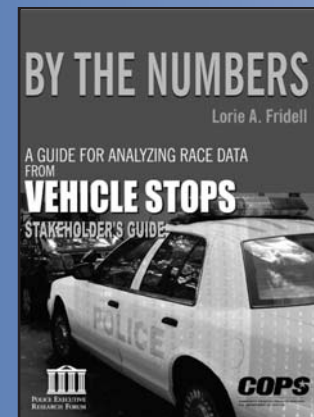


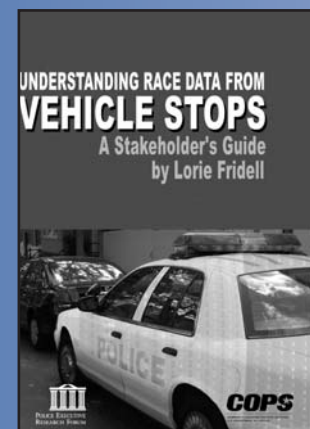
RACIALLY BIASED POLICING: GUIDANCE FOR ANALYZING RACE DATA FROM VEHICLE STOPS

BY THE NUMBERS: A GUIDE FOR ANALYZING RACE DATA FROM VEHICLE STOPS



AND

UNDERSTANDING RACE DATA FROM VEHICLE STOPS: A STAKEHOLDER'S GUIDE



EXECUTIVE SUMMARY

BY LORIE A. FRIDELL

Racially Biased Policing: Guidance for Analyzing Race Data from Vehicle Stops

EXECUTIVE SUMMARY

OF

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

AND

Understanding Race Data from Vehicle Stops:
A Stakeholder's Guide

Lorie A. Fridell



The two reports are available as free downloads from the websites of Police Executive Research Forum (www.policeforum.org) and the U.S. Department of Justice Office of Community Oriented Policing Services (www.cops.usdoj.gov).

This project, conducted by the Police Executive Research Forum, was supported by Cooperative Agreement #2001-CK-WX-K046 by the U.S. Department of Justice Office of Community Oriented Policing Services (COPS). Points of view or opinions contained in this document are those of the author and do not necessarily represent the official position of the U.S. Department of Justice or the members of PERF.

Police Executive Research Forum, Washington, D.C. 20036

Published 2005

Cover art adapted from design by Marnie Kenney

Interior design by David Williams, adapted from design by Automated Graphic Systems

Foreword

Issues related to racism in American society have been debated for decades, and law enforcement has not been exempt from assertions that race has been an inappropriate factor in how they provide services to their communities. In the late 1990s there were renewed allegations of disparate treatment of minority citizens by police and with those allegations emerged a new label—“racial profiling.” Efforts to address racial profiling and perceptions of racial profiling have been marked by controversy and growing tensions over the intervening years. To enhance law enforcement responses to the issues surrounding racial profiling, also referred to as “racially biased policing,” the Police Executive Research Forum (PERF) and the U.S. Department of Justice Office of Community Oriented Policing Services (COPS Office) have partnered to provide resources to support law enforcement and the communities they serve in their efforts to more effectively address racial issues.

Many agencies around the country are collecting data on drivers’ race when their officers make vehicle stops as one component of their response to racially biased policing and perceptions of its practice. The primary purpose of these data collection efforts is to assess whether racially biased policing is occurring in the jurisdiction. To ensure that these efforts are undertaken responsibly and effectively, PERF and COPS have produced two documents that are summarized in this executive summary. The first book is entitled *By the Numbers: A Guide for Analyzing Race*

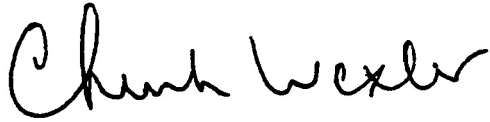
Data from Vehicle Stops and the second is *Understanding Race Data from Vehicle Stops: A Stakeholder’s Guide*. *By the Numbers* is a detailed how-to guide on data collection and analysis. It is written for the people—usually social scientists—who will actually be conducting the analyses and issuing the reports. In contrast, the *Stakeholder’s Guide* addresses the same topics, but is written for the stakeholders who will make or otherwise have an impact on decisions regarding data collection, and who will be the consumers of the reports emanating from those efforts. This includes law enforcement chief executives; local, state and federal policy makers; advocacy groups; the media; and other concerned community members.

