By the Numbers:
A Guide for Analyzing Race Data
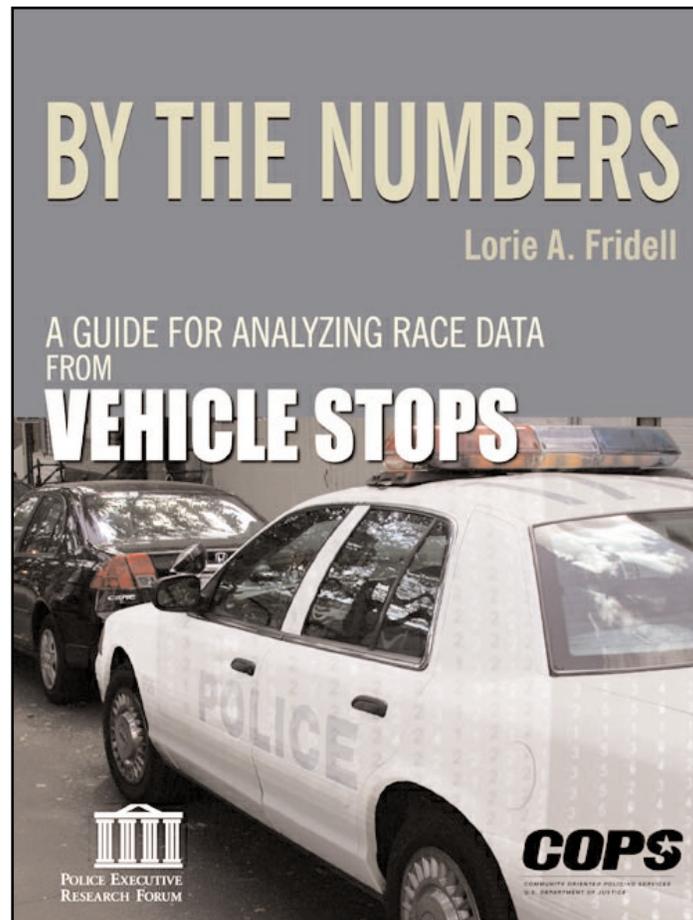# From Vehicle Stops

Executive Summary

By the Numbers:

A Guide for Analyzing Race Data from Vehicle Stops

Executive Summary

With funding from the U.S. Department of Justice's Office of Community Oriented Policing Services (COPS), the Police Executive Research Forum (PERF) has completed this guide for law enforcement agencies and other stakeholders on how to analyze, interpret, and understand vehicle stop data being collected on drivers' race. The first volume (Chapters 1-10) is available on the PERF Web site at www.policeforum.org, and the second volume (Chapters 11-13) will be available later in 2004. A softbound version of *By the Numbers* containing all thirteen chapters will be published by PERF in 2004.

This report will be of greatest value to the people who are analyzing the data—namely, research staff for law enforcement agencies, social scientists, and researchers associated with interest groups. A forthcoming companion document, *Understanding Race Data from Vehicle Stops: A Stakeholder's Guide,* will be less technical and directed to a broader audience of police professionals, concerned residents, advocacy groups, the media, and local, state, and federal policy makers. It will discuss the challenge of benchmarking, how to assess the quality of benchmarks, how to interpret results responsibly, and how to use the data for constructive dialogue and reform.

Background

While the vast majority of police officers in this country are principled men and women who are committed to serving all citizens respectfully and fairly, incidents and perceptions of racially biased policing, such as those reported in the national media, cannot be ignored. The issues involved in "racial profiling" and racially biased policing are not new. They are the latest manifestation of a long history of sometimes tense and even volatile police-minority relations. Yet police are more capable than ever before of addressing these longstanding issues in an effective way. In the past few decades, there has been a revolution in the quality and quantity of police training, in the standards for recruiting and hiring officers, and in policing procedures and accountability measures. An important part of this revolution has been the widespread adoption of community policing.

Indeed, agencies throughout the United States have implemented reforms to respond to the issues related to racially biased policing and the perceptions that it is practiced.   These reforms include adopting policies, implementing training, reaching out to minority communities, changing recruitment and hiring procedures, and improving supervision and accountability measures. Many agencies also are collecting information on stops made by police to assess whether police are inappropriately using race as a factor in their decision making. Some are collecting the data voluntarily; others are required by local mandate or state legislation to do so.   By early 2003, approximately half of the states had adopted legislation related to racial profiling; most of these laws include data collection requirements. Similar legislation is pending in other states.

