but this did not prove to be statistically significant ($p = .35$).

Table 5 Logistic regression analysis predicting speed status from race/ethnicity, age, and sex

Turnpike speed limit	Variable	Odds ratio (without age and sex)	Odds ratio (with age and sex)
65 mph	*Race*		
	White	1.00	1.00
	Black	1.96**	1.64**
	Hispanic	1.27*	1.07
	Other	1.34**	1.18**
	Age		
	16–45	-	3.15**
	Over 45	-	1.00
	Sex		
	Male	-	1.20**
	Female	-	1.00
55 mph	*Race*		
	White	1.00	1.00
	Black	0.96	0.86
	Hispanic	0.79	0.69*
	Other	0.75*	0.66**
	Age		
	16–45	-	2.15**
	Over 45	-	1.00
	Sex		
	Male	-	0.99
	Female	-	1.00

*$p < .05$; **$p < .005$.

These findings were not replicated in the 55 mph speed limit zones. There, no statistically reliable difference was found between White and Black drivers. Hispanic and drivers classified as Other were, however, less likely to be speeders than White drivers. Again, young drivers were more likely to be speeders, but there was no difference between men and women.

One possible explanation for the lack of racial/ethnic differences found in the 55 mph areas is that, in those areas, a much larger proportion of the vehicles are traveling higher than the criterion set as the definition of speeder (13.0 percent vs. 1.7 percent in the 65 mph zone). It does not take a great deal of racial/ethnic differences in speeding rates to produce a dramatic overrepresentation in a small faction of the drivers, as is the case in the 65 mph zones. However, where a substantial number of speeders exist, very large disparities between the races would be required to produce similar overrepresentations.

Figure 1 shows the percentage of each racial/ethnic group at varying speeds in the 65 mph zone. The figure reveals that in 65 mph zones, Black and White drivers maintain a relatively consistent representation until approximately 77 mph. However, beyond this point the proportion of Black drivers increases with speed. In 55 mph zones, however, there was no upward trend in the proportion of Black motorists as speed increases. This implies that the lack of racial disparities in speeding rates observed in the 55 mph zone is not due to the speed criterion set, but instead a function of differing driving behaviors in that zone.

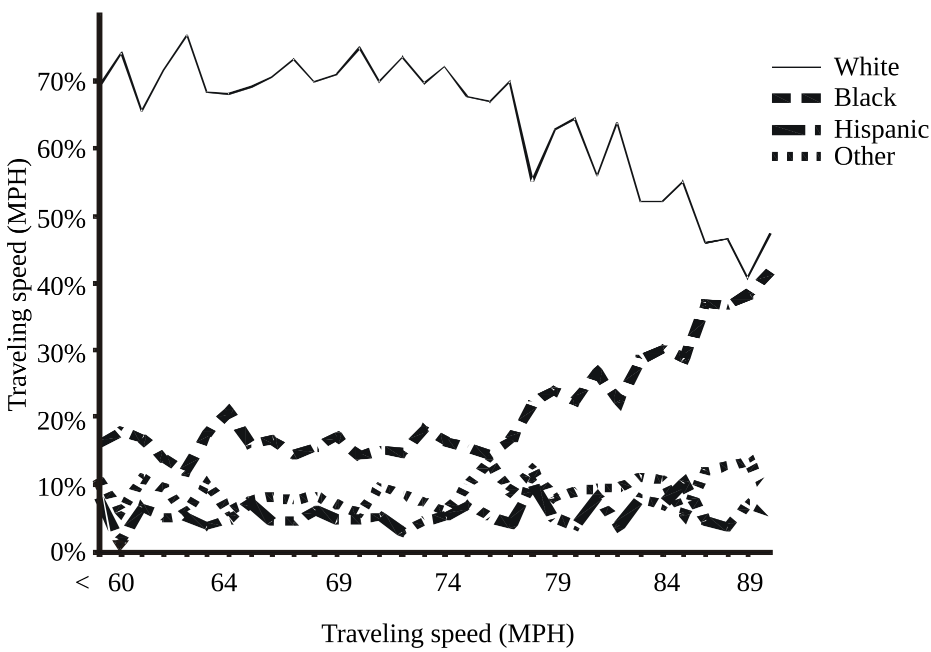

Figure 1 Race/ethnic distribution of vehicles traveling at various speeds in the 65 mph zone.

