

UNDERSTANDING RACE DATA FROM **VEHICLE STOPS** A STAKEHOLDER'S GUIDE by Lorie A. Fridell



POLICE EXECUTIVE
RESEARCH FORUM



COMMUNITY ORIENTED POLICING SERVICES
U.S. DEPARTMENT OF JUSTICE

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

Lorie A. Fridell



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

Library of Congress Number 2005905945

ISBN 1-878734-89-X

Cover art adapted from design by Marnie Kenney

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

Contents

Chapter 1: Introduction	1	Foreword	v
Chapter 2: The Benchmarking Challenge	5	Acknowledgements	vii
Chapter 3: Getting Started	21	References	89
Chapter 4: Data Analysis Guidelines for All Benchmarking Methods	29	Resources	92
Chapter 5: Methods for Benchmarking Stop Data	35	About the Author	93
Chapter 6: Guidelines for Analyzing Poststop Activities by Police	55	About COPS	94
Chapter 7: Drawing Conclusions from the Results	65	About PERF	95
Chapter 8: Using the Results for Reform	79		

Foreword

Longstanding tensions between police and minority groups within the community received renewed attention in the late 1990s with charges that law enforcement engages in “racial profiling.” Although allegations of disparate treatment of minority citizens by police are not new, the police response in the 21st Century to these issues is superior to that witnessed in previous eras. Around the country, executives of law enforcement agencies (public safety directors, chiefs and sheriffs) are joining with concerned residents to address the important issues of racially biased policing and the perceptions of its practice.

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 these efforts. With COPS Office funds, PERF first published *Racially Biased Policing: A Principled Response*, which provides law enforcement executives with guidance in implementing a comprehensive response to these important issues. Chapters address change efforts that can occur in the realms of policy, supervision and accountability, education and training, recruitment and hiring, and outreach to diverse communities. The final chapter in the book addresses data collection on vehicle stops. Many agencies around the country are collecting data on the stops of drivers made by their officers. Officers record information—usually on paper forms—regarding the race/ethnicity and other demographic characteristics of the person they have stopped, the reason for the stop, whether a search was conducted, the disposition of the stop (e.g., arrest, ticket), and so forth. The primary purpose of these data collection efforts is to

assess whether racially biased policing is occurring in the jurisdiction.

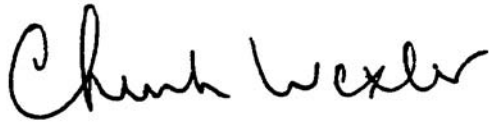
PERF and the COPS Office joined to produce two resources to guide data collection efforts. The first book is entitled *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* and the second is this one, *Understanding Race Data from Vehicle Stops: A Stakeholder’s Guide*. *By the Numbers* 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, this 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.

There were two main purposes for writing both books. One was to temper people’s expectations regarding the conclusions that can be reached with vehicle stop data. Many concerned stakeholders have been too optimistic regarding the ability of these data to determine the existence or lack of racially biased policing. Both books describe the potential and constraints of these data for measuring racially biased policing. The second purpose was to describe the various ways that data can be analyzed and the conclusions that can and cannot be drawn from the results.

We hope that this guide will assist stakeholders in making informed decisions regarding data collection efforts, understanding the conclusions that can and cannot be

drawn from the results, and using the results for constructive change. We also hope that this book will lead stakeholders to understand that data collection is only *one response* to the critical issue of racially biased policing. Data collection seeks to *measure* racially biased policing (and is limited in its ability to achieve that objective). Chiefs and other stakeholders should implement measures to *address* (not just measure) the problem of

racially biased policing and perceptions of its practice (for instance, through policy, training, supervision, and community outreach). 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 this document, along with the others in the series, will significantly advance these efforts.



Chuck Wexler
Executive Director
PERF



Carl Peed
Director
COPS Office

Acknowledgements

On behalf of PERF, I convey great appreciation to Director Carl Peed, Deputy Director Pam Cammarata, and other leaders in the U.S. Department of Justice Office of Community Oriented Policing Services (COPS Office). They shared our belief that there was a significant need for *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* and for this companion stakeholder’s guide. They provided the necessary funding to bring these documents to fruition. I am particularly grateful for the very constructive guidance and unwavering support I received throughout this project from Tamara Clark Lucas and Amy Schapiro, our COPS Office monitors.

Since this *Stakeholder’s Guide* is an abbreviated version of *By the Numbers*, I acknowledge here the people who were instrumental in making *By the Numbers* a reality and, by extension, the *Stakeholder’s Guide* as well. An advisory board made up of both practitioners and social scientists met early on in the project to guide the content and form of *By the Numbers* and the *Stakeholder’s Guide*. Many read chapters or sections of *By the Numbers* as they were completed. Members of this board included Dr. Gary Cordner, Eastern Kentucky University; Dr. Scott Decker, University of Missouri, St. Louis; Assistant Chief John Diaz, Seattle (WA) Police Department; Captain Thomas Dione, Montgomery County (MD) Police Department; Dr. Don Faggiani, PERF; Dr. Amy Farrell, Northeastern University; Secretary Ed Flynn, Massachusetts Executive Office of Public Safety; Assistant Chief John Gallagher, Miami (FL) Police Department; Former Chief Mark Kroeker, Portland (OR) Police Bureau; Dr. John Lamberth, Lamberth Consulting; Dr. Catherine Lawson, State University of New York at Albany; Dr. Steve Mastrofski, George

Mason University; Dr. Jack McDevitt, Northeastern University; Antony Pate, Private Consultant; Former Assistant Deputy Superintendent Karen Rowan, Chicago (IL) Police Department; Professor Margo Schlanger, Harvard University; Dr. Ellen Scrivner, Former Deputy Director of the COPS Office; Dr. Deborah Thomas, University of Colorado at Denver; Dr. Sam Walker, University of Nebraska at Omaha; Dr. Charles Wellford, University of Maryland; and Dr. Matt Zingraff, North Carolina State University.

A group of academics—most of them members of the advisory board—stand out for their contributions to the content of both books. They are the key social scientists around the country who are analyzing and interpreting police-citizen contact data for various law enforcement entities and who have been instrumental in advancing the methods used to assess racial bias. Members of this group were generous with their time and knowledge. They shared their draft and completed reports from their own research, responded to phone and email inquiries, served as a “sounding board” for particularly vexing issues, and reviewed critically the various chapters of *By the Numbers* that reported on their work and/or otherwise reflected most closely their experience and expertise. In alphabetical order, these people are

Geoff Alpert, *University of South Carolina*
 Gary Cordner, *Eastern Kentucky University*
 Scott Decker, *University of Missouri at St. Louis*
 Robin Engle, *University of Cincinnati*
 Amy Farrell, *Northeastern University*
 David Harris, *University of Toledo*
 John Lamberth, *Lamberth Consulting*

Catherine Lawson, *State University of New York at Albany*

Jack McDevitt, *Northeastern University*

Michael Smith, *University of South Carolina*

Samuel Walker, *University of Nebraska at Omaha*

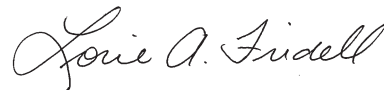
Matthew Zingraff, *North Carolina State University*

It is no exaggeration to say that there would be no content for these documents without their research efforts in the field and no summaries of the promising approaches without their direct assistance.

Also assisting me in various ways were Ian Ayres of Yale University; R. Richard Banks of Stanford Law School; Deputy Chief Michael Berkow of the Los Angeles Police Department; Jennifer Calnon of Pennsylvania State University; Jerry Clayton and Karl Lamberth of Lamberth Consulting; Rod Covey, Former Assistant Director of the Arizona Department of Public Safety; Scott Henson of the ACLU of Texas; Chief Ellen Hanson of the Lenexa Police Department; Chief Glenn Ladd of the North Kansas City Police Department; Nicholas Lovrich of Washington State University; Mike Maltz of the University of Illinois at Chicago; Ken Novak of the University of Missouri, Kansas City; Tim Oettmeier,

Executive Assistant Chief of the Houston Police Department; Nicola Persico of the University of Pennsylvania; Jeff Rojek of St. Louis University; Matt Salazar of the Austin Police Department; and Jane Wiseman of the Massachusetts Department of Public Safety. To all of these people, I convey my sincere appreciation.

I owe a tremendous debt to Barbara de-Boinville who provided expert editing of both books—ensuring that complicated, sometimes dense material was organized, accessible and clear. At PERF my thanks go to Executive Director Chuck Wexler who pushed me to complete the documents while being completely supportive with the encouragement and resources I needed to do so. I'm grateful to Martha Plotkin and David Edelson who provided their usual high-quality, expert advice on both the form and content of both documents and oversaw all aspects of their publication. Also at PERF, colleague Don Faggiani, reviewed various sections of the documents and was on call day and night to provide advanced methodological guidance. Finally, special thanks to all of my staff in the Research Unit at PERF who are wonderful in many, many ways and, for this particular project, thoughtfully and patiently let me hole up without interruption for days at a time to write.



Lorie A. Fridell

I Introduction

Racially biased policing and the perceptions of its practice are critical issues facing jurisdictions across the country. The issues involved in “racial profiling” and racially biased policing are not new. They are the latest manifestations of a long history of sometimes tense, and even volatile, relations between police and minorities. This need not be viewed, however, as proof of the problem’s intractability. Police are more capable than ever of effectively addressing police racial bias in their ranks. In the past few decades there has been a revolution in the quality and quantity of police training, the standards for hiring officers, procedures and accountability mechanisms, and the widespread adoption of community policing.

In the Police Executive Research Forum’s first book on this topic, *Racially Biased Policing: A Principled Response* (Fridell et al. 2001),¹ the various ways that police can and should respond to racially biased policing and the perceptions of its practice were set

forth.² Specifically, the book discussed methods of reform and prevention in the areas of accountability and supervision, policies to address racially biased policing, recruitment and hiring, education and training, outreach to minorities in the community, and the collection of data on police-minority contacts. Interestingly, this latter intervention, the collection of vehicle stop data, was the response of choice for many jurisdictions. Many jurisdictions have started collecting data on the race/ethnicity of drivers stopped and/or searched by police.

The collection of data reflects positively on policy makers in several ways. It shows their commitment that biased policing will not be tolerated, and it conveys a willingness on the part of the law enforcement agency to be held accountable to the public for its actions. In addition to wanting to convey concern and accountability, policy makers promote data collection because they want to determine the nature and extent of racially

¹ This book, funded by the U.S. Department of Justice Office of Community Oriented Policing Services (COPS) (Grant 1999-CK-WX-0076), is available as a free download from the PERF website at www.policeforum.org.

² Racially biased policing is defined as the inappropriate consideration by law enforcement of race or ethnicity in deciding with whom and how to intervene in an enforcement capacity. Mirroring the U.S. Census we use “ethnicity” to refer to whether a person is of Hispanic or non-Hispanic origin.

biased policing. Although many jurisdictions are eager to analyze data on vehicle stops to achieve this goal, little information has been available to them regarding how to proceed. There also have been overly optimistic expectations regarding the ability of social science methods to turn these data into meaningful conclusions regarding the existence of racially biased policing.

To address these issues, PERF with funding from the COPS Office has published *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* (Fridell 2004). *By the Numbers* is written for the people who are actually conducting the analyses of the data—whether they are associated with police departments, academic institutions, or stakeholder groups. This technical guide provides practical, hands-on instruction for analyzing both stop and poststop data, and it helps the researcher understand what conclusions can and cannot be drawn based on the results.

Understanding Race Data from Vehicle Stops: A Stakeholder's Guide, addresses the same issues as *By the Numbers*, but is directed toward a different audience. This book is written for the policy makers who are considering data collection as a response to concerns about racial bias in policing in their jurisdiction or who are otherwise linked to data collection efforts, either as consumers of the reports generated by researchers or as participants of jurisdiction task forces on racial profiling. This target population includes executives of state or local law enforcement agencies (for instance, chiefs, sheriffs, superintendents), other state or local policy makers (for instance, mayors, city council members, staff of Attorney's General offices, state or federal legislators), representatives of the media, advocacy groups (for instance, the American Civil Liberties Union, Amnesty International, the NAACP, La Raza, the Urban League), and concerned residents unaffiliated with organized groups.

THE RESPONSIBILITIES OF STAKEHOLDERS

The purpose of this guide is to help the targeted stakeholder groups meet two important responsibilities. First, they need to develop an understanding of both the potential and constraints associated with data collection; stakeholders need to know what conclusions can and cannot be drawn from the data.

Many different people in different professions share a desire to become knowledgeable about the analysis and interpretation of vehicle stop data. They want to gain this information for varied reasons. Agency executives and other policy makers seek information to help them decide whether to collect vehicle stop data, what data to collect, and how to “benchmark” the data. They want to know how to interpret and act upon the results. Media representatives want to know how to make sense of competing interpretations of results. Advocacy groups want objective information regarding all of these issues so that they can be constructive in what they convey to their constituencies. Reflecting on this responsibility of advocacy groups, Ron Davis (quoted in McMahan et al. 2002, 96) makes these comments:

Civil rights and community-based organizations ... have the responsibility of obtaining ‘expert’ knowledge and understanding about racial profiling, biased-based policing, and data collection and analysis before launching discrimination allegations. It is a disservice to the community for reputable organizations, whether civil rights or community-based, to accuse law enforcement of racism and/or discrimination based on statistical disparities or the implementation of non-bias traffic enforcement programs.

Although stakeholders should be knowledgeable about the potential and constraints associated with data collection, this is not

their only responsibility. A second responsibility of stakeholders is to work with each other. Law enforcement executives should reach out to other stakeholders for help in designing and implementing the data collection system and for their views on ways the agency can make progress on this long-standing issue affecting police and the community. Residents, local officials—indeed, all parties—need to put aside notions of “we-they” and instead come together in the spirit of cooperation to further the best interests of the jurisdiction as a whole.