The agencies collecting data require officers to report information on all traffic-related stops or on all vehicle stops (that is, traffic-related stops and stops to investigate a possible crime).[1] The information collected by officers includes the race/ethnicity of the driver and other information about the stop, such as the reasons for the stop, the disposition of the stop (a citation or warning, for example), whether a search was conducted, and the outcome of the search. Data collection is meant to help administrators determine whether police decisions to stop drivers are influenced by racial bias.

Although jurisdictions nationwide have invested considerable resources to collect race data from vehicle stops, most jurisdictions do not know how to analyze the collected data properly. They are either ill-equipped to do the analysis, or they are misinformed about what should be done. An overwhelming majority of the data analyses reviewed by PERF staff for this project were based on substandard methods. Most agencies are using models for their analyses that fall far short of minimal social science standards. In jurisdictions across the country, reports prepared by agencies or external groups (for example, some civil rights groups) draw conclusions wholly unsupported by the data. Other reports indicate that despite all the efforts and resources that were dedicated to the data collection, no conclusions can be drawn. These failures can largely be explained by the complexity of the task of measuring whether policing in a jurisdiction is racially biased. A tremendous number of factors other than bias can legitimately influence police decisions to stop drivers, and these "alternative hypotheses" must be ruled out before the "bias hypothesis" can be tested. A lack of understanding about which benchmarking methods will yield the most valid interpretations of the data is hindering agencies' efforts to reach valid, responsible conclusions.

---

[1] The term "vehicle stop" is used to denote any stop made by police of a person in a vehicle. The term "traffic stop" denotes a vehicle stop for the purpose of responding to a violation of traffic laws (including codes related to quality/maintenance of vehicles). A minority of agencies are also collecting data on pedestrian stops.

Many agencies that have already initiated data collection will continue to do so for at least several years to come; and, through choice or mandate, many more agencies will begin collecting race data. It is important that these agencies understand how to analyze and interpret their data in a manner that reflects accepted social science standards. *By the Numbers* will help law enforcement agencies and other stakeholders avoid past mistakes in data collection and implement the best practices that experts have developed. Following is an explanation of the specific purposes and contents of this "how to" guide.

The "How To" Guide for Agencies

The purposes of this COPS-supported PERF document are

(1)   to describe the social science challenges associated with data collection initiatives so that agencies and other stakeholders can be made fully aware of both the potential and limitations of police-citizen contact data collection; and

(2)   to provide clear guidelines for analyzing and interpreting the data so that the jurisdictions collecting them can conduct the most valid and responsible analyses possible with the resources they have.

Chapter 1 of *By the Numbers* offers a general introduction to the collection of race data for the purpose of measuring whether policing in a jurisdiction is racially biased. Chapter 2 describes the specific social science challenges associated with analyzing and interpreting the police-citizen

contact data. It also presents a scheme for evaluating the strength of various benchmarking

methods.[2]

As Chapter 2 explains, a key aspect of analyzing vehicle stop data is to determine whether the

driver's race/ethnicity has an impact on police stopping decisions. In order to assess whether

there is an impact, however, we must exclude or "control for" factors other than race/ethnicity

that might legitimately explain police stopping decisions. For example, the reports of most

jurisdictions regarding their police-citizen contact data state that men are stopped by police more

than women. Assume a jurisdiction finds that 65 percent of its vehicle stops by police are of male

drivers and 35 percent are of female drivers. Does this indicate gender bias by police? It is

unclear from these data, but most of us are disinclined to jump to that conclusion because we can

think of factors other than police bias that could account for the disproportionate stopping of male

drivers.   That is, alternative hypotheses for the results exist. One possibility is that men drive

more than women (the quantity factor). Another possibility is that men violate traffic laws more

often than women do (the quality factor). A third possibility is that more males than females

drive in the areas where police stopping activity tends to occur (the location factor). We do not

know if these possibilities are true, but we must consider these alternative explanations in our

research design because it is logical to assume that

- people who drive more should be more at risk of being stopped by police,

- people who drive poorly should be more at risk of being stopped by police, and

- people who drive in locations where stopping activity by police is high should be more at
  risk of being stopped by police.

---

[2] As described further below, benchmarking methods help researchers compare the racial/ethnic
composition of drivers stopped by police to the racial/ethnic population of drivers at risk of being stopped
by police if police bias is not a factor.