We have no data on hand to adequately test possible hypothesis that would explain the differing results between 55 mph and 65 mph speed-limit zones.

Race and driving distance

A plausible hypothesis for the racial disparity in speeding is that driving distances vary by race. If, due to varying geographic distributions of racial groups along the Turnpike, Black drivers were more likely to be driving long distances when traveling in the Southern or Central segments, then one might expect that speeding behavior may also vary. Two assumptions are necessary for this to be a logical explanation: (1) those traveling long distances will speed more than those using the Turnpike to travel only limited distances; and (2) there are travel-distance disparities among racial/ethnic groups of drivers. We have data from our Tollbooth Survey that can test the second assumption; unfortunately, we have no data to test the first assumption.

Participant drivers in the Tollbooth Survey were asked their Turnpike entry point. Using this information, distance traveled by racial/ethnicity group could be calculated. The results suggest that Black, Hispanic, and Other drivers may be more likely to be traveling long distances than White drivers in the Southern and Central segments. These data are presented in Table 6. Black drivers were more than twice as likely as White drivers to be traveling to or from another

Table 6 Turnpike distance traveled by driver race/ethnicity

Race	Average travel distance (miles)	Inter-segment travel relative probability
Southern segment travel distance analysis		
White	52.6 ± 7.7	1.00
Black	60.4 ± 9.7	1.19 *ns*
Hispanic	60.2 ± 13.9	1.45 *ns*
Other	55.4 ± 8.7	1.48 *ns*
Central segment travel distance analysis		
White	31.8 ± 3.3	1.00
Black	38.5 ± 7.8	2.29*
Hispanic	30.6 ± 7.5	1.94*
Other	30.2 ± 5.7	1.38 *ns*
Northern segment travel distance analysis		
White	21.6 ± 3.9	1.00
Black	25.6 ± 7.1	0.77 *ns*
Hispanic	21.6 ± 5.8	0.86 *ns*
Other	20.8 ± 4.7	1.04 *ns*

Note. *ns*: Difference from 1.00 is not statistically reliable.
*Difference from 1.00 is statistically reliable, $p < .05$.

segment when in the Central segment. In the Northern segment, Black and Hispanic drivers are actually less likely to be traveling to another segment.[5] The disparity seen in the Central segment and trends found in the other segments suggest that Black and Hispanic drivers may be more likely to speed in the Southern and Central segments because they are more likely to use that portion of the Turnpike for long-distance travel than White drivers. However, we have no data to test this hypothesis further as our Speed Survey did not gather information on traveling distance.

Discussion

Our results reveal that in southern and central New Jersey, Black drivers were overrepresented among speeders (relative to nonspeeders). More importantly, however, police stop rates matched very closely the rates at which drivers exceeded the speed limit by 15 mph. These results suggest that during the period of data collection, New Jersey State Troopers assigned to the Turnpike stopped Black drivers in approximate proportion to their representation among speeders. The results offer a plausible alternative explanation for the high police stop rates of Black drivers relative to regional census counts. What these studies *do not* show, nor were they intended to show, is that racial profiling is nonexistent. The key conclusion that we can draw from this research is that the typical method of assessing racial profiling on the precinct- or jurisdiction level is not adequate; the racial distribution in the population of driving nonviolators cannot be assumed to reflect the racial distribution in the population of driving violators, and it is from this latter population that police stops should be drawn.