This guide will help stakeholders meet the important responsibilities we have described. It provides a clear explanation of the social science challenges associated with data collection initiatives so that readers can separate meaning from myth with regard to police-citizen contact data. This book also explains how police and other stakeholders can come together constructively to conduct data collection/analysis and, more importantly, to address the critical problems of racially biased policing and the perceptions of its practice.

The technical information presented in *By the Numbers* is summarized here so that people who have a *stake* in data analysis but who are not themselves *conducting it* can understand the material. The book discusses 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.

Readers will come to appreciate that analyzing vehicle stop data is a complicated business. The information will affirm what many have already recognized: quality analysis of vehicle stop data is not as simple as comparing vehicle stop information to basic census information. In fact, some frustration will be generated by a major theme of this book: data collection cannot provide unequivocal answers to questions about the

existence or lack of racial bias by police in a jurisdiction. This fact does not, however, negate the value that can be accrued from collecting data on vehicle stops. Even equivocal results can serve as a basis for constructive dialogue between the police and residents—producing enhanced trust and cooperation as well as action plans for reform. Equipped with user-friendly knowledge regarding data collection and analysis, police and other stakeholders can join forces to address racially biased policing and the perceptions of its practice.

CONTRIBUTIONS OF SOCIAL SCIENTISTS

In producing this document and *By the Numbers*, PERF relied upon the valuable expertise of an advisory board. Its members—listed in the acknowledgements section—include the key social scientists around the country who are analyzing and interpreting police-citizen contact data, experienced law enforcement practitioners, and personnel within research units. PERF also has been aided by members of advocacy groups committed to fair and impartial law enforcement. Therefore, the pronoun “we” is used in both books to acknowledge that their contents reflect this collective wisdom.

CONTENTS OF THE BOOK

Chapter 2 describes the social science challenges associated with analyzing and interpreting the police-citizen contact data collected to measure racially biased policing; specifically, it explores the goal, the potential, and the limitations of what has come to be called “benchmarking” the data. Chapter 3, “Getting Started,” explains the steps agencies should take when they begin to collect data on police-citizen contacts. For example, it discusses how to develop a data collection plan, how and why to involve residents and police personnel from all levels of the agency in planning and implementing the

data collection system, and how to select a benchmarking method. Chapter 4 examines issues that are relevant to every benchmarking method, such as the need to review data quality, select a reference period, and analyze subsets of data on the jurisdiction. Chapter 5 presents information on 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 considering this question, a researcher is attempting to understand whether racially biased policing is manifested in the decisions of officers regarding whom to

stop. Chapter 6 addresses a second research question: “Does a driver’s race/ethnicity have an impact on police behaviors/activities during the stop?” It describes how to assess the impact of race/ethnicity on searches and dispositions (for instance, arrest, citation, no action)—activities that occur after the stop is made. In Chapter 7 we describe and compare the various calculations a researcher might use to measure disparity between racial/ethnic groups. Finally, in Chapter 8 we discuss how police and other stakeholders can come together to use the results from data collection to achieve reform.

The Benchmarking Challenge

Jurisdictions collecting police-citizen contact data are calling upon social science to determine whether there is a cause-and-effect relationship between a driver's race/ethnicity and vehicle stopping behavior by police. In analyzing the data, researchers have attempted to develop comparison groups to produce a "benchmark" against which to measure their stop data. If an agency determines that, say, 25 percent of its vehicle stops are of racial/ethnic minorities, to what should this be compared? In other words, what percentage would indicate racially biased policing? This is the question at the core of benchmarking. To determine an answer, researchers have compared the demographic profiles of people stopped by police to the demographic profiles of the residential population of the jurisdiction, to the demographic profiles of residents with a driver's license, and to the demographic profiles of people observed driving on jurisdiction roads—to name a few comparison groups.

THE OBJECTIVE OF BENCHMARKING

Before we discuss the various methods for benchmarking, it is constructive to consider our objectives when analyzing police-citizen contact data. Then we can outline how benchmarks vary in their ability to achieve

these objectives. We start with two conceptual models. Figure 2.1 shows a model of the first research question: Does a driver's race/ethnicity have an impact on the decisions police make with regard to whom to stop? We want to know if X (driver race/ethnicity) has any causal impact on Y (police decisions to stop drivers). To determine causality, however, we must exclude or "control for" rival causal factors—factors other than the race/ethnicity of the driver—that could explain police stopping decisions (see the model in Figure 2.2). In attempting to test whether X causes Y, we need to rule out alternative hypotheses that A, B, C, and Z—either alone or together or in interaction with X—cause Y.

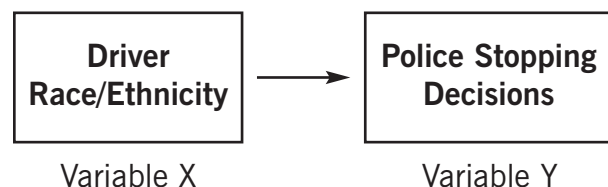


Figure 2.1. Model of First Research Question: Does Driver Race/Ethnicity Affect Vehicle Stopping Decisions Made by Police?

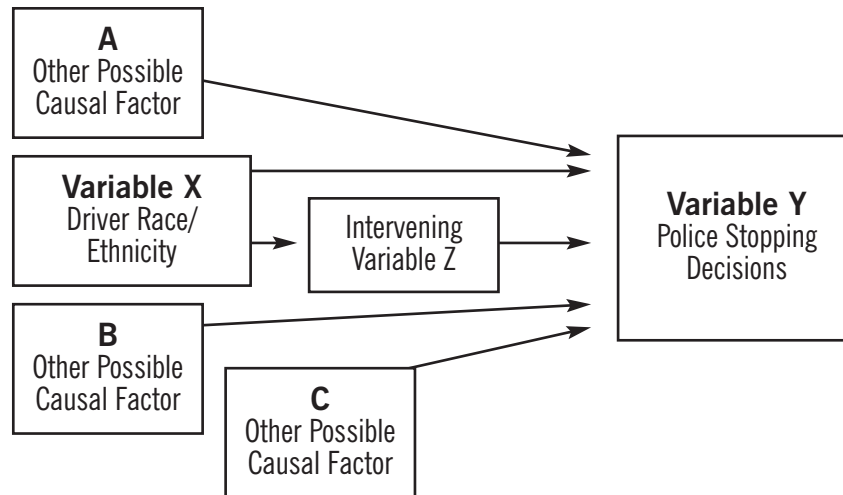


Figure 2.2. Model of Factors, Other than Bias, that Might Affect Stopping Decisions Made by Police

The following example clarifies why rival causal factors must be ruled out in any analysis of police-citizen contact data. Let us say that parents are concerned that the grading by math teachers at a high school reflects teachers' bias against females. The parents' allegation is that these math teachers believe boys are better than girls at math and that—consciously or unconsciously—these attitudes are reflected in the grades being given to the students.

Our basic conceptual model is that gender (X) has a causal impact on the grades given by teachers (Y). To test this scientifically, however, we cannot conduct analyses that consider only X and Y. We cannot, for instance, look only at the percent of females who got A's and B's in a course and the percent of males who got A's and B's in that course and draw any conclusions regarding teachers' gender bias. Instead, we must consider other factors that affect grading behavior. A key variable, of course, would be students' math performance. Our analyses must control for math performance (for example, scores on objective tests). In other words, our research design or statistical techniques must remove or "neutralize"

the impact of performance on grades. If, after we have controlled for math performance, we still find that males get better math grades than do females, then we must seriously consider the possibility of gender bias by teachers.

Now let us return to the first research question concerning who is stopped by police. Police can have various legitimate reasons for deciding to stop a vehicle. These reasons are the rival causal factors that would become the A, B, and C of Figure 2.2. Let's again consider gender but in the context of analyzing police stopping behavior, not math grades.

The reports of most jurisdictions regarding their police-citizen contact data state that men are stopped by police more than women. For instance, a jurisdiction may find 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 on the part of the police? It is unclear from these data, but most of us are disinclined to jump to that conclusion because factors other than police bias could account for the disproportionate stopping of male drivers.

That is, alternative hypotheses for the results exist. Men may drive more than women (the quantity of driving factor). Or men may violate traffic laws more often than women do (the quality of driving factor). A third possibility is that more males than females drive in the areas where stopping activity by police tends to occur (the location of driving 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.

The objective of benchmarking in our example is to see if gender bias is at work. If we could develop a gender profile of the people who *should be* more at risk of being stopped by police, we could compare it to the gender profile of the people who *are* being stopped by police. That is, if we managed through our research design to determine that men *should comprise* 65 percent of the police stops because of their driving quantity, quality, and location, and if indeed they *do comprise* 65 percent of the police stops (based on the stop data collected), then we could report to the jurisdiction that gender bias did not appear to affect stopping behavior by police.

Let us review now the key principles stated above. Researchers' goal is to develop a racial/ethnic profile of the people who should be at risk of being stopped by police in a jurisdiction, assuming no bias. Benchmarking is the essential tool used by researchers to achieve this goal. Benchmarks vary in quality. Their quality 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 following example will help clarify what we mean by benchmark quality. If a researcher uses road-side observers to develop a demographic profile of drivers who violate traffic laws, the researcher has produced a benchmark that represents fairly well the group of people who should be at risk of being stopped by police if no bias exists (at least at the observed location). On the other hand, if that same researcher used instead U.S. Decennial Census data to develop a demographic profile of people who live in the jurisdiction, the researcher has produced a benchmark that does not represent well the people at risk of being stopped by police if no bias exists. The next section on the bias hypothesis and the alternative hypotheses expands upon this discussion of benchmark quality. As we will demonstrate in this report, the variation in quality across benchmarks is great. This means that there are significant differences in what researchers can conclude about racial bias and policing in a jurisdiction. Findings based on a high-quality benchmark are more legitimate than findings based on a low-quality benchmark.

THE BIAS HYPOTHESIS AND THE ALTERNATIVE HYPOTHESES

Alternative hypotheses are hypotheses other than the one that reflects the possibility of police bias. Law enforcement agencies and their partners must consider these hypotheses when analyzing their jurisdiction's police-citizen contact data. These hypotheses reflect drivers' driving quantity, quality, and location—the key factors that could legitimately influence whom police stop.

The benchmarking method chosen by a jurisdiction must be evaluated in terms of its ability to address the alternative hypotheses. Findings that are based on a method that does not address these hypotheses are less valid

than those based on a method that does address them. The results of benchmarking analysis must be interpreted cautiously. Before drawing conclusions about racial bias, researchers must evaluate how successful they were in achieving their goal. Their goal is to find out what the racial/ethnic profile of drivers stopped by police would look like assuming no bias.

To make our point we begin with a simple example. We ask why—in a jurisdiction made up of Caucasians, African Americans, Hispanics, and Asians—the police do not report that 25 percent of their traffic stops are of Caucasians, 25 percent are of African Americans, 25 percent are of Hispanics, and 25 percent are of Asians? One hypothesis is that *police are racially/ethnically biased in their decisions regarding whom to stop*. Competing alternative hypotheses are as follows:

- Racial/ethnic groups are not equally represented as residents in the jurisdiction.
- Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.
- Racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.
- Racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.

In order to draw valid conclusions regarding whether racial bias is occurring in this hypothetical jurisdiction, we would need to rule out the other possible, legitimate explanations for disparity. We refer here to the disparity between the racial/ethnic profile of the people stopped by police for traffic violations and the racial/ethnic profile of the benchmark population. Ideally, our analysis and interpretation of stop data would encompass all of the factors reflected in those alternative hypotheses.¹

Researchers cannot presume that no differences exist between racial/ethnic groups in the quantity, quality, and location of their driving. (That is, they cannot presume the *null hypothesis*.) Below we consider each of the four alternative hypotheses every jurisdiction must address. For each of the hypotheses, there is evidence that differences do exist between groups, or at least there is insufficient information to prove to any acceptable degree of certainty that no differences exist. Unless research shows there are no differences between groups as pertains to these hypotheses, we must assume that there are differences. Again this requires researchers to use methods that consider the factors encompassed in the alternative hypotheses or, at the very least, interpret their results responsibly in light of any deficiencies in their chosen methodology.

¹ If we address the second hypothesis—*racial/ethnic groups are not equally represented as drivers on jurisdiction roads*—we need not concern ourselves with the first hypothesis—*racial/ethnic groups are not equally represented as residents in the jurisdiction*. That is, for purposes of identifying who is at risk of being stopped by police in a vehicle, if we know who is driving on jurisdiction roads, we do not need to know who lives in that jurisdiction. Similarly, addressing the third hypothesis—*racial/ethnic groups are not equivalent in the*

nature and extent of their traffic law-violating behavior—arguably negates the need to address the first two. It can be argued that knowing who is engaging in law-violating behavior negates the need to know who is on the road. Police are not told to pull over “people on the road” but rather “people who are violating laws.” The fourth hypothesis—*racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high*—stands alone and must be addressed independently of the other three.

Hypothesis 1: Racial/ethnic groups are not equally represented as residents in the jurisdiction.

It goes without saying: the demographic profile of people who live in a jurisdiction will affect the demographic profile of drivers on the jurisdiction's roads. Thus, the above hypothesis is indirectly related to the "quantity" factor, and we need to include it in anticipation of our later discussion of census benchmarking (a comparison of the demographic profile of people stopped by police to the demographic profile of jurisdiction residents as measured by the U.S. Census Bureau). That racial/ethnic groups are not equally represented among residents in jurisdictions is, of course, quite obvious to all. According to the 2000 Decennial Census, 75.1 percent of the U.S. population is White, 12.3 percent is Black or African American,² and 3.6 percent is Asian; 9.0 percent of the population self-identify as being of more than one race. Just over 12 percent (12.5 percent) of U.S. residents (of all races) are of Hispanic origin. Although figures for different jurisdictions will deviate from this breakdown of the total U.S. population, we can confidently state that no jurisdiction has equal representation in its population of racial/ethnic groups.