In developing "benchmarks," the researcher is attempting to construct a comparison group that represents the drivers at risk of being stopped by police—absent bias. This group is compared to the group of drivers *actually* stopped to help determine whether racial bias may have been a factor in police officers' decision-making process. The variation in quality across benchmarks is directly related to how closely each benchmark represents the group of people who should be at risk of being stopped by police if no bias exists. The strongest benchmarks take into consideration variations in driving quality, driving quantity, and driving location.

It is not difficult to measure whether there is *disparity* between racial/ethnic groups in stops made by police; the difficulty comes in identifying the *causes* for any disparity. For instance, a jurisdiction might compare the demographic profile of people stopped by police to the demographic profile of residents as measured by the census. The results might show "disparity"; that is, the results might show that some groups are stopped disproportionate to their representation in the residential population. The jurisdiction, cannot, however, identify the causes of that disparity using this measure. Only after controlling for driving quantity, driving quality, and driving location, can a researcher who finds that minorities are disproportionately represented among drivers stopped by police conclude with reasonable confidence that the disparity reflects police bias in decision making. Similarly, if no disparity was found, the researcher can fairly confidently conclude that bias was not a part of police decision making. If, on the other hand, the researcher finds disparity in the results after controlling for only driving quantity and driving location, the legitimate conclusions that can be drawn are limited: the researcher can conclude only that disparity exists and that the disparity could be the result of police bias or of differential driving quality. The researcher cannot pinpoint a single cause and must report that these two possible explanations for the disparity remain.

Chapter 3, "Getting Started," discusses important decisions agencies must make when they begin collecting and analyzing police-citizen contact data, including what stop information to collect, whether and how to involve residents and police personnel in the planning process, and what benchmark(s) to select. We emphasize that an agency should, if feasible, select a plan for analyzing the data at the same time that the decision makers decide what stops to target and what information to collect on stops.[3] We recommend that decision makers select all traffic stops or all vehicle stops, and not a subset of these categories as defined by their outcomes (for example, citations, arrests). Some jurisdictions (indeed, some entire states) are collecting data only on subsets of stops, such as traffic stops that result in a citation. In Chapter 3 we explain why this practice produces substandard data for analysis.

In Chapter 3 we also encourage agencies to involve residents and agency personnel from all levels in planning data collection and analysis. Police personnel—particularly line personnel—can bring valuable information and an important perspective to the table. These agency representatives have a critical stake in ensuring a high-quality initiative, and they should have the opportunity to raise any concerns they may have about the integrity and fairness of the data collection and analysis system. Employees' involvement can also facilitate "buy in" by the line officers upon whom the agency will rely to collect the data.

The involvement of residents (particularly minority residents) in data collection planning can improve police-citizen relations, enhance the credibility of the research efforts, and increase the likelihood that the community will view the findings as legitimate. Involving jurisdiction residents in discussions regarding data analysis/interpretation has an additional advantage: a core

---

[3] For information on what stops to target for data collection and what information to obtain for each stop, see PERF's first report on racial profiling entitled *Racially Biased Policing: A Principled Response* (Fridell et al. 2001, Chap. 8).

group of residents becomes knowledgeable about the complexities and constraints of the data collection process. Later on, when the results are released to the public, these residents can affirm the integrity of the analysis and the responsible interpretation of the results.

Before conducting the analysis, a law enforcement agency must decide whether to partner with an external social scientist. We recommend, if resources allow, that an agency obtain assistance for analyzing its police-citizen contact data. There are two major reasons for partnering with social scientists:

- Partnering with an individual or a team external to the agency can add credibility to the process and results.
- The skills of trained social scientists can supplement the internal resources available for research.

A key decision departments must make is which benchmark or benchmarks to select for analyses. In Chapter 3, we set forth the factors that an agency should consider in selecting a benchmark: (1) level of measurement precision desired, (2) agency resources, (3) data elements collected by the officers for each contact, and (4) availability of the information required for the various benchmarks.

Law enforcement agencies, regardless of the benchmarking method they choose for evaluating whether policing in their jurisdiction is racially biased, should follow certain guidelines on the analysis of police-citizen contact data. Chapter 4 presents these guidelines. The issues addressed are relevant to all analysis efforts, regardless of their particular focus or the benchmarking method selected. Topics include reviewing data quality, selecting reference periods (that is, selecting the length of time to collect data before analyzing it), and analyzing subsets of data.