Strengths and Weaknesses

The Speed Study was designed with the same general aims as Lamberth's "rolling" survey of speed violators. Our data-collection method, however, was the first of its kind for this purpose. There are several strengths to our methodological approach. First, although we dichotomized drivers in terms of speeder versus nonspeeder, we captured actual vehicle speed for each vehicle. This allowed us to examine results using different speed thresholds for the definition of speeder. Second, by photographing vehicles, we were able to expose the pictures to a panel of raters to ensure reliable race and ethnicity ratings. Third, we used sophisticated radar equipment to obtain population counts of speeders and nonspeeders. Fourth, we collected data from locations throughout the

5. Although these results point to a correspondence between driving distance and likelihood of speeding, not all of the results are statistically significant. The Tollbooth Survey was not originally intended to uncover segment-by-segment differences in driving distance behaviors. A larger sample may have produced similar results that would have been judged statistically reliable.

Turnpike, and through proper weighting and analysis, we were able to make population estimates of speeders versus nonspeeders.

Despite the strengths of this methodology, there are limitations worth noting, some within the scope of the specific methodology and others outside the scope. In the Tollbooth Study, participant race and ethnicity was determined by self-report, whereas the race of individuals stopped by police was based on police officers' perceptions. This may limit the comparability of the two distributions. The nature and degree of the bias between the perceived race and self-reported race are not clear. However, it is plausible that such bias would not only contribute to random error (and thus affect statistical power) but also affect the actual point estimates. Had our survey staff recorded their perceptions of drivers' race and ethnicity in addition to the drivers' own indication, we would be better able to quantify this bias.

It is important to consider, however, that the possibility of this bias underscores the inadequacy of traditional methods for assessing racial profiling. Studies that compare police stop data to regional demography also rely on different methods to measure race, as police stop data are based on observation and census data are based on self-report. Even if police stops reflected an unbiased sample of the population, we still might expect to find differences between the two sets of data given that different methods were used to measure race.

In the Speed Study, three trained coders examined each photograph to identify the race/ethnicity, sex, and age of the driver; a rating was considered reliable only when two of three judges agreed. Driver race in only a relatively small proportion (less than 9 percent total) of the cases was judged as unreliable, indicating good overall reliability.

Conversely, a moderate proportion of images (22.3 percent nonspeeders and 25.2 percent speeders) were judged to be unusable (i.e., at least two of three raters agreed that that they could not make a racial determination). Typically, cases were rated as unusable due to glare, tinted windshields, and equipment malfunction. The high proportion of unusable cases, however, does not necessarily threaten the interpretation of the results.

Although the racial and ethnic composition of these unusable cases is unknown, there is no clear reason to expect the racial distribution of the unusable cases to differ from that of the reliable cases. This is particularly true because our analysis revealed that the difference in usability between conditions (speeder vs. nonspeeders) could be accounted for by camera position. However, even if the racial composition of unusable cases did differ from reliable cases, there is no reason to believe that this bias would also differ across conditions. Because our conclusions were drawn from a comparison between speeders and nonspeeders, the results would only be mitigated if both (1) driver race/ethnicity was correlated with usability and (2) usability was correlated with speeder/nonspeeder status. The authors cannot generate a plausible explanation that accounts for such a relationship between picture usability, race/ethnicity, and speeder status. However, we cannot determine empirically whether such a relationship exists.

That the Speeder and Tollbooth Surveys demonstrated convergent validity with respect to overall estimates of Black and White drivers makes us more confident that a peculiar interaction such as that is unlikely. Given that there is no easily generated plausible explanation for any potential racial/ethnic differences in usability rate and that we can statistically explain the differences in usability rate between speeder and nonspeeders (through analysis of lane selection), the fact that a large proportion of total cases were rated as unusable should not threaten the validity of the findings.