Hypothesis 2: Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.

Not only are racial/ethnic groups not equally represented among residents in the jurisdiction (Alternative Hypothesis 1), but their representation as residents might not match their representation as drivers using jurisdiction

roads for two reasons: (1) racial/ethnic differences in driving quantity and/or (2) racial/ethnic differences in the population of people who do not live in the jurisdiction but drive in it. This is relevant to the analysis of vehicle stops by police. If one demographic group has more presence on the road than another, it should be more at risk of being stopped.

Driving Quantity

There is evidence that racial/ethnic groups differ in the amount of their driving. National data from the U.S. Decennial Census and from the National Household Transportation Survey (NHTS) indicate that racial/ethnic minorities are under-represented as drivers relative to their residential populations.³ The U.S. Decennial Census provides data on the percent of households that do not own vehicles, an indirect measure of driving quantity. In his comprehensive report on commuting patterns based on 1990 Census data, Pisarski (1996, xv) reports that "on average, more than 30 percent of Black households do not own vehicles, and in central cities the number is over 37 percent."⁴ Nationally, 19 percent of Hispanic households do not own vehicles; in central cities that number rises to 27 percent. In contrast, just under 9 percent of White non-Hispanic households are without vehicles, with a corresponding figure of 15 percent for central cities (Pisarski 1996, 36).

Vehicle ownership is an indirect measure of driving quantity. Information from the National Household Transportation Survey provides more direct measures of driving

² The terms *African Americans* and *Blacks* are used interchangeably in this book.

³ The National Household Transportation Survey (previously called the Nationwide Personal Transportation Survey and the American Travel Survey) is conducted by the U.S. Department of Transportation. See www.bts.gov/nhts.

⁴ Some cities have "extraordinary levels of Black households without vehicles" (Pisarski 1996, 36). In New York, 61 percent of Black households are without vehicles. The corresponding figures for Philadelphia, Chicago, and Washington, D.C., are 47 percent, 43 percent, and 43 percent, respectively.

quantity. Its data indicate that nonminorities drive more than minorities. For instance, the 1995 NHTS indicated that African Americans average fewer “trips per day” (including fewer vehicle trips) than do Caucasians and that Hispanics are twice as likely as non-Hispanics to use public transportation (instead of privately owned vehicles).

While the 2000 Census data on vehicle ownership and NHTS data on driving quantity both imply that minorities are under-represented as drivers relative to their representation in the U.S. population, other research reminds us that this is not going to be true in all places at all times. For instance, research conducted by the United Kingdom's Home Office (MVA and Miller 2000) found that minorities were over-represented as drivers relative to their representation in the residential populations in the areas studied.⁵ In Sacramento, California, Howard Greenwald compared the demographic profiles of drivers at various intersections (using observation) to the demographic profiles of residents in the same areas (using census data); he found over-representation of minorities as drivers in some areas and under-representation of minority drivers in others (Greenwald 2001). These two small-scale studies, although of less weight than the large-scale research findings of the NHTS and U.S. Census, nonetheless support our simple point: jurisdiction-level studies of racially biased policing must consider the possibility that racial/ethnic groups are not equally

represented as drivers on jurisdiction roads because of differences in their quantity of driving.

The extent to which racial/ethnic groups drive on roads will vary across locations within a jurisdiction. Because of this fact, we recommend that researchers not conduct a single analysis for the entire jurisdiction but numerous analyses within various geographic subareas of the jurisdiction (see Chapter 4).

Driving by Nonresidents

There is another reason—other than differences in driving quantity of jurisdiction residents—that racial/ethnic groups may not be equally represented as drivers on jurisdiction roads (and why their representation on the roads may not reflect their representation as residents). Racial/ethnic groups may not be equally represented among the *nonresidents* who drive in the jurisdiction; that is, racial/ethnic groups may not be equally represented among the people who live outside of the jurisdiction but drive into it.⁶ The extent to which nonresidents drive within the jurisdictions that are collecting police-citizen contact data will vary greatly, as might the demographic profile of those drivers. The influx of nonresident drivers will be particularly significant in the big cities that draw commuters in from surrounding jurisdictions, especially the suburbs, during the daytime hours.⁷ Additionally, nonresidents will drive into the “target jurisdiction” (the jurisdiction that is

⁵ The Home Office of the United Kingdom is the government department responsible for promoting safe communities. Its closest equivalent in the United States is the National Institute of Justice.

⁶ In its first annual report regarding police–citizen contact data, the Denver Police Department (Thomas 2002) revealed that 62.5 percent of the Whites stopped in their vehicles by police were nonresidents compared to 32.8 percent of the Blacks who were stopped and 35.2 percent of the Hispanics who were stopped.

⁷ In 1993, 43 percent of the traffic tickets given in Seattle were given to nonresidents (Scales 2001). The Denver Police Department (Thomas 2001) reported that from June 2001 through May 2002 (the reference period for its second summary report) over one-half of its traffic stops were of nonresidents. In Louisville (Edwards et al. 2002a) and Iowa City (Edwards et al. 2002b), fewer than two-thirds of all drivers stopped were city residents.

the subject of police-citizen contact data analysis) to shop, seek entertainment, vacation, travel on to another jurisdiction, and for other reasons. These nonresident drivers will affect the demographic profile of drivers on the roads of the target jurisdiction.

The frequency with which drivers are stopped by police is affected by driving quantity. *Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.* This is a viable alternative hypothesis that should be accounted for in the analysis of police-citizen contact data. This book will describe how jurisdictions can incorporate this alternative hypothesis into their study design.

Hypothesis 3: Racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.

Police are directed to stop drivers because of their driving behavior. Therefore, researchers must recognize this variable (traffic law-violating behavior) in their research unless we are quite confident that there are no differences across racial/ethnic groups. Excluding driving behavior from the model is equivalent to excluding math performance from the earlier analysis that tested gender bias in math teachers.

Vehicle stopping behavior by police may not be equivalent across racial/ethnic groups because racial/ethnic groups violate traffic laws at different rates or at different levels of seriousness. These possibilities must be recognized. Concerned stakeholders have questioned the inclusion in our analysis of the third hypothesis (*racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*). They have asked the author whether the unstated implication 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). The following information calls the null hypothesis into question.

Information on the Equivalence of Driving Behavior

Large-scale, quality research on driving behavior and race/ethnicity is scarce, but this does not negate the importance and viability of Alternative Hypothesis 3. In fact, it does just the opposite: what is important for our purposes is the absence of sufficient research to *rule out* the possibility of racial/ethnic differences in the nature and extent of law-violating behavior. Again, even if we had national data pointing to equivalent driving behavior or pointing to one particular direction or the other, we could not presume that those results were applicable to all times and all places.

The information on the equivalence of driving behavior across racial/ethnic groups is mixed. Research in the transportation field, albeit not substantial, indicates some differences across racial groups with regard to certain traffic violations. For instance, Feest (1968) found that Whites were more likely than minorities not to stop at stop signs. Other researchers analyzing police-citizen contact data have produced information indicating other differences in violating behavior across racial/ethnic groups. For instance, Lange, Blackman, and Johnson (2001) found that along segments of the New Jersey turnpike where the speed limit was 65 miles per hour rather than 55 miles per hour, African Americans were disproportionately

represented among the few speeders.⁸ Other researchers have also identified differences in speeding behavior across races (Engel, Calnon, Liu and Johnson 2004 and Smith et al. 2003). In contrast, Lamberth (1996a, 1996b) conducted research in New Jersey and Maryland and found no differences in the demographics of speeders versus nonspeeders. He reports that all racial/ethnic groups were speeding in high, and similar, proportions.⁹

In citing these mixed findings, we are not trying to argue that there are differences in violating behavior across racial/ethnic groups. Quite the contrary: we do not know whether differences exist or not. Because the research does not allow us to rule out the possibility of differences in driving quality across racial/ethnic groups, we contend that research analyzing police-citizen contact data should address the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*.¹⁰

Youthfulness and Driving Behavior

Youthfulness has been linked to law-violating behavior. If a racial/ethnic group has proportionately more young people than another racial/ethnic group, age becomes an important “intervening variable”¹¹ in the analysis model. (It is a potential “Variable Z” in Figure 2.2.) We must consider whether the breakdown of age groups in a jurisdiction (or in the subareas being analyzed) varies across racial/ethnic groups. For example, if 30 percent of

the minority population of an area is young (24 years of age or less) and only 20 percent of the Caucasian population is young, this phenomenon would lead to more drivers who violate the law in the minority population than in the nonminority population, assuming the link between poor driving and age.

When researchers use police-citizen contact data to measure racially biased policing in a jurisdiction, they may get results that suggest bias when none exists. Disproportionate numbers of young drivers in racial/ethnic groups in a jurisdiction can produce misleading results. We illustrate this hazard with hypothetical data in Table 2.1 and Figure 2.3. In our example, two assumptions are made: (1) there are equal numbers of Caucasian and minority drivers on the road in hypothetical Jurisdiction Q, and (2) there is equivalence of driving behavior across these two racial/ethnic groups.

As shown in Table 2.1, Caucasians and minorities each made up 50 percent of the driving population. (There are 1,000 drivers in each group.) Among the Caucasian drivers, 300 or 30 percent were between the ages of 15 and 24, and 700 or 70 percent were 25 or older. (We use age 15 as the lower cut-off point to include only people of driving age.) The corresponding percentages for the minority group of drivers were 60 percent and 40 percent. That is, 600 of the drivers were between the ages of 15 and 24, and 400 were 25 years of age or older.

⁸ This study was criticized for various aspects of its methodology and the high proportion of missing data produced by those methods.

⁹ These studies defined speeding so broadly (1 mile per hour over the speed limit in Maryland and 5 miles per hour over the speed limit in New Jersey) that speeders included most drivers. This broad definition reduced the researcher's ability to detect any existing, finer distinctions in driving behavior across groups.

¹⁰ See *By the Numbers* (Fridell 2004), Appendix D, which challenges this view in the context of discussing the observation method of benchmarking. Appendix E summarizes arguments for and against the methodological consideration of variations in driving quality.

¹¹ We use the term “intervening variable” to refer to a variable (measured or unmeasured) that is linked causally to one or more other variables in an equation or model.

Table 2.1. Representation of Caucasian and Minority Drivers in the Driving Population and Population of Stopped Drivers, by Age, Hypothetical Jurisdiction Q

Age Group	Caucasians (n=1,000)			Minorities (n=1,000)		
	Number of Drivers	Percent Stopped	Number Stopped	Number of Drivers	Percent Stopped	Number Stopped
15-24	300	20%	60	600	20%	120
25+	700	10%	70	400	10%	40
Total	1,000	13%	130	1,000	16%	160
	Percentage of all stops: 45.61%			Percentage of all stops: 56.14%		

The police in hypothetical Jurisdiction Q are completely devoid of racial/ethnic bias, and they legitimately stop, as a result of the young drivers' poorer quality driving, twice as many drivers between the ages of 15 and 24 as drivers 25 years of age and older. Twenty percent of the young Caucasians were stopped ($0.2 \times 300 = 60$), and 20 percent of the young minorities were stopped ($0.2 \times 600 = 120$). Police stopped 10 percent of the Caucasian drivers age 25 or above (producing 70 stops) and 10 percent of the minority drivers age 25 or above (producing 40 stops).

The effect of the differential representation of young people among the minority drivers can be seen when we look at the overall representation of Caucasians and minorities among the drivers stopped by police (Figure 2.3). Caucasians made up 50 percent of the drivers (1,000 of the total 2,000) and only 46 percent of the stops. Minorities made up the other 50 percent of the drivers but 56 percent of the stops. Even though racial bias is not manifested by the police (equivalent stopping decisions across racial/ethnic groups), our data indicate (falsely) that disparity exists. If the researcher for Jurisdiction Q did not, as we did, analyze the data within age groups to confirm a lack of disparity, the researcher would have mistakenly concluded

that there was disparity across racial groups. The disproportionate representation of youth in the minority population and the increased likelihood of young people being stopped by police produced the misleading results shown in Figure 2.3: minorities appeared to be over-represented among people stopped relative to minorities' representation in the driving population.

In sum, the strongest research methodologies will address the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*. Theoretically, driving behavior is quite relevant to decisions by police to stop drivers, and the research that has been conducted on the relationship between driving quality and race/ethnicity is not sufficient for us to assume no differences across groups. Complicating matters as pertains to this "quality of driving" factor is the link between age and driving behavior.

Hypothesis 4: Racial/ethnic groups are not equally represented as drivers on jurisdiction roads where stopping activity by police is high.

Vehicle stops by police may occur more in some areas of a jurisdiction than in others. Indeed, the level of stops may vary quite

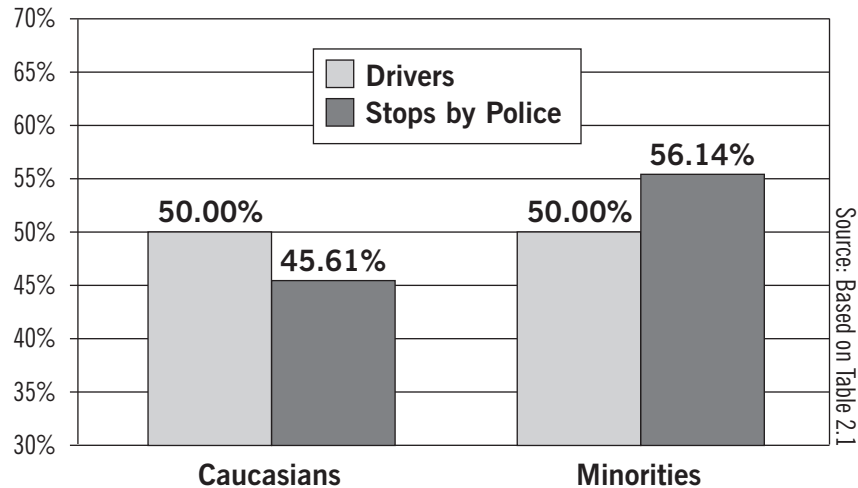


Figure 2.3. False Indication of Racial/Ethnic Bias Based on Age Differences of Drivers in Hypothetical Jurisdiction Q

legitimately from area to area.¹² For example, citizens may be concerned about the high number of accidents in a particular area, and they may ask the local police department to crack down on speeders there. People who drive in areas where stopping activity by police is high are at greater risk of being stopped than are people who drive in areas with low stopping activity.¹³ If the demographic composition of these areas also varies, the police-citizen contact data gathered by the jurisdiction can appear to indicate racial bias by police where none exists. Therefore, we strongly recommend that researchers conduct analyses within subareas of the jurisdiction. If they do not—if they analyze data on stops for the jurisdiction as a whole—results that indicate disparity may reflect not racial/

ethnic bias but very legitimate variations in police practices.