These documents were developed with the assistance of an advisory board made up of both law enforcement practitioners and social scientists. These professionals helped to outline the two documents and many read chapters as they were completed. Of particular assistance were the social scientists around the country who are analyzing and interpreting police-citizen contact data for various jurisdictions and who have been instrumental in advancing the methods used to assess racial bias. The PERF author, Lorie Fridell, read the draft and completed reports produced by these social scientists and engaged in considerable discussions with these experts about vexing and controversial issues. The accumulated wisdom of the researchers working in this realm has been documented in the two PERF/COPS books so

that agencies and other entities analyzing vehicle stop data can learn from their advances, as well as missteps.

Police and other stakeholders must collaborate to identify concerns about law

enforcement practices and think comprehensively about how they will be resolved. PERF and the COPS Office hope the documents described here will significantly advance these efforts.

A handwritten signature in black ink that reads "Chuck Wexler". The script is fluid and cursive, with the first name "Chuck" being more prominent than the last name "Wexler".

Chuck Wexler
Executive Director
PERF

A handwritten signature in black ink that reads "Carl Peed". The signature is written in a bold, cursive style, with the first name "Carl" and the last name "Peed" clearly distinguishable.

Carl Peed
Director
COPS Office

Executive Summary

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. At this writing, approximately half of the states have 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 and/or reported conclusions that were not supported by the analyses and results. Most agencies are using models for their analyses that fall short of minimal social science standards. In jurisdictions across the country, reports prepared by agencies or community groups draw conclusions wholly unsupported by the data. These failures can largely be explained by the complexity of the task of measuring whether policing in a jurisdiction is racially biased. A 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 the strengths and weaknesses of the various benchmarking methods is hindering agencies’ efforts to reach valid, responsible conclusions.

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

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.

these agencies understand how to analyze and interpret their data in a manner that reflects accepted social science standards. Two books are available to assist in this effort. *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* is a “how to” guide written for people actually analyzing the data; it describes in detail how to analyze and report vehicle stop data. *Understanding Race Data from Vehicle Stops: A Stakeholder's Guide* summarizes the content of *By the Numbers* so that people who have a *stake* in data analysis but who are not themselves *conducting it* can understand the material. Both books discuss the challenge of benchmarking, how to assess the quality of benchmarks, various benchmarking options that jurisdictions can choose, and how to interpret the research results responsibly.

The purposes of both of these COPS-supported books 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 both books 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, researchers must exclude or “control for” factors other than race/ethnicity that might legitimately explain police stopping decisions. For example, the reports of virtually all jurisdictions regarding their police-citizen contact data show 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 most people 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,³ and

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 assuming no bias.

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

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 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 of the disparity.

In both books 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. The author emphasizes 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.⁴ She recommends that decision makers select all traffic stops or all vehicle stops, and not a subset of these categories as defined by their

³ Concerned stakeholders have asked the author whether the unstated implication of this assumption is that minorities violate more. Indeed, no direction is implied by its inclusion. Minorities may violate traffic laws with *less* frequency than do majority populations. (In fact, this could be the case in light of minorities’ concern about racial profiling and the increased attention they perceive they get from police.) If minorities do violate less, then it is important that this information be incorporated into the analysis to appropriately determine the rate at which they should be stopped by police

in light of their driving quality. Driving behavior cannot be removed from our analysis unless there is clear evidence in support of the null hypothesis (no differences between racial/ethnic groups exist).

⁴ 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). This book is available on the PERF website at www.policeforum.org.