We start by explaining how the data that have been collected from officers can be checked for quality, an important first step in any type of social science research and not unique to the analysis of police-citizen contact data. Although there is no cost-effective way to ensure that the data are 100 percent accurate, the methods described in the chapter can help the researcher check for and enhance the quality of their data. A range of methods can be used to ascertain whether officers are submitting forms to the agency for each and every stop targeted for data collection. Additionally, there are methods for assessing the level and source of missing data, errors, and intentional misstatements of facts. When selecting reference periods we recommend that, if economically and politically feasible, agencies collect one year of data before analyzing it. Agencies are advised to delay the start of the reference period for several months after data collection begins. In the first few months officers can become accustomed to the data collection process, and their data should be reviewed to identify particular problems (such as large amounts of missing data on certain variables or missing forms). Once the problems appear to be resolved, the reference period should begin.

For many reasons, it is appropriate for agencies to analyze subsets of their police-citizen contact data. In Chapter 4 we describe why a researcher might choose not to analyze all of the data submitted during the reference period but only a portion, and how and why a researcher might conduct separate, multiple analyses using subsets of the data. For example, the researcher might choose to analyze for his or her report only proactive stops (stops in which police have discretion regarding whom to stop); then the researcher might choose to conduct separate analyses of these data within geographic subareas of the jurisdiction. We discuss subsets based on (1) whether stops are proactive or reactive, (2) whether the officer could discern the driver's race/ethnicity, (3) whether the driver appears in the database once or multiple times, (4) geographic locations of

stops (to allow for analyses within subareas of the jurisdiction), and (5) whether the stops are for traffic violations or for the purpose of investigating crime.

The final section of Chapter 4 explains the need for comparability of the stop data and benchmarking data or what we call "matching the numerator and the denominator." The "numerator" refers to the data collected on stops by the police, and the "denominator" refers to the data collected to produce the comparison group, or benchmark. To "match the numerator and the denominator" the researcher adjusts the stop data to correspond to any limiting parameters of the benchmark or vice versa. For instance, in the observation benchmarking method, researchers collect data from the field regarding the race/ethnicity of drivers. Placed at various locations, the observers count the drivers in different race/ethnicity categories. This process produces a racial/ethnic profile of drivers observed at these locations that can be compared to the people who are stopped by police. Since the "denominator" (observation data) pertains only to certain areas, the relevant analysis will only include in the "numerator" the police stops in that area. Using this method, the researcher will compare the demographics of the people who are observed driving through Intersection A, for example, to the demographics of the people stopped by police in and around Intersection A. (This type of analysis will be conducted separately for each intersection.)

The numerator and denominator must be matched with regard to other parameters as well. For example, if observation data were collected from January through May 2002, the analysis should involve only police stops that occurred during roughly that same time period. If the researchers collected observation data only during daylight hours because of visibility issues, then the analysis should include in the numerator only those stops that occurred during daylight hours.

Chapters 5 through 10 target some of the mistakes often made when comparing stop data to commonly used benchmarks. For example, many law enforcement agencies and outside analysts will compare the percentage of stops that involve African Americans or other minorities to the racial make-up of the residents of a particular area as measured by census data. More often than not, the mass media, civic groups, and citizens draw conclusions from this comparison regarding the existence or lack of racially biased policing in the jurisdiction; *these conclusions are wholly unsupportable using this method of analysis.* Frequently, no mention is made of non-race-related explanations for the disparity between the census population and the population of stopped drivers, explanations that relate to driving quantity, driving quality, and driving location. These are all factors that legitimately affect stopping behavior by police.

These last six chapters of Volume I cover the following topics:

- θ Chapter 5: Benchmarking with Adjusted Census Data
- θ Chapter 6: Benchmarking with DMV Data
- θ Chapter 7: Benchmarking with Data from "Blind" Enforcement Mechanisms
- θ Chapter 8: Benchmarking with Data for Matched Officers or Matched Groups of Officers
- θ Chapter 9: Observation Benchmarking
- θ Chapter 10: Other Benchmarking Methods and Analysis Tools.

Readers are given clear and specific information regarding how to implement each benchmarking method. Equally important, they learn what conclusions regarding the existence or absence of racially biased policing can and cannot be drawn from each method. This information is particularly valuable because it will enable law enforcement agencies to report legitimate findings rather than misinterpretations of police-citizen contact data.