Table 2.2 and Figure 2.4, analogous to the earlier example that focused on differences in age demographics across racial/ethnic groups, illustrate how misleading indicators of racial/ethnic disparity can easily emerge. The racial/ethnic profile of driving-age residents and the racial/ethnic profile of the drivers stopped in hypothetical Jurisdiction R (composed of Area A and Area B) are shown in Table 2.2. There are an equal number of people of driving age in each area (1,000 each), but 80 percent (800) of driving-age residents in Area A are Caucasians, and 80 percent (800) of driving-age residents in Area B are minorities. In each area, the demographic profile of the drivers stopped by police matches the demographic profile of the

¹² These variations in stops across areas within a jurisdiction would *not be* legitimate if the differential enforcement were based on inappropriate factors such as racial/ethnic bias. To discern whether bias is a factor, the researcher could assess whether legitimate factors (such as calls for service, traffic accidents) adequately predict levels of stops.

¹³ Heavy levels of police deployment will not necessarily coincide with high levels of vehicle stops for traffic violations. In fact, in some high-crime areas where police deployment is likely to be correspondingly high, traffic enforcement may be a low priority in light of the more critical problems that need to be addressed.

Table 2.2. Representation of Caucasian and Minority Drivers in the Driving Population and Population of Stopped Drivers, by Subarea, Hypothetical Jurisdiction R

Types of Drivers	Area A				Area B				Total Jurisdiction			
	No. of Driving-Age Residents	Percent of Residents	No. of Stops	Percent of Stops	No. of Driving-Age Residents	Percent of Residents	No. of Stops	Percent of Stops	No. of Driving-Age Residents	Percent of Residents	No. of Stops	Percent of Stops
Caucasians	800	80%	80	80%	200	20%	40	20%	1,000	50%	120	40%
Minorities	200	20%	20	20%	800	80%	160	80%	1,000	50%	180	60%
Total	1,000	100%	100	100%	1,000	100%	200	100%	2,000	100%	300	100%

driving-age adults in the area. That is, in Area A, 80 percent of the residents are Caucasians, and 20 percent are minorities; similarly, 80 percent of the drivers stopped by police are Caucasians and 20 percent are minorities.¹⁴

In Area B, like Area A, the demographic profile of the drivers stopped by police matches the demographic profile of the residents. In short, the results as analyzed within Area A and Area B indicate no disparity. Note, however, that more traffic stops are made in Area B than in Area A. In our hypothetical example, the reason is legitimate. Concerned citizens have requested greater enforcement of speeding laws in this part of town. Because of this, twice as many stops are made in Area B (200 stops) than in Area A (100 stops). If the researcher had not controlled for police activity within the two areas (by conducting separate analyses for Area A and Area B) but instead had presented data for the whole jurisdiction, the results would have appeared to signal racially biased stopping decisions. In the lower, far right column of Table 2.2, note the disproportionate

representation of minorities among drivers stopped by police relative to their representation among residents (60 percent versus 50 percent, respectively). Those misleading results are obtained when the absolute numbers of stops across areas are summed, and the demographic profile of the drivers who are stopped is compared to the demographic profile of the residential population. Minorities comprise 50 percent of the jurisdiction population but 60 percent of all stops. This is shown in Figure 2.4. Anyone looking at this figure could easily jump to the conclusion that police were picking on minorities, stopping them in numbers that were disproportionate to minorities' representation in the population of the jurisdiction. But was this the case? No. Even if all officers' decisions were based on legitimate factors, and even if the increased traffic enforcement activity in Area B was completely legitimate, the results shown in Figure 2.4 would be the same. Such results will mislead researchers unless they take into account Alternative Hypothesis 4.

In sum, people who drive in areas where stopping activity by police is high are at

¹⁴ We use this particular benchmark, residential population, for purposes of making our point—not to promote it as a method.

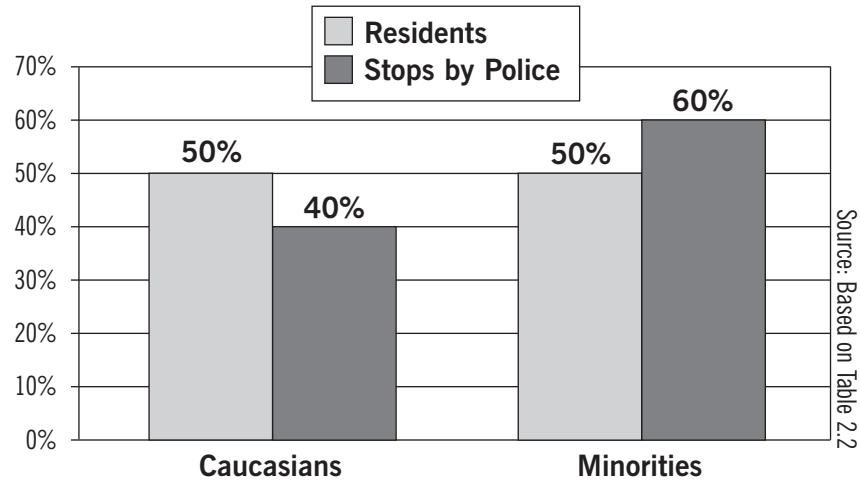


Figure 2.4. False Indication of Racial/Ethnic Bias Based on Differential Stopping Activity by Police across Subareas in Hypothetical Jurisdiction R

greater risk of being stopped than those who drive in areas where stopping activity is low. Stops may legitimately vary across geographic areas where the demographic composition also varies. Therefore, analyses of police-citizen contact data should reflect consideration of the hypothesis that *racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high*. The example given in this section highlights why researchers should conduct analyses within geographic subareas of their jurisdiction, and they should select those subareas in a way that allows them to hold constant (or “control for”) the exposure of drivers to stopping activity by police.

SUMMARY OF THE BENCHMARKING CHALLENGE

To measure whether racially biased policing is occurring in a jurisdiction, researchers must develop a “benchmark” against which to compare the racial/ethnic profile of drivers stopped by police. This benchmark is the racial/ethnic composition of the drivers who are at risk of being stopped, assuming no bias by police. The key factors that influence this

risk are driving quantity, driving quality, and the location of driving. Stopping activity by police is affected by these three causal factors. In order to determine whether there is a cause-and-effect relationship between the race/ethnicity of drivers and police stops (the bias hypothesis), researchers must be able to rule out alternative hypotheses that reflect the factors that increase the risk of being stopped. The alternative hypotheses are

- racial/ethnic groups are not equally represented as residents in the jurisdiction,
- racial/ethnic groups are not equally represented as drivers on jurisdiction roads,
- racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior, and
- racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.

It is not difficult to measure whether police stop drivers of one racial/ethnic group more or less than another; the difficulty

comes in identifying the causes for disparity. The alternative hypotheses present potential causes that need to be ruled out before a researcher can claim that any identified disparity is likely the result of police bias. After controlling for driving quantity, driving quality, and driving location (as pertains to levels of police stopping activity), a researcher who finds that minorities are disproportionately represented among drivers stopped by police can conclude with reasonable confidence that the disparity reflects police bias in their decision making. 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, s/he can report that disparity exists and that the results can be explained either by police bias or differential driving quality. That is, the researcher could not pinpoint a single cause (for example, bias) but must report that (at least) two possible explanations for the disparity remain.

Even results showing no disparity would need to be qualified if all factors were not controlled for. If, for instance, results indicated no disparity in stops, but driving quality had not been considered, the researcher cannot rule out the possibility of racial/ethnic bias in stopping behavior. We explore this possibility further in our discussion below of “masking.”

A benchmark’s value depends on the extent to which it addresses the alternative hypotheses. The higher the quality of the benchmark, the more confidence a researcher can have in the results. The need to rule out alternative hypotheses shows how much more complex benchmarking is than many have previously thought.

THE PROBLEM OF INCONCLUSIVE RESULTS: A CENSUS BENCHMARKING EXAMPLE

Researchers’ failure to address the alternative hypotheses can lead to inconclusive results. In this section we use the census benchmarking method of analyzing police-citizen contact data to illustrate this point. In census benchmarking, a jurisdiction compares the demographic profile of the drivers stopped by police to the demographic profile of the residents of the jurisdiction as measured by the U.S. Decennial Census. Regardless of the results of this comparison (minorities are over-represented, minorities are under-represented, minorities are proportionately represented), *researchers can draw no definitive conclusions regarding racially biased policing.*

Let us suppose that a law enforcement agency finds that minorities are over-represented among drivers stopped by police relative to minorities’ representation among jurisdiction residents. The racial/ethnic disparities manifested in this comparison might reflect racially biased policing, or they might reflect variation in the demographic profiles of (1) drivers on jurisdiction roads, (2) traffic law violators, or (3) drivers driving in locations where stopping activity by police is high. A comparison of stop data to census data indicates disparity, but the *causes* of that disparity have not been identified. We know that we have “disparate impact” (using the social science rather than the legal definition of the phrase), but we do not know if we have unjustified disparate impact in the form of racially biased policing. Because of these limitations, no conclusions can be drawn with regard to the existence or absence of racially biased policing in the jurisdiction.

Census benchmarking (assuming no adjustments of the census data)¹⁵ takes into

¹⁵ Chapter 5 discusses ways that census data can be adjusted by researchers in an attempt to encompass factors related to driving quantity.

consideration only one of the four alternative hypotheses presented in this chapter—the hypothesis that *racial/ethnic groups are not equally represented as residents in the jurisdiction*. Census benchmarking does not address hypotheses related to demographic variations across driving quantity, quality, or location. Unaware of these shortcomings of the methodology, public officials, law enforcement executives, civil rights group representatives, journalists, and other stakeholders often draw wrong conclusions from the results of census benchmarking. Some of those false conclusions are expressed in the benchmarking “myths” described below.

BENCHMARKING MYTHS

Myth 1: No racial/ethnic disparity means no racially biased policing.

The results produced by benchmarking with unadjusted census data, regardless of whether they show that minorities are under-represented, over-represented, or proportionately represented among drivers stopped by police, cannot enable researchers to draw conclusions about racially biased policing. This is an important truth, but some have contradicted it. In some reports, the authors correctly acknowledge that census benchmarking cannot produce conclusions regarding *the existence* of racially biased policing (because the alternative hypotheses have not been ruled out), they argue, however, that it can prove *the absence* of racially biased policing. A finding of disproportionately high minority representation among persons stopped does not prove racially biased policing, they say, but a finding of disproportionately low minority representation or proportionate minority representation does prove that racially biased policing does

not exist. This argument—that a method is valid for one result although not for another—is not true.

The adequacy of a law enforcement agency’s benchmark is the same for all results. The researchers who put forth the argument that, regardless of benchmark quality, a showing of no disparity means no racially biased policing fail to recognize that an inadequate benchmark can “mask” (or hide) disparity. The following example shows how.

Let us say that a jurisdiction uses census benchmarking and finds that the racial/ethnic profile of residents matches perfectly the racial/ethnic profile of people stopped by police. It is still possible that policing in the jurisdiction is racially biased. This is surprising but true. How can it be possible? It is possible if minorities’ representation in the residential population is a higher percentage than minorities’ representation in the population of drivers or a higher percentage than minorities’ representation in the population of drivers violating the law. Then a finding that minorities are stopped proportionate to their residential representation may indicate racially biased policing. Indeed, the existence of racially biased policing may be masked by flaws inherent in the benchmark. For instance, a researcher conducting census benchmarking would not have the information needed on violating behavior and therefore could easily misinterpret the results.

Figure 2.5 presents a finding of no disparity between minorities and nonminorities that some mistakenly argue indicates an absence of racially biased policing. It shows that 25 percent of the residents are racial/ethnic minorities (left bar of three showing minority information) as are 25 percent of the people who are stopped by police for traffic

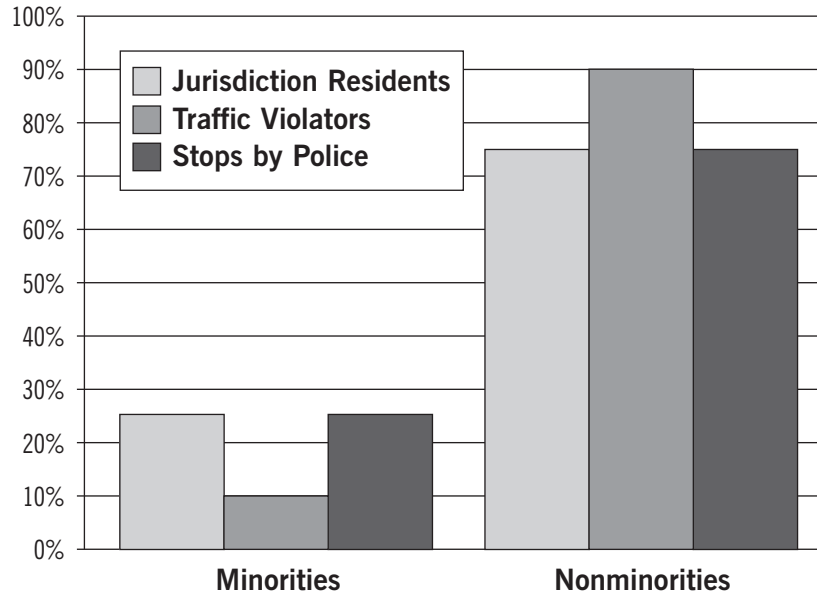


Figure 2.5. Racially Biased Policing Masked in Hypothetical Jurisdiction S

violations (right bar of three showing minority information).