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. Chapter 3 explains why this practice produces substandard data for analysis.

In Chapter 3 the author also encourages 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 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. 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, the author sets 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.

The author starts 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, by using the methods described in the chapter, researchers can 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 the author recommends 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 the author describes 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. Viable subsets include those based on (1) whether stops are proactive or reactive, (2) whether the officer could discern the driver's race/ethnicity, (3) geographic locations of stops (to allow for analyses within subareas of the jurisdiction), and (4) 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 the author calls "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.

Chapter 5 in the *Stakeholder's Guide* describes the various major benchmarks used to analyze stop data. In *By the Numbers* these methods are described in much more depth in Chapters 5 through 10. The chapter(s) attempts to help agencies avoid some of the frequent mistakes associated with benchmarking. 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.

The benchmarking chapter(s) covers the following topics:

- Benchmarking with Adjusted Census Data
- Benchmarking with DMV Data
- Benchmarking with Data from “Blind” Enforcement Mechanisms
- Benchmarking with Data for Matched Officers or Matched Groups of Officers
- Observation Benchmarking
- Other Benchmarking Methods.

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 of both books 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 or of jurisdiction

residents of driving age. A straight comparison between the demographics of these two groups is called “unadjusted” census benchmarking—a method that is not recommended. 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.

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.

Next, each book 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. This method is comparable to adjusting census data for vehicle ownership. To implement this method, drivers' license data in the state must be linked to racial/ethnic information.

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.

The books also describe 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). 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.

Several other analysis methods are based on the assumptions underlying benchmarking with “blind” enforcement mechanisms. For instance, some agencies compare stops in which officers exercise a high degree of discretion to low-discretion stops and a team from the RAND Corporation benchmarked “daylight” stops against “darkness” stops in Oakland, CA. These benchmarking methods also are explained.

Another benchmarking method involves comparing 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 for implementing this method 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 officer’s exposure to drivers at risk of being stopped; *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: 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, another benchmarking method discussed in Chapter 5 of *Stakeholder’s Guide* and Chapter 9 of *By the Numbers*, 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:

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

The coverage of this method also explains how social scientists have addressed these questions in the context of their research.

The numerator and denominator data 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 produce a strong benchmark representing the people at risk of being stopped by police absent bias. 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.

To complete the coverage regarding ways to benchmark stops, each book briefly describes

- Crime data benchmarking,
- Crash (auto accident) data benchmarking,
- Transportation data benchmarking, and
- Survey data benchmarking.

Researchers can benchmark police stop data against crime data, but only certain stops by police can be used in this analysis. Benchmarks based on crime data can 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.

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 data on crime in the jurisdiction (the denominator or crime data). 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.

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). The author presents information on the types and sources of crash data and describes two major studies—one conducted in North Carolina that developed its

benchmark using all people 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).

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. The author outlines the advantages and disadvantages of using survey data to assess the existence of racially biased policing.

Overall, Chapter 5 and Chapters 5 through 10 in *Stakeholder's Guide* and *By the Numbers* respectively 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?" Chapter 6 of the *Stakeholder's Guide* and Chapter 11 of *By the Numbers* address a second research question, "Does a

driver's race/ethnicity have an impact on police behaviors/activities during the stop?" In each book this chapter describes methods for analyzing search and stop disposition (for instance, citation, arrest, warning) decisions. The chapters explain that many stakeholders have inappropriately drawn conclusions regarding the existence or lack of racially biased policing from "percent searched" data. "Percent searched" measures are produced by calculating for each racial/ethnic group the percentage of stopped drivers who are searched. If during a specified period, 100 minorities were stopped in their vehicles and 20 of them were searched, the percent searched is 20. In many jurisdictions higher proportions of stopped minorities are searched than stopped Caucasians. However, analysts, stakeholders, and reporters are mistaken if they conclude that this disparity between the frequency of searches of minorities and searches of Caucasians necessarily indicates bias on the part of police. Such conclusions are not supported by "percent searched" information. "Percent searched" information may show disparity, but it cannot identify the cause of disparity between searches of racial/ethnic groups.