Chapter 5 warns against the most commonly used benchmarking method, unadjusted census benchmarking, and provides detailed guidance on how law enforcement agencies can modify or "adjust" census data to reflect factors that can legitimately influence police decisions to stop drivers. In traditional census benchmarking, law enforcement agencies compare the demographic profile of drivers stopped by police to the U.S. Census Bureau demographic profile of jurisdiction residents. A straight comparison between the demographics of these two groups is called "unadjusted" census benchmarking—a method that we do not recommend. Chapter 5 highlights valuable adjustments that can be made. For example, researchers may adjust the census data on the demographics of residents to take into consideration who, among those residents, owns a vehicle. This adjustment reflects the fact that not every resident owns a vehicle, and people without vehicles are clearly at less risk of being stopped in vehicles by police. Census benchmarking with this adjustment is a stronger method than unadjusted census benchmarking for assessing the nature and extent of racially biased policing. Innovative researchers have also incorporated information regarding the influx of drivers from neighboring jurisdictions. Chapter 5 provides information on how to measure race/ethnicity using census data and how and why to conduct analyses within subareas of the jurisdiction as well.

Despite the weaknesses of using census data as a diagnostic tool, some jurisdictions (limited by resources or time) may have no option other than to use this method. This will be particularly true of researchers charged with analyzing data for an entire state. The obligation of the researcher in this position is to ensure that the results are conveyed in a responsible fashion. In fact, this obligation falls to all stakeholders, including concerned citizens, civil rights groups, and the media. *No one interpreting results based on census benchmarking—even adjusted census benchmarking—can claim they have **proved** the existence or lack of racially biased policing.*

This caveat is not unique to adjusted census benchmarking, and the inability to identify a causal connection between driver race/ethnicity and police decisions does not mean that data collection is without value. Even if the results from data collection do not provide definitive conclusions, they can serve as a basis for constructive discussions between police and citizens regarding ways to reduce racial bias and/or perceptions of racial bias.

Chapter 6 describes how some researchers have compared the racial/ethnic profile of licensed drivers who reside in a jurisdiction (using DMV data) to the profile of the drivers stopped by police. Like adjusting census data for vehicle ownership, this method produces an indirect measure of driving quantity. It accounts, in part, for the possibility that racial/ethnic groups are not equally represented as drivers on jurisdiction roads. This method is preferable to adjusting census data for vehicle ownership, if the necessary information is available to the jurisdiction. (To implement this method, drivers' license data in the state must be linked to racial/ethnic information.)

A North Carolina team of researchers (Zingraff et al. 2000; Smith et al. 2003) faced the challenge of attempting to measure racial profiling in fifty state trooper districts in the state. Using citations issued by police in the target jurisdiction, they estimated the representation on target jurisdiction roads of both residents and nonresidents. From demographic data for people with a driver's license who lived in the various districts included in their analyses, they developed a racial/ethnic profile of the population of "drivers driving" on jurisdiction roads.

Benchmarking with DMV data, like benchmarking with adjusted census data that takes into account vehicle ownership, imperfectly assesses who is driving on jurisdiction roads. The caveats associated with this method reflect three truths: not everyone with a driver's license drives, some people drive even though they do not have a driver's license, and some jurisdiction

residents (particularly students and military personnel) have a driver's license from another state. Most importantly, having a driver's license is a very crude measure of driving quantity. Residents of various racial/ethnic groups who have a driver's license may drive in different amounts. Agencies that have implemented benchmarking with DMV data cannot draw conclusions regarding the existence or lack of racially biased policing in their jurisdiction. Nonetheless, the results can be valuable as the basis for discussions between police and citizens about racially biased policing and the perceptions of its practice.

Chapter 7 describes how law enforcement agencies can use "blind" enforcement mechanisms (red light cameras, radar, air patrols) to produce a benchmark against which they can compare their data on stops by patrol officers. With this method the racial/ethnic profile of technology-selected drivers is compared to the racial/ethnic profile of human-selected drivers (that is, traffic law-violating drivers stopped by police). Some agencies compare stops in which officers exercise a high degree of discretion to low-discretion stops. This benchmarking method also is explained in Chapter 7.