The proportion of minorities and nonminorities who are *traffic violators* (center bar of three showing minority information) is information that would not be available to the researcher who conducted only census benchmarking. This information indicates that minorities are over-represented among the drivers who are stopped (the right bar is higher than the center bar). If minorities comprise only 10 percent of the traffic violators (that is, 10 percent of the population *legitimately at risk of being stopped by police*), but 25 percent of the population that *is stopped by police*, racial bias may be a factor. The key here is that the researcher conducting census benchmarking would not have had the information (on violating behavior that is shown with the center bar) necessary to interpret either results that showed disparity or results that showed no disparity.

Researchers who are assessing police-citizen contact data should remember that (1) a weak benchmark is weak for all results,

and (2) their benchmarking method can mask racially biased policing.

Myth 2: Results from a weak methodology become more worthy over time.

It is not true that results from a weak methodology, or benchmark, can become a worthy baseline for interpreting data in subsequent years—at least not for the purpose of assessing the existence of racially biased policing. An example will help dispel this myth. Let’s say that a jurisdiction uses census benchmarking and determines that racial/ethnic minorities are over-represented among people stopped by police relative to their representation in the residential population as measured by the census. As explained above, these results indicate the existence of a disparity but not its cause. The temptation for stakeholders, and even some researchers, is to equate the disparity with racially biased policing and to desire a reduction in that disparity in subsequent years. That is, they might acknowledge that their benchmark is

weak but falsely claim that the results produced during the first year of analysis can be used to assess and evaluate change in subsequent years. Because of the weak method used, the researcher cannot equate the disparity with racially biased policing and therefore should not presume that a reduction in disparity the following year would be desirable and that it would indicate reduced bias. Similarly, a jurisdiction that finds no disparity as a result of its census benchmarking analysis the first year and *does* find disparity the second year should not blame the police department. Again, because of the methods used, this disparity cannot be equated with police bias.

In sum, a benchmark that cannot pinpoint cause cannot produce explanations of cause over time. A reduction in disparity is not always a legitimate goal. Disparity may reflect wholly legitimate factors at work, but this cannot be known with some benchmarking methods.

Myth 3: Results from a weak methodology become strong if replicated in multiple geographic areas.

A police department that conducts census benchmarking within multiple subareas of the city (say, within each police district) and finds no evidence of racial/ethnic disparity in each one might conclude that the city as a whole is not encountering biased policing. The police spokesperson might acknowledge the weaknesses of census benchmarking but discount those weaknesses and claim that because the results are consistent throughout the city, this proves policing in the city is not racially biased. Such a claim would be in error. The results from a weak methodology are not validated if the results are consistent across multiple geographic areas.

If a methodology can measure only disparity and not the cause of that disparity, this

limitation persists even when the methodology is used over and over again in multiple areas. In a contrasting example, a researcher may find disparity in all or most of the subareas within a jurisdiction. Again, however, multiple measures of disparity do not accumulate to provide a cause for that disparity; they continue to represent only multiple measures of disparity.

CONCLUSION

The challenge of analyzing stop data is to determine a cause-and-effect relationship between drivers' race/ethnicity and stopping decisions by police. This requires that researchers develop a racial/ethnic profile of drivers at risk of being stopped by police, assuming no bias. Several factors can legitimately increase or decrease the likelihood that drivers will be stopped. The "alternative hypotheses" to the "bias hypothesis" take into account these factors. The strength of a benchmark depends on the degree to which it encompasses the factors (driving quantity, quality, and location) associated with the alternative hypotheses. In Chapter 5 we discuss the major benchmarking methods: adjusted census benchmarking, benchmarking based on a comparison of licensed drivers and drivers stopped by police, benchmarking based on blind versus not-blind enforcement mechanisms, internal benchmarking, and observation-based benchmarking. Each benchmarking method's ability to address the alternative hypotheses is explained. Relatedly, we make recommendations regarding how the results of the police-citizen contact data analysis can be responsibly conveyed to the public. However, before we turn to these various benchmarking methods, we discuss how agencies mandated or choosing to collect data initiate collection (Chapter 3) and prepare the data for analysis (Chapter 4).

III Getting Started

This chapter describes the preliminary steps a jurisdiction should take when collecting police-citizen contact data. It also explains how and why a jurisdiction might involve residents, police personnel from all levels of the department, and independent social scientists in these efforts. The factors these jurisdiction stakeholders should consider before choosing a benchmark for analyzing the data are specified.

Any jurisdiction team that is planning to collect data needs to address the following questions:

- On what law enforcement activities should data be collected?
- What information should be collected regarding those activities?
- How should the data be analyzed and interpreted?

Building upon the work of Ramirez, McDevitt, and Farrell (2000), Fridell et al.

(2001, Chap. 8) discuss the options available to agencies regarding the first two questions.¹ For instance, the 2001 PERF book reviews the considerations for deciding whether to collect data on traffic stops only, all vehicle stops, or all detentions (including pedestrian stops). Also discussed are the data elements that agencies should consider for inclusion in their protocol (for example, the date, time, and reason for the vehicle stop; the race, ethnicity, age, and gender of the person stopped; information regarding stop dispositions and search activity). We do not repeat those discussions here. Agencies in the first stages of planning data collection will find these previously published sources helpful. (Again, the Fridell 2001 document can be downloaded from www.policeforum.org.) It also may be constructive for them to contact peer agencies and request to review their “forms.”² Be sure to ask relevant personnel what, in hindsight, they would change about their forms.

¹ For the sake of simplicity, we refer to law enforcement agencies as the primary actor in setting up a data collection system. As discussed in Chapter 1 and later in this chapter, we recommend that the agency work with resident stakeholders in making these decisions.

² Not all agencies are using paper forms to collect their data. Some agencies ask their officers to submit data by using handheld or in-car computers; in other agencies, officers verbally submit the stop information over the radio. The word “forms” used throughout this report denotes all methods of data submission.

**DEVELOPING THE
DATA COLLECTION PROTOCOL:
TWO RECOMMENDATIONS**

We offer two important recommendations related to developing the data collection protocol. First, plans for how an agency will analyze its data should be developed, if feasible, at the same time the decision makers develop the overall strategy for collection. Uninformed or after-the-fact decisions in these matters can lead to unnecessary tensions between residents (particularly racial/ethnic minority residents) and policy makers and/or between police officers and policy makers. Both jurisdiction residents and officers have a strong stake in the highest quality analyses of the data. Officers, in particular, can be legitimately skeptical of—even strongly opposed to—data collection efforts if they lack assurances that the data will be analyzed using the best social science methods available or, at least, responsibly interpreted. An early designation of the method of analysis and a commitment to responsible interpretation can mitigate these concerns. In the same vein, it is important for the agency and other jurisdiction stakeholders to confirm early on that sufficient resources are available to meet their objectives. Otherwise, a jurisdiction may make a significant investment in a data collection system only to find out that analyses of the quality it desires cannot be implemented. Some of the methods that can be chosen to analyze police-citizen contact data rely on particular data elements in the forms that officers complete. This is another reason for comprehensive, early planning.

Second, we strongly advise that, in identifying which activities a jurisdiction will target for data collection, the decision makers select all traffic stops, all vehicle stops, and/or all detentions and not a subset of any of these categories as defined by their outcomes.³

Some agencies (indeed, some states) are collecting and analyzing data only from the traffic stops that *result in citations*. (That is, instead of collecting and analyzing data from all traffic stops, these jurisdictions are focusing on a subset of traffic stops as defined by the outcome, a citation.) This common practice is convenient because it does not add paperwork for the officers. They can use existing, albeit possibly modified, forms. But the practice is not recommended. The resulting data exclude stops by police that may be at heightened risk of being racially motivated. A data collection system based on citation stops alone excludes stops of law-violating drivers who should have received a citation but did not, and it may include law-abiding drivers who should not have been stopped in the first place. These drivers—the fortunate drivers and the illegitimately stopped drivers—could have been “selected” by police based on the drivers’ race/ethnicity. By excluding drivers who do not receive citations, a jurisdiction severely jeopardizes its ability to assess the existence of racially biased policing, regardless of the strength of the benchmark used. The researcher could, with these limited data, identify bias where none exists or conclude there is no bias when, in fact, there is.

This faulty methodology (the limitation of data to traffic stops that result in citations) is

³ We use the term “vehicle stop” to denote any stop made by police of a person in a vehicle; we use “traffic stop” to denote a vehicle stop the stated purpose of which is to respond to a violation of traffic laws (including codes related to quality/maintenance of

vehicles). We use “investigative (vehicle) stop” to denote police stops of people in vehicles when there is at least reasonable suspicion of criminal activity. The term “detentions” includes both vehicle and pedestrian stops.

analogous to assessing the impact of race on prison sentences by focusing only on those who are in prison. For example, by examining only the racial makeup of the prison population and comparing length of prison sentences across races, a jurisdiction will be unable to reach sound conclusions. It must also assess whether or not there are racial differences with regard to who gets sentenced to prison (versus sentenced to jail or to probation, for example).

If a jurisdiction is collecting data only on subsets of stops, the report released to the public needs to include a strongly stated caveat regarding the stops that are excluded from its research. This limitation on the data concerning who is stopped will also affect the analysis of poststop activities and outcomes. This is because some people who were stopped by police—some of whom were searched and maybe even detained for long periods of time—will not be included in the data set being analyzed.

INVOLVING RESIDENTS AND POLICE PERSONNEL IN PLANNING DATA COLLECTION AND ANALYSIS

It is advantageous for jurisdictions to involve residents and a cross-section of law enforcement agency employees in planning how the data will be collected and analyzed. (Regarding the latter, we note that even if a jurisdiction did not involve residents and police in planning the data collection system, it could still involve them in discussions about the data's analysis and interpretation.)

Police personnel—particularly line personnel and representatives of labor associations—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 of their concerns 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 whole community will view the outcome as legitimate. Involving jurisdiction residents in discussions regarding data analysis/interpretation has the additional advantage of educating a core group within the community about the complexities and constraints of the process. These residents can serve as important voices affirming the integrity of the analysis and the sound interpretation of the results when reports are released to the public.

In the interest of responsible social science, the caveats associated with various benchmarking methods should be included in jurisdiction reports. The caveats should convey why the results may not provide definitive proof of racially biased policing or its absence in the jurisdiction. Coming only from the police department spokesperson, these caveats may be interpreted by skeptical residents as defensive excuses for why results showing disparity (if they do) are not proof of racial bias. Although the use of independent social scientists to conduct analyses will add credibility to these caveats, the additional voices of respected residents who understand the methodological constraints will increase the likelihood that the results and the conclusions drawn from them will be viewed as legitimate by the general public and the media. “If the community understands benchmarks and the variables that skew aggregate data there is less likelihood the information will be misinterpreted and misused,” writes McMahan et al. (2002, 94). One way to make sure residents understand data analyses is to set up a local task force on racial profiling or an advisory committee.

As recommended in PERF's first report on the topic of racially biased policing, these task forces should be composed of fifteen to twenty-five people with representatives from both the department and the community (Fridell et al. 2001, Chap. 7). In selecting community members, decision makers should focus on those people who are most concerned about racial bias by police. The task force should include representatives from the jurisdiction's various minority groups and representatives from civil rights groups. Consideration should be given to media representatives as well because these professionals will be in the important position of conveying the results to jurisdiction residents. Police personnel selected for the task force should represent all departmental levels, particularly patrol.

Citizens and police should be involved because they can contribute information of unique value in planning the data analyses and interpreting the results. What they know about the jurisdiction's characteristics, residents, and police activities can be of great help to the researchers charged with actually implementing the analysis plan. For instance, their knowledge of jurisdiction roads may be helpful to a researcher trying to choose representative intersections where observers will document the race/ethnicity of drivers. (See discussion of the observation method of benchmarking in Chapter 5.) Or their knowledge that a particular downtown entertainment area with a large number of minority residents draws many white suburbanites on Saturday nights because of its entertainment venues can help a researcher interpret the results for that area.

PARTNERING WITH SOCIAL SCIENTISTS

If resources allow, a jurisdiction should consider obtaining the assistance of independent social scientists 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 law enforcement agency (and independent of other stakeholders) can add credibility to the process and thus to the results.
- The skills of trained social scientists can supplement the research skills/resources of law enforcement agencies and/or stakeholder groups.

Data collection to assess racially biased policing is a social science research endeavor and a political endeavor. Thus, a law enforcement agency that chooses to collect vehicle stop data, or is mandated to do so, must attend to both social science and political objectives in developing and implementing an analysis plan. An agency could use internal staff to conduct a high-quality analysis but lose in the political arena because the jurisdiction's residents did not consider the internally conducted analysis to be credible.

Many law enforcement agencies (especially small and medium-size ones) do not have the in-house expertise to analyze and interpret police-citizen contact data. A social science partner may be essential to supplement agency resources and perform these functions. The analyst(s) should be trained in social science methods and understand (that is, should have demonstrated knowledge of) the specific issues associated with analyzing police-citizen contact data (Fridell et al. 2001, Chap. 8). Ideally, this "demonstrated knowledge" would come from having conducted similar analyses for other jurisdictions. Capable analysts are most likely to be associated with a college or university or with an independent research firm. The individual social scientist or the research team will play a major role in educating jurisdiction residents about the benchmarking methods that can be used for analysis and the strengths and weaknesses of each.