Another way to analyze searches is to look at "hit rates." A hit rate is the percent of searches in which the officers find something (for instance, contraband or other evidence of crime) upon the people being searched. Agencies compare the hit rates across racial/ethnic groups. For all types of searches, hit rates provide descriptive information regarding whether or not there is disparity in the "productivity" of searches. This is valuable information, but bias may not be the cause of this disparity. The chapter explains, however, that for certain categories of searches, hit rates provide more valuable information. Lower hit rates for minorities than for Caucasians for certain categories of searches are cause for concern. These results are a warning signal or "red flag" requiring

the serious attention of law enforcement agencies. They are, however, not *proof* of racially biased policing.

For “evidence-based searches” researchers can say with reasonable confidence that any identified disparity in hit rates across racial/ethnic groups is unjustified and likely (although not certainly) caused by bias. The “outcome test” from economic theory explains why the hit rates for evidence-based searches provide us with this important information. The author explains the outcome test and its application to police searches.

The author also provides information on how to analyze and review stop disposition data. Most agencies collecting vehicle stop data obtain information from officers on the disposition of the stop, such as whether the officer gave the driver a ticket, made an arrest, provided a verbal or written warning, or gave “no disposition.” In their analysis of disposition data, like vehicle stop data, researchers can identify “disparity” in police actions or the lack thereof. They can calculate the percentage of various dispositions across drivers within various racial groups. Like the “percent searched” data, however, disposition data can identify disparity in police actions but not the cause of that disparity. Not all stopped drivers are at equal risk of receiving the various dispositions. In disposition data analysis, the more legitimate factors the researcher can rule out for the officers’ choice of disposition, the more confidence the researcher can have that disparity in police decisions is due to bias. The chapters describe various ways that researchers can rule out some of the legitimate factors that might impact on officer disposition decisions and convey results appropriately and responsibly.

Chapter 7 (*Stakeholder’s Guide*) and Chapter 12 (*By the Numbers*) report how researchers can present their results. In the previous chapters discussing the benchmarking of stops and the analysis of poststop data, the author reported on how researchers

can produce results that may or may not show “disparity” across racial/ethnic groups. These chapters begin by explaining four ways that disparity can be conveyed: through absolute differences in percentages between those stopped by police and the benchmark population, relative differences in percentages, disparity indexes, and ratios of disparity. In their analysis of stop, search, and disposition data, researchers can choose one or more of these measures of disparity.

Social scientists analyzing vehicle stop data have differences of opinion regarding whether researchers should report multiple measures of disparity or just one. Those who advocate the selection and reporting of a single measure (for instance, the disparity index) point out that multiple measures could confuse the residents, policy makers, and other stakeholders who read the agency’s report. Other social scientists favor reporting two, three, or even all four of the measures of disparity. They claim it is better to provide report consumers with more information, not less, including information on how various measures can produce different results in different circumstances. Indeed, different measures *do* produce different results, and researchers and jurisdiction stakeholders need to understand this important fact. If the percentages of minorities (or of Caucasians) in the population of stopped drivers or in the benchmark population are not very high or very low, the researcher’s choice of one measure of disparity over another will not have strong ramifications for the results. On the other hand, when a researcher is dealing with very high or very low percentages of minorities (or of Caucasians), the selection of one measure over another will lead to very different interpretations of the results.

The chapters describe how researchers might use contingency tables to identify disparity and outline the benefits of multivariate analyses, but caution 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.

When does disparity equate to bias? There is no simple answer to this question. Some researchers set a cut-off point: they decide that disparity levels above this point indicate racial bias. Others believe it is impossible, and therefore inappropriate, to set a cut-off point. The author evaluates these opinions and explains useful tools that researchers can use to interpret data.