This research also is limited in that it addresses only speeding, just one of many legitimate reasons for a traffic stop. We conjecture that speeding is by far the most common reason for a vehicle stop, but because the police stop data only indicated that the offense was a moving violation, without specifying the offense we cannot empirically examine this possibility. *State v. Kennedy* (1991) indicates that a comparison of the racial composition between speeding violations and citations for *that offense* is an appropriate benchmark for measuring racial profiling. Our study does not quite accommodate *State v. Kennedy* in that our police stop data contain stops for moving violations other than speeding.

More importantly, these data do not address police behavior other than traffic stops (e.g., vehicle consent searchers, citations, or arrests). Thus, even if police officers legitimately stopped a higher proportion of Black drivers than are represented in the population, our data are silent with respect to whether vehicles driven by Black drivers are disproportionately searched (for more details, see Harris, 2002).

Racial Differences and Similarities

Our research demonstrated differences in the proportion of speeders (relative to nonspeeders) between Black and White drivers. The magnitude of these differences, however, can be misleading in terms of the raw numbers of speeding violators. Among all ethnic and racial categories, the majority of drivers were nonspeeders: only 1.4 percent of White drivers and 2.7 percent of Black drivers drove at or above 80 mph. The average speed for Black and White was very similar as well: 66.3 and 66.8, respectively, in a 65 mph zone. This suggests that Black and White drivers, on average, were quite similar, and only at the upper extreme end of the tail[6] did the racial populations differ.

Our decision to define speeders using the 80 mph cutoff was made a priori and was indicated in our initial research proposal to the New Jersey Office of the Attorney General. The results of this research may have produced very different conclusions had the speeding-limit threshold been established at a lower value, such as 75 mph or 70 mph. The discrepant representation between Black and

6. The lower extreme end of the speed distribution was artificially constrained by congestion, etc., so it is not easily interpretable.

White drivers among speeders vanishes as the speeding-limit criterion approaches the legal limit but becomes exaggerated as the limit increases to above 80 mph. In this respect, our results do not differ greatly from Lamberth's study; Lamberth used a value of 5 mph above the speed limit to define speeding. Using 70 mph as a cutoff, our data reveal roughly equal proportional racial/ethnic representation between speeders and nonspeeders. In fact, it is not until the threshold is slightly below 80 mph that Black drivers became overrepresented.

It is noteworthy that our Tollbooth Survey indicated that Black drivers were younger, on average, than White drivers. Given evidence that younger drivers are statistically at higher risk for traffic violations, it is plausible that age may account for the overrepresentation of Black drivers among speeders. In our logistic regression, age was measured poorly; with a better measure, the differences between Black and White drivers may have been attenuated.

Although we conducted a logistic regression to identify demographic characteristics that uniquely predict speeder status, results indicating that race does (or does not) *uniquely* predict speeding has little bearing on our basic conclusion. Even if the disproportionate representation of Black drivers among speeders was explained entirely by age (or some other variable), it would not attenuate the finding that the racial distribution of police stops approximated the racial distributions of speeders. It is quite possible, even likely, that other variables can account for the differences between Black and White drivers documented in this research. However, regardless of the *cause* of this racial discrepancy, the main conclusion of this research—that it is plausible for Black drivers to be overrepresented in police stops even if police behavior is unbiased—would nevertheless remain tenable.

Conclusion

The results of our research offer a plausible explanation for the findings that Black drivers are represented among traffic stops at a higher rate than they are represented in the population. Although this does not rule out the plausibility of racial profiling or of discriminatory behavior on the part of the police, the results underscore weaknesses in the traditional method of assessing the extent of racial profiling by police precinct or within jurisdiction. One cannot assume that the racial/ethnic distribution of residents will match the racial/ethnic distribution of traffic violators, particularly at extreme ends of the distribution, where it is easier for disproportionate differences to occur. Benchmarks for measuring racial profiling (comparing with police stop rates) should be based on estimates of traffic violations rather than on geographic census counts.