Enforcement using red light cameras is "blind" because traffic law violators are detected and "ticketed" in a manner that does not allow for the intrusion of bias. The analyst compares the racial/ethnic profile of the drivers ticketed by the camera technology to the racial/ethnic profile of the drivers stopped by police. If officers are as "blind" to race/ethnicity as are the cameras, the demographic profile of the people stopped for red light violations by the officers should match the demographic profile of the people ticketed by the cameras in the same area. If, however, officers are targeting minorities for stops, minorities may compose a larger percentage of stops by the officers than by the technology.

Researchers implementing this benchmarking method, like others, must match the numerator and denominator. For example, the location of the red light cameras and the location of stops by police should be matched.

Radar enforcement is "blind" to the racial/ethnic characteristics of traffic law-violating drivers only if it is used in certain ways. The radar must be directed at all cars in a particular area, or the officer with the discretion to direct the radar at some cars and not at others must not be able to identify (because of light or distance) the racial/ethnic characteristics of the drivers. Air patrols are another "blind" enforcement mechanism. Air patrol officers identify speeders and direct patrol officers on the ground to stop the violators. The instructions to researchers regarding the use of radar and air patrol data resemble the instructions provided on the use of red-light-camera data. When implemented in accordance with our recommendations, benchmarking with "blind" enforcement mechanisms enables a jurisdiction to conduct a strong assessment of biased policing. The results, however, are strong only for specific locations and for particular types of stops. In other words, the rigor of the methodology comes at the cost of scope. A law enforcement agency that has chosen this benchmarking method must include an essential caveat in its report of results: *Jurisdictionwide* conclusions about the presence or absence of racially biased policing cannot be drawn.

Chapter 8 describes benchmarking with data for matched officers or matched groups of officers. Specifically, law enforcement agencies can compare stops by individual officers to stops by other officers, or they can compare stops by a group of officers to stops by other groups of officers. These comparisons must be made across "matched" sets of officers or groups of officers to control for legitimate factors (driving quantity, quality, and location) that increase the likelihood that a driver will be stopped. For instance, an agency might compare the racial/ethnic profile of people stopped by individual patrol officers who work the same shift in the same precinct. If a

particular officer stops proportionately more minority citizens than does his or her matched peers, further exploration of this officer's policing activities and decisions could be warranted. This method has also been referred to as "internal benchmarking."

Most of the recommendations in this chapter are geared toward ensuring that the researcher is comparing "similarly situated" officers or groups of officers. The goal is to compare officers (or units of officers) similar to one another in terms of the people at risk of being stopped by them. It is important to note that *the finding of disparate results does not prove the officer is acting in a racially biased manner.* The degree of confidence analysts can have that policing by the identified officer is racially biased is entirely dependent upon the strength of the match. Perfect matches would fully account for the legitimate factors that increase the risk of a driver being stopped (driving quantity, quality, and location); *but no match is perfect.* For instance, in a large geographic area within which officers are being compared, the racial/ethnic profile of drivers to which particular officers are exposed may differ. Even officers in the same area with the same general assignment of "patrol" may be directed toward different activities in the course of their work. Therefore, they would not be exposed to identical populations at risk of being stopped.

A subsequent review of officers (or of units of officers) who stop proportionately more minorities than their matched counterparts would explore whether the identified disparity is the result of bias or alternative, legitimate reasons. Supervisors should meet with the officer to discuss possible reasons for the disparity and review other sources of data before drawing conclusions regarding the cause of the disparate results.

There is a major caveat associated with internal benchmarking—one that must be highlighted in a law enforcement agency's report of its findings to the public. This method uses information on stopping behavior by police as both the numerator and denominator. In an officer-level match,

the numerator is one officer's stop data, and the denominator is the same type of data from other similarly situated officers in the same department. Although this method of analysis can identify "outliers," it cannot determine whether or not all units used in the comparison (all officers in an officer-level analysis or all groups in a group-level analysis) are practicing biased policing because, in this method, the department is compared to itself. Using internal benchmarking in conjunction with other methods allows the researcher to address this weakness while taking advantage of this method's strength.

In the observation method, the subject of Chapter 9, researchers compare the racial/ethnic profile of drivers observed at selected sites to the racial/ethnic profile of drivers stopped by police in the same vicinity. The observation data (the denominator) is used as a benchmark for the stop data (the numerator). Agencies usually hire one or several researchers to help them with this assessment. Observations are conducted by individuals trained by the researchers.