Importantly, the social scientists become “partners” with the agency or, preferably, *with the jurisdiction task force* in the data collection/analysis effort. They are not just handed the data to analyze as they see fit in the privacy of their university or agency offices. The analysis plan should be agreed upon by all parties, and the social scientists should communicate with their agency and/or task force partners throughout their work. The researchers should share preliminary results, soliciting perspectives from their police and resident partners who will have superior knowledge regarding local conditions that may be pertinent to the interpretation of the data.

SELECTING BENCHMARKS

Chapter 5 describes the various benchmarks that law enforcement agencies and their partners can use to analyze and interpret vehicle stop data. These benchmarks vary considerably in terms of their ability to address the alternative hypotheses discussed in Chapter 2. In deciding which benchmark(s) to use, decision makers should consider the following factors: the level of measurement precision they desire, the financial and personnel resources that are available, the data elements that must be collected, and the availability of other data that may be required for using a particular benchmark.

Level of Measurement

Precision Desired

The higher the quality of the benchmark, the greater the ability of the researcher to “measure” racially biased policing and draw conclusions from the data. High-quality analysis can provide meaningful information not only on whether a problem is indicated, but also on the nature of the problem and the specifics of its manifestation (in terms of particular geographic areas, shifts, or officers). Imperfect data, however, can still provide a solid base for constructive dialogue between

police and other stakeholders. Results showing “disparity” that cannot be linked to a particular “cause” (such as bias) can still lead to a meaningful discussion of possible causes and desirable reforms. Importantly, these discussions can lead to the collection of other forms of “data,” including that which comes from an open and frank sharing of concerns by citizens.

The institution conducting the analysis need not pick one of the most precise methodologies (coming as these do with generally higher complications and sometimes higher costs) in order to make its data collection system successful and constructive. The keys to success for a jurisdiction picking a benchmark are (1) responsible interpretation and (2) constructive discussion among stakeholders concerning the weaknesses of the benchmark that is selected.

For each benchmark described in Chapter 5, we provide information related to its strengths and weaknesses. Because the extent to which each benchmark addresses the alternative hypotheses will determine the legitimacy of conclusions about police bias, reports must include information on the alternative hypotheses to ensure responsible interpretation of the data. This means that the jurisdiction’s report must include information on whether and how the benchmarking method takes into account driving quantity, quality, and location—factors other than bias that can explain stopping decisions by police.

In short, what can be known about the possibility of police bias in a jurisdiction depends on the benchmarking method that is chosen. Because of the difficulty of addressing all of the alternatives to the bias hypothesis, jurisdictions may not be able to pinpoint cause of disparity. Nevertheless, the analysis of police-citizen contact data can yield very positive fruit. Commenting on the value of police-citizen contact data for facilitating police-citizen dialogue, Farrell, McDevitt, and

Buerger (2002, 365) report: “The most effective and productive use of racial profiling data is not its ability to determine if racial profiling exists but rather its ability to provide concrete information to ground police-community discussions about patterns of stops, searches, and arrests throughout local communities.” This important dialogue is a major focus of Chapter 8, “Using the Results for Reform.”

Required Agency Resources

In selecting a benchmark for analyzing police-citizen contact data, an agency or jurisdiction should consider not only the level of measurement precision it desires but also the resources it has available. Not surprisingly, the most effective benchmarks usually (but not necessarily) require the most resources in terms of finances and personnel. A jurisdiction will want to select the most effective method given its resources and objectives.⁴

An important responsibility of stakeholders is to ensure that the police department or other entity responsible for the analyses is provided with sufficient resources. A number of concerned stakeholders across the nation (most notably local and state legislators) have mandated that law enforcement agencies collect and analyze vehicle stop data, but they have not provided the resources to ensure quality processes and products. Without appropriate resources, jurisdictions cannot conduct high-quality analyses. Jurisdiction stakeholders who push for data collection should also push for resources to fund the analyses of those data.

Data Elements

The use of some benchmarks is dependent on the inclusion of particular elements on the data collection form completed by police officers. If the jurisdiction is in the early stages of developing the data collection protocol, decisions regarding how to analyze/interpret the data should be made in conjunction with decisions about the content of the form (that is, what data elements to include). If a jurisdiction has already developed the form, decision makers will need to ensure that the data requested on the form match the data that are needed to conduct the benchmarking method selected. For example, as noted in Chapter 5, some jurisdictions have compared the demographic profiles of drivers stopped for speeding by police unaided by radar to the demographic profiles of drivers stopped because of radar measurements of their speed. (The radar stops are conducted in a manner so that the radar operator cannot discern the driver's race/ethnicity.) To compare these two sets of profiles, the jurisdiction must be able to identify, from data on the forms, which stops were conducted with and without radar. This is only one example of the necessity of advance planning.

For all benchmarking methods we advocate analyses of jurisdiction-level data within specific geographic subareas.⁵ Therefore, the location of the stop is an important data element to include on the police-citizen contact data form. For purposes of reviewing and monitoring data for quality, a unique identifier (number) on the form also is helpful. Most advantageous is an incident number or similar identifier that corresponds

⁴ We do not have reliable information regarding the costs that are associated with the various benchmarking methods. Many jurisdictions seeking to hire outside analysts issue requests for proposals and then review the proposals. In the review they balance the strength of the methodology against the resources required to conduct the analysis.

⁵ Analyzing data within subareas of jurisdictions (for instance, counties, municipalities) is unwieldy for the researcher who is charged with analyzing state-wide data.

to information about the event that is contained in other data sets, such as computer-aided-dispatch (CAD) data and citation data.

The Availability of Other Data

Some benchmarking methods are dependent upon the availability of information from outside sources. Jurisdictions should make sure they can get the data a particular benchmark requires before going full steam ahead with efforts to gather police-citizen contact data.

For example, benchmarking with blind enforcement mechanisms (say, enforcement cameras) is a method that would be available only to jurisdictions that (1) have enforcement cameras in place at controlled intersections to detect and ticket red-light violators or speeders and (2) are in states that have racial/ethnic information for owners of registered vehicles. Clearly, a jurisdiction that chooses this benchmarking method is reliant on data from a source outside the law enforcement agency. With the data provided by the cameras, the jurisdiction can compare the racial/ethnic profile of drivers who are identified as traffic violators by enforcement cameras to the racial/ethnic profile of drivers who are identified as violators by officers on patrol in the same area as the cameras. The camera provides a photo of the license plate number, which enables the jurisdiction to determine the race/ethnicity of the vehicle owner from DMV data. (See Chapter 5 for a more detailed explanation of this benchmarking method.)

Other Considerations

A jurisdiction may decide to use multiple benchmarks. For example, it might implement “internal” benchmarking and some “external” method as well. Internal benchmarking is a strong benchmark for identifying which police officers or units may be stopping minorities at higher rates than their “similarly situated” counterpart officers or units. A drawback to internal benchmarking, however, is that it only compares parts of the

law enforcement agency to itself. For this reason, the jurisdiction might choose—in addition—to compare the agency’s performance to some outside benchmark, such as that provided by the blind versus not-blind enforcement method, or the observation method. Thus, a jurisdiction might implement both internal benchmarking and some external method as well.

A jurisdiction might also decide to implement a relatively simple benchmark (for example, benchmarking with adjusted census data) in all the subareas of its jurisdiction and then invest in a more complicated and more effective benchmark (for example, the observation methodology) in those subareas identified by the simpler benchmark as having the greatest racial/ethnic disparities.

INFORMING THE PUBLIC OF DATA COLLECTION EFFORTS

Some law enforcement executives, when announcing the agency’s data collection efforts, have referred to the initiative as an opportunity to “prove” that policing in their jurisdiction is not racially biased. Such a prediction of research results is *inappropriate*. While a particular executive might be justified in having confidence that racially biased policing is neither systematic nor widespread within his or her jurisdiction, the executive is naïve to claim absolutely that it never occurs. Such a statement is almost certain to offend racial/ethnic minorities who perceive otherwise. Our society has serious racial/ethnic biases, and the police profession—like every other profession—hires from a population with these prejudices. Even in a department in which racial bias is neither systematic nor widespread, it is likely that biased decisions occur in some places, at some times, by some individual officers. Finally, such a strong claim (the police executive’s use of the word “prove”) conveys to the public that police-citizen contact data can provide definitive answers—which they cannot. As is true of

social science in general, even strong methods of benchmarking will not provide definitive proof of the existence or lack of racially biased policing.

Although data collection as a response to racially biased policing has had important benefits, one side effect has been negative: the inherent implication that some agencies are “guilty” of racial bias and others are “innocent.” Equally unfortunate is the related implication that the “guilty” agencies are the only ones that should implement reforms. In fact, all agencies committed to democratic policing, not only agencies supposedly “proven guilty” of bias through data collection, can make progress on this longstanding issue; all agencies can implement action plans to help them move closer to the ideal of bias-free policing by every officer. Because of the inability to provide definitive answers about the existence of racial bias in a jurisdiction and because all agencies can make progress on this issue, data collection and analysis should not be viewed by law enforcement executives and other stakeholders as a pass-fail test but rather as a means of identifying priorities for change.

CONCLUSION

This chapter reviewed important considerations for jurisdictions that are getting started with data collection. It offered concrete suggestions that will help jurisdictions develop a useful form for recording police-citizen contact data. It also provided guidelines for deciding which types of activities to target for data collection. We encourage the involvement of residents and police personnel from all levels in making decisions regarding the data collection system, and we explain the circumstances in which jurisdictions might want to involve independent social scientists.

The selection of a benchmarking method should be based on several factors: the level of measurement precision desired, the financial and personnel resources of the jurisdiction, the data elements specified on the form for recording police-citizen contacts, and the availability of other data. A police executive announcing data collection plans to the public should not vow that the initiative will “prove innocence” before the fact. Like society at large, an agency is rarely bias free. Neither should that agency executive, and those residents and other stakeholders partnering with the agency, await the results of data collection—whatever they might be—to implement reforms. It is never too soon to address racially biased policing or perceptions of its practice.

IV

Data Analysis Guidelines for All Benchmarking Methods

Researchers, regardless of their benchmarking method for evaluating whether policing is racially biased in a jurisdiction, should follow certain guidelines when analyzing data on police-citizen contacts. This chapter summarizes the guidelines discussed in much more depth in *By the Numbers* (Fridell 2004, Chapter 4).

Data that have been collected from officers should be checked for quality by the researcher or under the supervision of the researcher. This is an important first step in any type of social science research and not unique to the analysis of police-citizen contact data, but it is a step that has been overlooked in many jurisdictions. In this chapter we also discuss “reference periods,” the length of time that agencies should collect data before they begin analyzing it. The chapter explains why it is advisable for researchers to analyze portions or “subsets” of the full data set. Subsets based on the type of stop (proactive or reactive), whether the officer could discern the driver’s race/ethnicity before the stop, and the geographic location of the stop are recommended. The final section of the chapter describes necessary adjustments for comparing the stop data and the

benchmarking data in any analysis, or what we call “matching the numerator and the denominator.”

REVIEW OF DATA QUALITY

Quality data are a prerequisite for quality research. For accurate results, social scientists in any research endeavor carefully review their data to check for and, if possible, correct errors before analyzing it. Once data collection is under way, researchers who are attempting to measure racial bias should “audit” the incoming data from officers for quality. Even if the department has been collecting data for a while, researchers are still advised to implement a data review and monitoring system.

Although there is no cost-effective way to ensure that the data are 100 percent accurate, researchers can use various auditing methods to improve the quality of the data. These audits have two objectives: (1) to ascertain whether line personnel in the police department are submitting data collection forms for all stops targeted for data collection (for example, all vehicle stops or all traffic stops) and (2) to ensure that officers are filling out the forms fully and accurately.¹

¹ As noted in Chapter 3, not all agencies use paper forms to collect their data. Some officers submit data

by using handheld or in-car computers, or they relay information on vehicle stops over the radio.

To achieve the first objective, researchers should try to identify a second source of data that tracks some or all of the stops that are targeted for data collection. That second source of data may be “computer aided dispatch” (CAD) data, citation data, written warning data, videotapes, or other departmental data. Researchers then can compare aggregate totals from the two data sets. For example, researchers could check whether there are as many police-citizen contact forms with citation dispositions as there are citation incidents in the citation database. Researchers also can compare the two data sources on a stop-by-stop basis. The latter method is preferable because it has the added advantage of identifying the source of any problems and can lead to interventions to improve data quality.

Agencies will not be able to compare two data sources on a stop-by-stop basis unless they institute at the start of the data collection system a mechanism for linking the stop data to secondary sources. (An example of this would be including a citation number on the vehicle stop form.) Most agencies, however, will have second sources of data to implement the first auditing method described above. For instance, most will be able to compare the number of citations given by police as recorded in the vehicle contact data to another source of citation information.

In addition to checking that forms are being submitted for all of the targeted stops, researchers should review the data to detect missing or potentially erroneous data. This review is particularly important during the first two months of data collection. If this review identifies significant amounts of missing data for particular variables or numerous apparent errors on forms, the agency should implement, early in the data collection process, certain corrective measures (for example, a remedial training program to ensure that officers know how to fill out the

forms and understand the importance of filling out the forms correctly).

It is impossible for researchers to detect all errors merely by reviewing the data that have been submitted. For example, a review of the data is unlikely to detect that the correct disposition of a traffic stop was a warning when the officer erroneously indicates on the form that a citation was given. However, a review is still worthwhile because it will improve the quality of the data, which in turn will improve the quality of the research results.

REFERENCE PERIOD FOR ANALYSIS OF DATA

A key question for many social science studies is “How much data should be collected before analysis begins?” We recommend that analysis be based upon the stops occurring within a full twelve-month period, if feasible. This reference period length will lessen the impact on the data of special events or circumstances, it will eliminate seasonal effects (since all seasons will be included), and it will increase the reliability of the data. A twelve-month reference period, however, may not be economically feasible or politically viable. Regarding the latter, residents may not expect to wait more than one year for the results of the analysis. If researchers choose a reference period of less than one year (for example, six months), their report to the public should include a caveat that the results do not necessarily generalize to the rest of the year for which data were not analyzed.