Chapters 7 and 12 explain measures of disparity and how they can be calculated. It does not provide definitive answers about when policing in a jurisdiction is characterized by racial bias. A theme of both books is that researchers can measure disparity easily, but identifying the cause of disparity presents a challenge. That theme continues through these chapters. No calculations of measures of disparity—however advanced—will themselves overcome this challenge. Those who have a stake in the results of benchmarking analysis—residents, local officials, members of the media, advocates for minorities, and others—seek definitive answers about whether policing in their jurisdiction is racially biased, but those definitive answers cannot be given. The reason is the impossibility of ruling out all of the legitimate (non-bias) factors influencing police decisions to stop a vehicle, conduct a search, or give a disposition (that is, arrest the driver, ticket the driver, warn the driver, or provide no disposition to the stopped driver). Benchmarking analysis can signal the possibility of biased policing, motivate jurisdictions to explore policing practices, and improve relations between police and the community. Definitive conclusions, however, cannot be drawn from the results.

The final chapter in each of the two books begins by acknowledging that stakeholders may well be frustrated by the message that vehicle stop data cannot be used to prove or

disprove racially biased policing. The concerned stakeholder might ask: Of what value are these results if researchers cannot report, with confidence, the existence or lack of racial bias in the jurisdiction? The answer is that they *can* be of significant value. These results can serve as a basis for constructive dialogue between police and residents, which can lead to (1) increased trust and cooperation and (2) action plans for reform. The chapter describes various ways that police and resident stakeholders can come together to reflect on the results of data collection efforts. Their ultimate aim is mutual understanding and reform. Specifically, the chapters indicate

- who should be brought together;
- what information—including vehicle stop and poststop results—this group might explore; and
- the types of changes the group might recommend.

As articulated by Chief John Timoney (2004) of the Miami Police Department, the reality is that “race is a factor in policing.” Every police executive needs to consider and address the issues of racially biased policing and the perceptions of its practice. Because all agencies can make progress on this issue and because the data will never “prove” or “disprove” racially biased policing, the author contends that vehicle stop data collection and analysis should never be viewed—either by police or resident stakeholders—as a “pass-fail test.” Instead, it should be viewed as a diagnostic tool to help pinpoint the decisions, geographic areas, and procedures that should get priority attention when the agency, in concert with concerned residents, identifies its next steps for addressing the problem or perception of racial profiling. The change initiatives outlined by the agency in cooperation with citizens might be specific to a particular “finding” (e.g., the disproportionate representation of minorities among those asked for

consent to search), or they might be of a general nature. The author provides examples of such initiatives and describes varied responses to racially biased policing that are also set forth in PERF's first DOJ COPS-funded report, *Racially Biased Policing: A Principled Response*, available on the PERF website (www.policeforum.org). They can be grouped in the following areas: supervision/accountability, policy, recruitment/hiring, training/education, and outreach to diverse communities.

References

- Alpert Group. 2003. *Miami-Dade Racial Profiling Study*. Draft of the methods section of the report to the Miami-Dade County (FL) 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.
- Timoney, John. 2004. Panelist at "Law Enforcement Use of Force" webcast discussion sponsored by the Office of Community Oriented Policing Services at their 2004 National Community Policing Conference, June 22.

Resources

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

McMahon, Joyce, Joel Garner, Captain Ronald Davis, and Amanda Kraus. 2002. *How to Correctly Collect and Analyze Racial Profiling Data: Your Reputation Depends on It! Final Report for Racial Profiling–Data Collection and Analysis*. Washington, D.C.: Government Printing Office. Available online at <http://www.cops.usdoj.gov/Default.asp?Open=True&Item=770>.

McMahon, Joyce and Amanda Kraus. 2005. *A Suggested Approach to Analyzing Racial Profiling: Sample Templates for Analyzing Car-Stop Data*. Washington, D.C.: Government Printing Office. Available online at: <http://www.cops.usdoj.gov/mime/open.pdf?Item=1462>.