The observation benchmarking method, if implemented in accordance with solid methodological standards, can be effective in controlling for the legitimate factors that affect stopping decisions by police (driving quality, quantity, and location). Answers to the following questions are provided in Chapter 9:

- How should the observations be conducted?
- What should be observed?
- What locations should be selected for observation?
- When should the observations be conducted?

The chapter also explains how social scientists have addressed these questions in the context of their research. A key point of controversy is whether to use as a benchmark all drivers at the

selected site or only traffic law-violating drivers at the site. We recommend that the observation benchmark be based on law-violating drivers, not all drivers, because this model encompasses the fact that drivers who drive poorly are at greater risk of being stopped by police. (We present the alternative viewpoint in an appendix.)

The numerator and denominator data, as the chapter explains, should be matched with regard to violations observed, geographic location, time of day, and reference period.   As in other benchmarking methods, matching reduces the scope of the analysis, but it increases the researcher's ability to draw conclusions regarding racially biased policing.

The observation method, when conducted in accordance with standard principles of social science, can provide meaningful information for a jurisdiction regarding the existence of racially biased policing.   Researchers using this method, however, are only able to conduct "spot checks" of racially biased policing. That is, they will have a strong assessment of racially biased policing but only in the geographic areas, during the time periods, and for the violations under study.

Chapter 10, the final chapter in the first volume of *By the Numbers,* examines

- Crime data benchmarking,

- Crash (auto accident) data benchmarking,

- Transportation data benchmarking,

- Survey data benchmarking,

- Geographic Information Systems (GIS) resources, and

- Other analytic tools.

Researchers can benchmark police stop data against crime data, but only certain stops by police can be used in this analysis. Specifically, we recommend that benchmarks based on crime data be

used only to evaluate *investigative* vehicle stops by police. Using crime data to benchmark *traffic* stops would require the researcher to make a tenuous assumption—namely, that the same people who commit traffic violations are the ones who commit crimes and vice versa. As some researchers have pointed out, certain traffic stops may be investigative stops: that is, on the pretext of a traffic violation, an officer may stop a driver to investigate a crime. These researchers point to pretext stops as justification for using crime data as a benchmark for traffic stops. But analyses based on this justification fail because of their inability to identify whether bias is a factor in police decision making.

Researchers conducting crime data benchmarking must decide carefully what measures of crime to use. To assess whether racial profiling in their jurisdiction exists, the researchers compare the racial/ethnic profile of drivers stopped by police in an investigation of possible criminal activity (the numerator or investigative stop data) to the racial/ethnic profile of people who appear in recorded data on crime in the jurisdiction (the denominator or crime data). Therefore, the first criterion for viable measures of crime is that they be linked to the race/ethnicity of the suspect or perpetrator. The second criterion is that the measures reflect as closely as possible *actual crime* as opposed to crime responded to by police. Chapter 10 discusses how agencies can meet this second criterion. It also describes how crime data were used in the comprehensive analysis of "stops and frisks" by the New York City Police Department.

In crash data benchmarking, researchers compare the racial/ethnic profile of drivers stopped by police (the numerator) to the racial/ethnic profile of drivers involved in crashes (the denominator).[4] We share information on the types and sources of crash data and describe two major studies—one conducted in North Carolina that developed its benchmark using all people

---

[4]Researchers variously describe the denominator data as crash data or vehicle accident data.

involved in crashes (Smith et al. 2003) and another conducted in unincorporated Miami-Dade County that used data only on the drivers adjudged not to be at fault in the crashes (Alpert Group 2003).

Like crash data, data collected for transportation assessment and planning may be useful for producing benchmarks to assess racially biased policing. Transportation data that include information about drivers' driving behavior and race/ethnicity are of the most value to researchers in this regard.

Some researchers have used *survey data* (from written surveys, telephone interviews, or face-to-face interviews) to assess whether policing in a particular jurisdiction is racially biased. The surveys are conducted of scientifically selected residents of the jurisdiction. Respondents are asked about (1) incidents over a specified time period in which they were stopped in their vehicles by police and (2) the quantity, quality, and location of their driving. In effect, these surveys collect both numerator and denominator data. The information on stops can be used instead of police-collected data to measure the nature and extent of vehicle-stopping behavior. The information on driving quantity, quality, and location provides the researcher with information on the various factors that can legitimately affect a driver's risk of being stopped by police. Chapter 10 outlines the advantages and disadvantages of using survey data to assess the existence of racially biased policing. In addition, it reports on the survey methods used by the team analyzing data for the North Carolina State Highway Patrol (Smith et al. 2003) and other examples of using survey data to assess racially biased policing and the perceptions of its practice.