Acknowledgments

The New Jersey Office of the Attorney General provided funds for this research. We also would like to thank David A. Harris for his review of earlier *drafts of the paper and for his insight into the empirical and legal history of the phenomenon. This report reflects the views of the authors and not necessarily those of the New Jersey Attorney General's Office.*

References

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, Intention, Efficiency, and Control in Social Cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition*. Hillsdale, NJ: Erlbaum.

Bargh, J. A., Lombardi, W. J., & Higgins, E. T. (1988). Automaticity of chronically accessible constructs in person X situation effects on person perception: It's just a matter of time. *Journal of Personality and Social Psychology,* 55, 599–605.

Bargh, J. A., & Pratto, F. (1986). Individual construct accessibility and perceptual selection. *Journal of Experimental Social Psychology,* 293–311.

Cordner, G., Williams, B., & Velasco, A. (2002 November). *Vehicle stops in San Diego: 2001* (Report). San Diego, CA: Eastern Kentucky University; Vanderbilt University; San Diego State University.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology,* 56, 5–18.

Eck, J. E., Liu, L., & Bostaph, L. G. (2003 October 31). *Police vehicle stops in Cincinnati: July 1–December 31, 2001* (Report). Cincinnati, OH: University of Cincinnati, the Cincinnati Police Department, and the City of Cincinnati.

Harris, D. A. (2002). *Profiles in injustice: Why racial profiling cannot work*. New York: The New Press.

Lundman, R. J., & Kaufman, R. L. (2003). Driving while black: Effects of race, ethnicity, and gender on citizen self-reports of traffic stops and police actions. *Criminology,* 41 (1), 195–220.

Meehan, A. J. (1998). The impact of mobile data terminal (MDT) information technology on communication and recordkeeping in patrol work. *Qualitative Sociology,* 21, 225–254.

Rodriguez v. California Highway Patrol, 891131 (N.D. Cal 2000).

State v. Kennedy, 588834, 840 (N.J. Super. Ct. App. Div. 1991).

State v. Pedro Soto, 734350 (N.J. Super Ct. Law Div. 1996).

Verniero, P., & Zoubek, P. H. (1999 04/20). *Interim report of the State Police Review Team regarding allegations of racial profiling*. Newark: State of New Jersey.

Wilkins v. Maryland State Police (D. Md 1993).

Wright, C. P., Tomaskovic-Devey, D., & Zingraff, M. T. (2000). *Self-reported police speeding stops: Results from a North Carolina record check survey*. Paper presented at the American Society of Criminology Meeting San Francisco, CA.

Appendix A. Race and Ethnicity of Participants and Refusals in the Tollbooth Survey

Table A1 reveals the racial composition of drivers in Tollbooth survey. Participants indicated their own racial or ethnic identification, whereas the race or ethnicity of refusals was estimated by the interviewer. Refusals were less likely to be categorized as Black, and more likely to be categorized as Other. However, the racial composition of the total dataset is comparable to that of the participants.

The age of refusals was also estimated by the tollbooth interview. However, age did not differ significantly between participants (M = 40.3) and refusals (M = 40.4), t (4491) = –.33, p = .75.

Table A1 Racial composition of participants and refusals from the tollbooth survey

	Total	Participants (self-reported)	Refusals (estimated)
White	2,759 (59.3%)	2,403 (59.6%)	356 (57.3%)
Black	711 (15.3%)	652 (16.2%)	59 (09.5%)
Hispanic	709 (15.2%)	605 (15.0%)	104 (16.7%)
Other	477 (10.2%)	375 (09.3%)	102 (16.4%)
Total	4,656	4,035	621