It is advisable to delay the start of the reference period for the analysis until officers have become accustomed to the data collection process. (That is, the first one or two months of data collection should not be included in the analysis.) As noted above, these first few months of data should be reviewed to identify problems (such as large

amounts of missing data on particular variables), and these problems should be resolved through communications with officers or other retraining. Once the problems appear to be resolved, the reference period should begin.

REASONS FOR ANALYZING SUBSETS OF DATA

For many reasons explained in this section, researchers should analyze subsets of the police-citizen contact data rather than the whole data set. Below we discuss subsets based on (1) whether stops are proactive or reactive, (2) whether the officer could discern the driver's race/ethnicity, (3) geographic location of the stops, and (4) whether the stops are for traffic violations or for the purpose of investigating crime.

Type of Stop: Proactive or Reactive

Researchers analyze police-citizen contact data to try to find out whether or not individual officers are making decisions to stop drivers based on racial/ethnic bias or based only on legitimate factors that should affect the selection of drivers they stop. Researchers can evaluate whether bias was a factor only on stop decisions where the police had a choice (proactive stops). On reactive stops by police (that is, stops in response to a traffic accident or stops in response to an order to stop drivers at a highway checkpoint), police have less discretion (often *no* discretion) in the selection of "who is stopped." Those reactive decisions to stop drivers are unlikely to be influenced by bias.

Law enforcement agencies that have designed their data collection process to target only proactive stops do not need to whittle down their data set. Agencies, however, that are mandated to collect or voluntarily collect data on proactive *and* reactive stops should include only proactive stops in the data given to researchers for analysis—if the agencies are able to separate the two groups of stops based on information provided on the forms.² The ability to create this subset of data is dependent on the inclusion of information on the form regarding the type of stop conducted by the officer.

Prestop Observability of the Driver's Race/Ethnicity

"Was the driver's race/ethnicity observable by police before the stop?" The answer to this question has significant relevance to an assessment of whether or not stopping decisions are based on bias. An officer who cannot discern the racial/ethnic characteristics of a driver cannot make a (biased) decision based on those characteristics. Therefore, it makes sense to exclude those incidents in which the officers could not discern (at the time the decision was made to stop the vehicle) the driver's race/ethnicity.³

The decision, however, to exclude data for stops for which officers said they could not discern the driver's racial/ethnic characteristics can have negative effects, political as well as statistical. In some jurisdictions, residents have questioned the inclusion of this variable (regarding whether the officer can discern driver characteristics) on police-citizen

² Both proactive and reactive stops should be included in the analysis of poststop variables (for example, length of stop, whether a search is conducted). Although officers had little discretion deciding whom to stop in reactive situations, they have considerable discretion in deciding how to proceed once the stop is made.

³ Both "observable characteristics" stops and "not observable characteristics" stops (like proactive and reactive stops) should be included in the analysis of poststop factors.

contact data forms because they doubt the validity of officers' responses. Those who mistrust the data submitted for this variable are likely to be skeptical of a decision to exclude all data for stops with "not observable characteristics." Another potential drawback is that the exclusion of stops for which officers report that they could not discern the race/ethnicity of the driver may reduce the size of the data set dramatically. We recommend that, if politically and numerically feasible, researchers include in their analysis of data regarding "who is stopped" only those stops for which the driver's racial/ethnic characteristics can be discerned. The ability to create this subset requires an element on the form soliciting information from the officer about his or her ability to discern demographic characteristics.

Geographic Location of Stop

If possible, researchers should create subsets of data based on type of stop (proactive stops only) and on whether the officer could observe the driver's race/ethnicity before the stop was made ("observable characteristics" stops only). A third subset of data is related to the geographic location of the stop.

Because it is likely that *racial/ethnic groups are not equally represented as drivers on jurisdiction roads where stopping activity by police is high* (see Chapter 2, Alternative Hypothesis 4), researchers should analyze data for geographic subareas of the jurisdiction.

To illustrate our point we use census benchmarking as an example. We recommend that researchers not compare the racial/ethnic profile of *all* drivers stopped in the jurisdiction to the racial/ethnic profile of *all* driving-age residents in the jurisdiction. Instead it is preferable to compare stop data and benchmark data within smaller geographic

areas of the jurisdiction. These subareas become subsets of the analyses.

Subarea analysis "controls for" the volume of stopping activity by police. Some areas may have many more stops per capita than others because these areas have a high rate of calls for service or a large volume of accidents. As suggested earlier, the race/ethnicity of the resident population in these high-activity and low-activity areas may vary considerably. Because there may be greater vehicle stop activity in Area A than in Area B, researchers should compare the racial/ethnic profile of drivers stopped in Area A to the racial/ethnic profile of drivers in the Area A benchmark group. Similarly, the demographics of drivers stopped within Area B should be compared to the demographics of the Area B benchmark group.

Type of Stop: Traffic or Investigative

The term "traffic stop" refers to a vehicle stop the stated purpose of which is to respond to a violation of traffic laws, including codes related to quality/maintenance of the vehicle; the term "investigative stop"—in a vehicle context—denotes police stops of people in vehicles when there is a reasonable suspicion of criminal activity. We now consider whether agencies collecting data on all vehicle stops should analyze traffic and investigative stops together as a group or separately.⁴

In answering this question, we distinguish between what is theoretically appropriate and what is practical in terms of measurement capabilities. At a theoretical level, traffic stops and investigative stops should be analyzed separately and alternative hypotheses developed for both categories. The factors that put a person at risk of being legitimately stopped by police for a traffic

⁴ This discussion is not relevant to agencies that are collecting data only on traffic stops.

violation are different from the factors that put a person at risk of being legitimately stopped by police for purposes of investigating criminal activity.

The police-citizen contact data that are collected, however, do not enable agencies to distinguish between the stops made for the purpose of enforcing traffic laws and those made for the purpose of investigating crime. Police can and do stop vehicles on the basis of legitimate traffic violations but *for the purpose of* investigating crime. In most agencies these “pretext stops”—as they are called—will be coded as traffic stops, even though at their core they are investigative stops. Because researchers cannot successfully distinguish the two types of stops, we recommend that all vehicle stops—those for traffic and for investigative purposes—be analyzed together. This is a necessary and practical means of resolving the research problem just described. It does, however, reduce the precision of the analyses.

Although pretext stops hinder researchers’ ability to determine which stops that were coded in the data as “traffic stops” were, in fact, “traffic stops” and not “investigative stops,” the converse is not true. Researchers can be confident that the stops coded as investigative stops are truly investigative stops and not a “ruse” for catching a traffic violator. For this reason, after researchers conduct their major analysis using all traffic and investigative stops (see discussion above), they can conduct an additional analysis of only investigative stops using appropriate crime-related benchmarks (see Chapter 5).

MATCHING THE NUMERATOR AND THE DENOMINATOR

Social scientists analyzing police-citizen contact data to measure racially biased policing emphasize the importance of “matching the numerator and the denominator.” In their specialized lingo, the “numerator” refers to

the data collected on stops by the police, and the “denominator” refers to the data collected to produce the benchmark or comparison data. To “match the numerator and the denominator” means the researcher should adjust the stop data to correspond to any limiting parameters of the benchmark or vice versa.

For example, the researcher benchmarking with census data adjusted for vehicle ownership should include in his or her analysis only the stops by police involving drivers who are residents of the jurisdiction. In this method of analysis, the researcher adjusts the census data on the demographics of residents to take into consideration who, among those residents, owns a vehicle. That is, the researcher compares the racial/ethnic profile of the people stopped by police to the racial/ethnic profile of people who not only live in the jurisdiction but who also have access to vehicles, according to the U.S. Census. The “numerator” is the stop data collected by police, and the “denominator” is the adjusted U.S. Census data. The denominator in this situation is restricted: it only includes people who live inside the jurisdiction. This parameter on the denominator must be applied to the numerator data. That is, the researcher must compare the census data only to the stops by police of *jurisdiction* residents. The researcher must select out of the numerator data all of the stops of drivers who do not live inside the jurisdiction. Nonresidents of the jurisdiction are excluded from the denominator, and therefore they must be similarly excluded from the numerator.

Sometimes, however, data must be excluded from the denominator. For example, with adjusted census benchmarking, there is an inherent limitation on the numerator (the stop data); only people of driving age will be included. The drivers stopped by police will usually be of legal driving age. Because only people of driving age will be represented in the numerator, the researcher must limit the

denominator data to people of driving age. Thus, in the example of census benchmarking, the researcher will not calculate the race/ethnicity of all residents of the jurisdiction but only of those residents who are of driving age (for example, age 15 and older).

Regardless of the benchmarking method, researchers must match the numerator data and the denominator data: the parameters that apply to one must be applied to the other. Let us consider how that works in the observation method. Placed at various locations in the jurisdiction—often at intersections, observers count drivers in various race/ethnicity categories. The result is a racial/ethnic profile of drivers around that intersection (the denominator data). Therefore, the numerator data (drivers stopped by police) must be limited to the same intersection. To conduct observation benchmarking, the researcher will compare the demographics of the people who are observed driving through Intersection A 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 under study in the jurisdiction.

“Matching the numerator and the denominator” applies to the time period during which the data are collected as well. In this observation methodology example, if the observation data are collected during January through May 2002, the analysis will involve

only those police stops that occurred during that same (or reasonably similar) 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.

CONCLUSION

The preceding guidelines for data analysis apply to all benchmarking methods. Researchers need to review data quality, choose a reference period, analyze subsets of data, and match the numerator data and the denominator data. Some of these guidelines also apply to the analyses of poststop data (see Chapter 6).

Jurisdictions that follow these guidelines will improve the validity and usefulness of their findings. They also will increase their costs. It is important for stakeholders to be aware that some of the steps recommended above will increase significantly the labor involved in the analysis of the vehicle stop data. For instance, conducting the recommended subarea analyses instead of a single, jurisdiction-wide analysis multiplies several times over the time needed for researchers to complete their work. Stakeholders aspiring to high-quality data and meaningful analyses need to ensure that there are sufficient resources available to finish well what they began.

Methods for Benchmarking Stop Data

A number of benchmarking methods have been created to help jurisdictions analyze and interpret vehicle stop data collected to measure racial bias. This chapter reviews the following methods and assesses their strengths and weaknesses: 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, and several other benchmarking methods and tools. For more detailed information on how to implement each of these methods, see *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* (Fridell 2004), Chapters 5 through 10.

Before explaining the first method, we review several general principles introduced in Chapter 2, “The Benchmarking Challenge.” In any benchmarking method, the researcher compares the racial/ethnic breakdown of drivers stopped by police (the stop data) to the racial/ethnic breakdown of people at risk of being stopped by police, assuming no bias (the benchmark data). The purpose of this comparison is to determine if disparity exists that might indicate racial bias. In creating a benchmark group that represents the people at risk of being stopped by police, the researcher considers important factors

influencing police decisions to stop someone. These factors are related to driving quantity, driving quality, and driving location. Ideally, the researcher’s benchmark would take into account 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. These factors underlie the alternative hypotheses described in Chapter 2.

BENCHMARKING WITH ADJUSTED CENSUS DATA

In census benchmarking, law enforcement agencies compare the demographic profile of drivers stopped by police to the demographic profile of jurisdiction residents as measured by the U.S. Census Bureau. A straight comparison between the demographics of these two groups is called “unadjusted” census benchmarking. The weaknesses of this method in ruling out alternative hypotheses were discussed in Chapter 2. Most jurisdictions appear to be benchmarking their police-citizen contact data against unadjusted census data. This is because their resources are limited, and unadjusted census benchmarking is the simplest and least costly

benchmarking method. That said, we do not recommend unadjusted census benchmarking. Agencies that must rely on census methods should use one of the various adjustment techniques described below.

Researchers should adjust the census data by incorporating into their benchmarking method information pertaining to one or more of the alternative hypotheses (such as quantity of driving). They can do this, for example, by taking into consideration who, among the residents listed in the census data for the jurisdiction, has access to a vehicle. People without access to a vehicle are clearly at less risk of being stopped in vehicles by police than are people with vehicle access. Census benchmarking with this adjustment is a stronger method than unadjusted census benchmarking for assessing the nature and extent of racially biased policing.

Adjusting Census Data on Jurisdiction Residents to Account for Vehicle Access

Researchers who adjust census data to account for vehicle access are improving the quality of their research by taking into account the alternative hypothesis that *racial/ethnic groups are not equally represented as drivers on jurisdiction roads*. To make this beneficial adjustment to the census data, they would subtract from the census population data for each racial/ethnic group in the jurisdiction the estimated number of people within each of those groups who do not have access to vehicles. To do this, the researcher would obtain the census information for the jurisdiction on vehicle-less households by race and ethnicity.

Figure 5.1 compares a hypothetical racial profile of people stopped by police in Area A to the benchmark data produced by adjusted census information. There is little racial disparity indicated. Caucasians represent 65 percent of the drivers stopped by police and 67 percent of jurisdiction residents with access to vehicles. African Americans represent 19 percent of drivers stopped by police

and a corresponding 19 percent of jurisdiction residents with access to vehicles.

Another recommended way that researchers can adjust census data on the jurisdiction is to take into account the drivers on jurisdiction roads who come from neighboring jurisdictions.

Adjusting Census Data on Jurisdiction Residents to Account for the Influx of Nonresident Drivers

The race/ethnicity of drivers on jurisdiction roads may not match the race/ethnicity of jurisdiction residents because of (1) racial/ethnic differences in driving quantity and/or (2) racial/ethnic differences in the population of people who do not live in the jurisdiction but drive in it. By adjusting for vehicle access, researchers can address indirectly the first possibility; by adjusting for the influx of nonresident drivers, researchers can focus on the second possibility.