Data from Geographic Information Systems (GIS) can be used by jurisdictions collecting police-citizen contact data. GIS employ a computer system to assemble, store, manipulate, and/or

display data on physical locations (geographic coordinates). Although these spatial data, by themselves, cannot provide evidence that racially biased policing is or is not occurring in a jurisdiction, they can be used to select subsets of stops for analyses and to produce maps for inclusion in a jurisdiction's report on its findings.

We close the chapter by briefly describing data from other analytic tools (such as in-car videos and police queries recorded on Mobile Data Terminals) that researchers have used to assess racially biased policing.

Chapters 5 to 10 present detailed information on benchmarking methods that can be used to address the first of two research questions, "Does a driver's race/ethnicity have an impact on vehicle-stopping behavior by police?" In the next volume, Chapter 11 will address a second research question, "Does a driver's race/ethnicity have an impact on police behaviors/activities during the stop?" The focus is on how to examine the disposition of the stop (citation, arrest, warning, no disposition) and information on search activity (Was a search conducted? On what legal authority? Was anything found?). A key method for assessing bias in post-stop activities is the calculation of "hit rates" for searches. Hit rates represent the percentage of discretionary searches by police that result in finding contraband. Researchers compare search hit rates across racial/ethnic groups to assess whether decisions to search may be racially biased. Chapter 11 details how and why minority hit rates that are *lower* than the corresponding hit rates for Caucasians can raise legitimate concerns about the existence of bias in search decisions.

From Chapter 12 analysts who are not advanced statisticians can learn how to present their results in the form of percentages and how to calculate measures of disparity between racial/ethnic groups including likelihood ratios. The chapter outlines the benefits of multivariate analyses but

cautions that multivariate analyses should not be oversold to agency executives as a method that magically overcomes the major challenges inherent in the quest to measure racial bias.

Chapter 13, the final chapter in *By the Numbers*, will discuss how law enforcement agencies can use the results from data collection to achieve reform. Even results based on weak benchmarking methods can stimulate productive discussions between police and residents about the issues of racially biased policing and the perceptions of its practice. The chapter suggests how these discussions can be structured to produce action plans for reform. We strongly recommend, however, *that agencies focus not merely on measuring racially biased policing but on responding to it*. Varied responses to racially biased policing are set forth in PERF's first DOJ COPS-funded report, *Racially Biased Policing: A Principled Response*, available on the PERF Web site. They can be grouped in the following areas: supervision/accountability, policy, recruitment/hiring, training/education, and outreach to diverse communities.

About the Author

Dr. Lorie Fridell is Director of Research for the Police Executive Research Forum (PERF) and a social scientist by training. Prior to joining PERF in 1999, she was a tenured associate professor of criminology and criminal justice first at the University of Nebraska   and then at Florida State University.   She has been conducting research on law enforcement for more than 15 years and is a national expert on racial profiling. The lead author of *Racially Biased Policing: A Principled Response* (PERF 2001), Fridell also has written extensively on such topics as police use of force, citizen complaints, police pursuits, and problem-oriented policing.

References

The Alpert Group. 2003. *Miami-Dade Racial Profiling Study.* Draft of the methods section of the report to be submitted to the Miami-Dade County (Florida) Police Department.

Fridell, Lorie, Robert Lunney, Drew Diamond, and Bruce Kubu. 2001. *Racially Biased Policing: A Principled Response*. Washington, D.C.: Police Executive Research Forum.

Smith, William R., Donald Tomaskovic-Devey, Matthew T. Zingraff, H. Marcinda Mason, Patricia Y. Warren, and Cynthia Pfaff Wright. 2003. *The North Carolina Highway Traffic Study*. Final report submitted to the National Institute of Justice, Grant No. 1999-MU-CX-0022. Washington, D.C.: National Institute of Justice.

Zingraff, Matthew T. and H. Marcinda Mason, William R. Smith, Donald Tomaskovic-Devey, Patricia Warren, Harvey L. McMurray, and C. Robert Fenlon. 2000. *Evaluating North Carolina State Highway Patrol Data: Citations, Warnings and Searches in 1998*. Report submitted to the North Carolina Department of Crime Control and Public Safety and North Carolina State Highway Patrol. Available online at http://www.nccrimecontrol.org/shp/ncshpreport.htm.