Appendix B. Sample Characteristics of the Speed Survey

Approximately 48 hours of data images were collected at each of the 14 locations along the Turnpike between March 31 and June 30, 2001. Images were taken at each location on a weekend and weekday. No sampling occurred on holidays or Mother's Day. There were 38,747 images conforming to the sampling structure defined here transmitted to the PIRE offices. Before raters determined the usability of cases, 600 cases were removed due to technical problems resulting in missing values, and 41 cases were removed because they were the sole representatives of 8 strata causing their weights to be excessively high, which resulted in unstable results. Thus, removing those 41 cases resulted in the elimination of 8 of the 331 strata combinations. In addition, two of the three raters judged 8,506 cases to be unusable and 3,266 cases to be unreliable (that is, at least two raters did not agreed upon race or sex of the driver). The number of cases from each sampling location by usability is presented in Table B1. The table displays the sample characteristics with nonweighted values in the cells, offered only to demonstrate characteristics of the data on hand.

Table B1 Total counts of available cases at sampling locations by usability

	Weekend			Weekday			
Direction and mile post	Total	Reliable	Unreliable /unusable	Total	Reliable	Unreliable /unusable	Grand total
N-MP 13.8	1,749	1,194	555	1,592	958	634	3,341
S-MP 23.3	1,061	851	210	1,299	1,028	271	2,360
S-MP 38.0	757	466	291	1,060	542	518	1,817
N-MP 40.8	1,206	532	674	1,572	1,029	543	2,778
S-MP 56.8	1,076	626	450	1,141	588	553	2,217
N-MP57.1	877	588	289	1,248	614	634	2,125
S-MP 73.9	1,232	998	234	1,294	1,043	251	2,526
N-MP 78.4	1,517	1,110	407	1,396	826	570	2,913
N-MP 96.2	1,044	640	404	1,342	928	414	2,386
S-MP 96.2	1,386	965	421	945	709	236	2,331
S-MP 104.0	1,495	1,242	253	2,376	1,298	1,078	3,871
N-MP 104.9	1,315	843	472	1,358	916	442	2,673
S-MP 115.6	1,652	1,272	380	1,912	1,531	381	3,564
N-MP 115.7	1,923	1,442	481	1,922	1,555	367	3,845
Total	18,290	12,769	5,521	20,457	13,565	6,892	38,747

Table B2 describes sample counts of participants by race/ethnicity. The data are presented in two ways. Along the rows are racial and ethnic categories (along with categories for unreliable and unusable) defined by agreement from at least two out of three raters; the columns contain racial and ethnic coding categories based on a more stringent criterion of three out of three raters agreeing on the category. For the rows, the unreliable category represents counts of cases where two out of three coders failed to agree on the racial or ethnic category. For the columns, the unreliable category represents counts of cases where coders failed to agree unanimously on the racial or ethnic category. In approximately 68 percent of cases two out of three coders agreed on the rating category. However, in only 40 percent of cases did all three coders agree on the same category.

Table B2 Sample counts by race/ethnicity and rater coding criterion

		Race/ethnicity: Three out of three coders agree						
		White	Black	Hispanic	Other	Unreliable	Unusable	Total
Race/ethnicity: At least two out of three coders agree	White	10,373	0	0	0	6,809	0	17,182
	Black	0	3,265	0	0	2,049	0	5,314
	Hispanic	327	0	287	1	904	0	1,519
	Other	0	0	0	998	1,323	0	2,321
	Unreliable	19	16	2	0	3,229	0	3,266
	Unusable	150	30	7	15	3,618	5,325	9,145
	Total	10,869	3,311	296	1,014	17,932	5,325	38,747

A review of a draft of this document produced a query about the use of a "majority" decision rule as opposed to a "unanimous" decision rule. Based upon this query, to ensure that our results were not an artifact of our predetermined decision rule, an additional set of analyses was run under the "unanimous" decision rule. These *post hoc* results are not presented in detail here. However, based upon the unanimous decision rule, additional cases removed from the usable data set due to lack of unanimity differed only fractionally for Black and White nonspeeders and Black and White speeders. Consistent with this relatively proportional distribution of non-unanimous cases, all of the relevant major findings presented in the report were replicated within the originally reported margins of error when using the "unanimous" decision rule for race.