Northwestern University Racial Profiling Data Collection Resource Center. Online at <http://www.racialprofilinganalysis.neu.edu/>.

About the Author

Dr. Lorie A. Fridell, a social scientist by training, was Director of Research for the Police Executive Research Forum (PERF) from 1999 to 2005. Prior to joining PERF, she was a 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, violence against police, and problem-oriented policing. She is currently an Associate Professor of Criminology at the University of South Florida in Tampa.

About the Office of Community Oriented Policing Services (COPS) U.S. Department of Justice

The Office of Community Oriented Policing Services (COPS) was created in 1994 and has the unique mission to directly serve the needs of state and local law enforcement. The COPS Office has been the driving force in advancing the concept of community policing, and is responsible for one of the greatest infusions of resources into state, local, and tribal law enforcement in our nation's history.

Since 1994, COPS has invested over \$11.4 billion to add community policing officers to the nation's streets, enhance crime fighting technology, support crime prevention initiatives, and provide training and technical assistance to help advance community policing. COPS funding has furthered the advancement of community policing through community policing innovation conferences, the development of best practices, pilot community policing programs, and applied research and evaluation initiatives. COPS has also positioned itself to respond directly to emerging law enforcement needs. Examples include working in partnership with departments to enhance police integrity, promoting safe schools, combating the methamphetamine drug problem, and supporting homeland security efforts.

Through its grant programs, COPS is assisting and encouraging state, local, and tribal law enforcement agencies to enhance their homeland security efforts using proven community policing strategies. COPS programs such as the Universal Hiring Program (UHP) has helped agencies address terrorism preparedness or response through community policing. The COPS in Schools (CIS) program has a mandatory training component that includes topics on terrorism prevention, emergency response, and the critical role schools

can play in community response. COPS also developed the Homeland Security Overtime Program (HSOP) to increase the amount of overtime funding available to support community policing and homeland security efforts. Finally, COPS has implemented grant programs intended to develop interoperable voice and data communications networks among emergency response agencies that will assist in addressing local homeland security demands.

The COPS Office has made substantial investments in law enforcement training. COPS created a national network of Regional Community Policing Institutes (RCPIs) that are available to state, local, and tribal law enforcement, elected officials and community leaders for training opportunities on a wide range of community policing topics. Recently the RCPIs have been focusing their efforts on developing and delivering homeland security training. COPS also supports the advancement of community policing strategies through the Community Policing Consortium. Additionally, COPS has made a major investment in applied research which makes possible the growing body of substantive knowledge covering all aspects of community policing.

These substantial investments have produced a significant community policing infrastructure across the country as evidenced by the fact that at the present time, approximately 86 percent of the nation's population is served by law enforcement agencies practicing community policing. The COPS Office continues to respond proactively by providing critical resources, training, and technical assistance to help state, local, and tribal law enforcement implement innovative and effective community policing strategies.

About PERF

The Police Executive Research Forum (PERF) is a national professional association of chief executives of large city, county and state law enforcement agencies. PERF's objective is to improve the delivery of police services and the effectiveness of crime control through several means:

- the exercise of strong national leadership,
- the public debate of police and criminal justice issues,
- the development of research and policy, and
- the provision of vital management and leadership services to police agencies.

PERF members are selected on the basis of their commitment to PERF's objectives and principles. PERF operates under the following tenets:

- Research, experimentation and exchange of ideas through public discussion and debate are paths for the development of a comprehensive body of knowledge about policing.
- Substantial and purposeful academic study is a prerequisite for acquiring, understanding and adding to that body of knowledge.
- Maintenance of the highest standards of ethics and integrity is imperative in the improvement of policing.
- The police must, within the limits of the law, be responsible and accountable to citizens as the ultimate source of police authority.
- The principles embodied in the Constitution are the foundation of policing.