Not all drivers on jurisdiction roads are residents of that jurisdiction, and the influx of nonresidents can affect the racial/ethnic profile of drivers “available” to be stopped by police.

Consider, for example, a municipal area with a substantial minority population; during the day many Caucasians from the suburbs drive into this jurisdiction for work. As a result, the percentage of drivers on jurisdiction roads who are Caucasian is higher than the percentage that would be predicted to be on jurisdiction roads based on residential population data alone. Of course, nonresidents also might enter a jurisdiction for other reasons—to shop, go to school, to seek entertainment, to travel on to another jurisdiction, or for other reasons.

Because of the influx of nonresidents, the racial/ethnic profile of residents produced by unadjusted census data is likely to be an *inaccurate* estimate of drivers who could be stopped by police, assuming no bias. Therefore, the adjustments described below are needed.

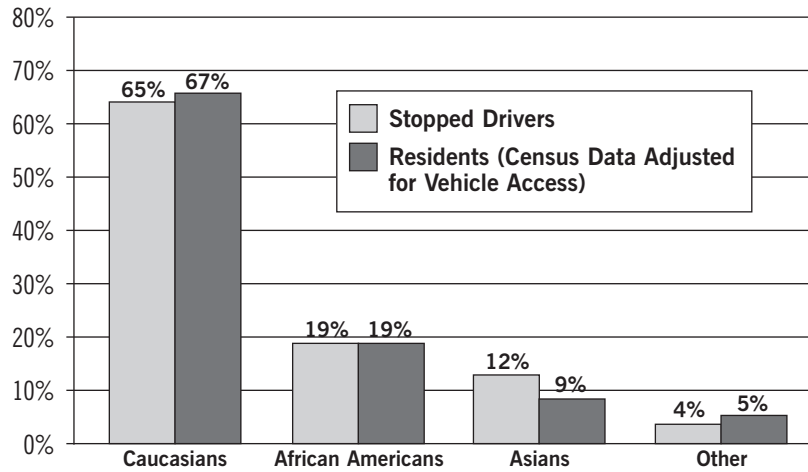


Figure 5.1. Drivers Stopped by Police in Hypothetical Area A and Area A Residents with Vehicles, by Race

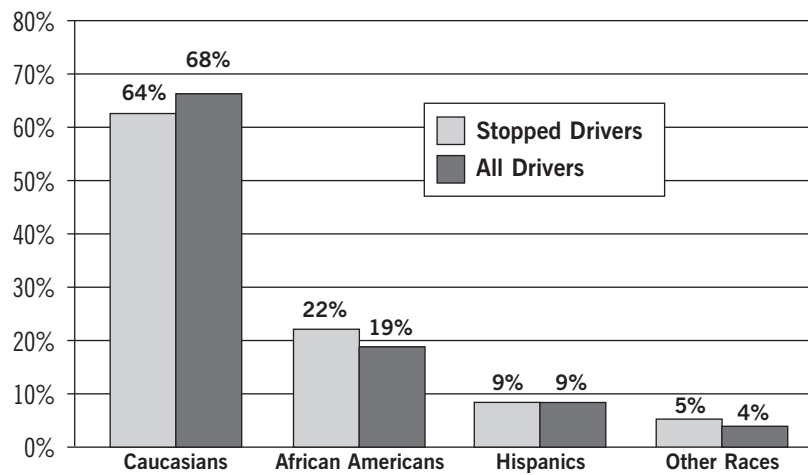


Figure 5.2. Drivers Stopped by Police in the Target City and All Drivers (Resident and Nonresident) in the Target City, by Race

Several researchers have come up with ways to estimate the race/ethnicity of resident and nonresident drivers in a jurisdiction. Here we describe the method developed by Novak (2004). First, Novak looked at the stop data in the jurisdiction (the so-called “target jurisdiction”) to estimate the extent to which nonresidents from various “outside jurisdictions” were represented on target jurisdiction roads. (He was able to do this because

the stop data for the jurisdiction included information on the jurisdiction-of-residence of the driver.) With this information, he produced an estimate of the extent to which residents and nonresidents drove on jurisdiction roads. Second, he used the census data for the target jurisdiction and for each of those outside jurisdictions to estimate the racial/ethnic profiles of these resident and nonresident drivers. Finally, he compared the

racial/ethnic profile of drivers stopped by police in the target jurisdiction to the racial/ethnic profile of all drivers (residents and nonresidents) in the target jurisdiction. Hypothetical results are shown in Figure 5.2 (previous page).

Drawing Conclusions from the Results

Adjusted census benchmarking can incorporate two kinds of valuable information related to driving quantity: information on vehicle access and information on the influx of nonresidents into the jurisdiction. This method addresses, in part, the alternative hypothesis that *racial/ethnic groups are not equally represented as drivers on jurisdiction roads*. Researchers conducting adjusted census benchmarking who are able to analyze subareas of the jurisdiction incorporate useful information related to driving location; subarea analyses enable these researchers to address the alternative hypothesis that *racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high*.

Another alternative hypothesis—*racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*—is not addressed, however, by the method of benchmarking with adjusted census data. Driving quality is an important factor influencing police decisions to stop drivers, but it is not taken into account by this method. For this reason, researchers benchmarking with adjusted census data cannot draw definitive conclusions regarding the causal link between the race/ethnicity of drivers and stopping behavior by police. Because of its importance, this statement bears repeating: researchers (and the law enforcement agencies and other

stakeholder groups citing the researchers' reports) cannot draw conclusions about whether policing in a jurisdiction is racially biased.¹

Despite these weaknesses of adjusted census benchmarking as a diagnostic tool, researchers limited by resources or time may have no option other than to use this method. In particular, researchers who are charged with analyzing data from all of the jurisdictions within a single state or many of them may have to rely on this method, despite its weaknesses. The obligation of the researcher in this position is to ensure that the results are conveyed in a responsible fashion. In fact, this obligation also falls to all stakeholders, including concerned residents, civil rights groups, local/state policymakers, and the media. No one interpreting results based on benchmarking with adjusted census data can legitimately draw conclusions regarding the existence or lack of racially biased policing.

BENCHMARKING WITH DMV DATA

Benchmarking with census data adjusted for vehicle access and benchmarking with data from the Department of Motor Vehicles (DMV) are similar methods. In one the researcher creates a comparison group based on people who live in the jurisdiction and have access to a vehicle. In the other the researcher creates a comparison group based on people who live in the jurisdiction and have a driver's license. Like adjusting census data for vehicle ownership, benchmarking with DMV data produces an indirect measure of driving quantity. It accounts, in part, for the alternative hypothesis that *racial/ethnic*

¹ What they can do is mention the disparities or lack of disparities shown by the data, and they can reference possible explanations for the results—using the alternative hypotheses as a guide.

*groups are not equally represented as drivers on jurisdiction roads.*²

Implementing the Method

To implement this method in its simplest form, researchers restrict their analysis to jurisdiction residents; they compare the racial/ethnic profile of licensed drivers who live in the jurisdiction to the racial/ethnic profile of the jurisdiction residents stopped by police.³ To implement a preferable and more sophisticated version of this method, researchers conduct subarea analyses and/or take into account the influx of nonresidents.

Earlier we described how Novak (2004) adjusted census data by measuring the influx of drivers from outside the jurisdiction. Using stop or citation information and census data, he was able to estimate the racial/ethnic profiles of residents and nonresidents on jurisdiction roads. These same techniques can be applied to the DMV benchmarking method by substituting the driver's license demographic data for the census demographic data.

Drawing Conclusions from the Results

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.

Using DMV data to benchmark police-citizen contact data is similar to using census data that have been adjusted for vehicle ownership. Both benchmarking methods produce a proxy measure for driving quantity by trying to determine who is and who is not driving on jurisdiction roads. The benchmarking method that uses adjusted census data considers a person a driver if the person has access to a vehicle. The method described in this section considers a person a driver if the person has a driver's license. This method will not produce conclusions regarding the existence or lack of racially biased policing in a target 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. We discuss how the results can be used to stimulate these discussions in Chapter 8.

BENCHMARKING WITH DATA FROM “BLIND” ENFORCEMENT MECHANISMS

Law enforcement agencies can use “blind” enforcement mechanisms (for example, red light cameras, radar, air patrols) to create benchmark data against which they can compare their data on stops by patrol officers. In 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

² To use this benchmarking method, a law enforcement agency must be able to obtain from the Department of Motor Vehicles (DMV) information on the race and/or ethnicity of the licensed drivers in the target jurisdiction (the jurisdiction being analyzed based on police-citizen contact forms). Additionally, the information from the DMV regarding race and/or ethnicity must be

compatible with the measurement of race and/or ethnicity on the law enforcement agency's data collection form.

³ Comparing *residents* with drivers' licenses to *residents* stopped by police reflects “matching the numerator and the denominator” as described in Chapter 4.

stops in which officers exercise a high degree of discretion to low-discretion stops. At least one jurisdiction benchmarked with “blind” data from a nontechnological source; they compared “daylight stops” and “darkness stops.” These benchmarking methods also are explained in this section.

Benchmarking with Data from Red Light Cameras

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. These cameras are placed at selected intersections that have a traffic light. A driver who runs the red light trips the camera, which takes a picture of the violator’s license plate.

In this benchmarking method researchers compare the racial/ethnic profile of the drivers “ticketed” by the camera technology to the racial/ethnic profile of the drivers stopped by police.⁴ The goal of benchmarking is to create a comparison group of people at risk of being stopped by police, assuming no bias. People “ticketed” by red light cameras do represent such a group. If officers are as “blind” to race/ethnicity as are the cameras, the demographic profile of the people stopped for red light violations (or comparable 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 humans than by the technology.

Matching the Numerator and the Denominator

The camera data can provide researchers with information on the race/ethnicity of people violating red light laws in certain locations.⁵ The researchers, however, must carefully match the numerator data and the denominator data (see Chapter 4). Specifically, they must match the stop data (numerator) to the red light data (denominator) in terms of location, time, and violations detected.⁶

This careful matching, while necessary, inevitably narrows the scope of the racial bias assessment. An example will help to convey this point. To maximize the match of the two groups, the researcher might use the same intersection (for example, Intersection A) to collect both camera data and officer stop data. Alternatively, in a near-ideal design, the researcher might compare red-light-camera data from Intersection B to officer stop data for red light violations at Intersection A. Intersections A and B would need to be similar in terms of (1) the race/ethnicity of the drivers (this might be more likely if the intersections are near each other) and (2) driving behavior because both intersections have the same type of traffic—residential, not commercial. In this way the researcher has maximized the match between the stop data and the benchmark data. As will be explained later, the laudable rigor of this match comes at a cost: the scope of the analysis is narrowed.

Having maximized the match between the stop data and the benchmark data, researchers can begin to conduct the comparison. They can compare the racial/ethnic profile of the people “ticketed” by the red

⁴ Actually, the person “ticketed” by the camera is the person to whom the vehicle is registered, not necessarily the driver. This is one drawback of this method.

⁵ To benchmark against red-light-camera data, a law enforcement agency must, of course, have red-light-camera technology in place. It also must be able to

access Department of Motor Vehicle (DMV) data that can link the license plate photographed by the camera to the race and/or ethnicity of the owner of the vehicle.

⁶ See Chapter 4 for a detailed explanation of the concepts of “numerator” and “denominator.”

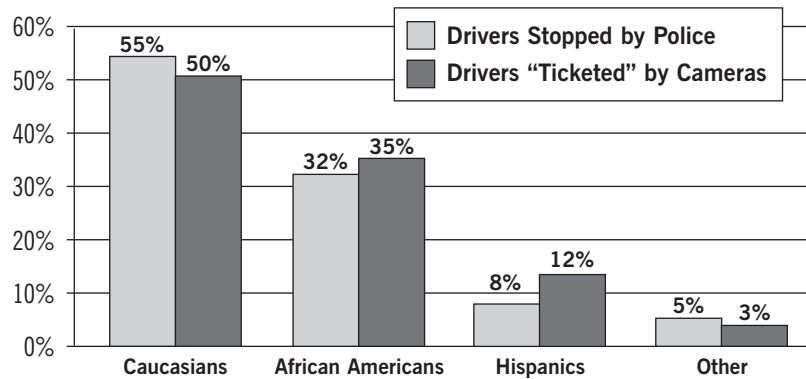


Figure 5.3. Drivers Stopped by Police for Red Light and Stop Sign Violations and Drivers "Ticketed" by Red Light Cameras, by Race/Ethnicity of Drivers (Hypothetical Data)

light cameras to the racial/ethnic profile of the people stopped by police for red light violations in the matched geographic area (see Figure 5.3). They will conduct these analyses for each red-light-camera intersection and its matched geographic area.

Drawing Conclusions from the Results

This benchmarking method has an important strength: it can create a comparison group, or benchmark, that reflects the people at risk of being stopped by police, assuming no bias. This method, however, has several important drawbacks.

First, the measure of race/ethnicity within the benchmark group is suspect. The benchmark group is composed of the *owners* of the violating vehicles, not necessarily the *drivers*. Therefore, a law enforcement agency cannot be sure that it has accurately measured the race/ethnicity of the people "ticketed" by the red light cameras. Second, any assessment of racially biased policing with this method is limited to certain locations for certain types of stops. The results in which an agency can have confidence relate only to the particular types of stops studied (red light and equivalent violations) and only to the specific intersections studied. To generalize from these

"spot checks" to other types of stops/violations requires an assumption without validity—namely, that the racial/ethnic profile of people who violate red light laws matches the racial/ethnic profile of people who commit all moving violations. Similarly, to generalize beyond the geographic test areas to the entire jurisdiction, a law enforcement agency must assume that those areas are representative of all areas of the target jurisdiction. This also is a shaky assumption for a number of reasons, including the likelihood that red light cameras are placed at intersections with higher than average traffic volume, violation behavior, and/or accidents.

This benchmarking method has addressed the following alternative hypotheses to the extent that it has created a match between the numerator and denominator data:

- *Racial/ethnic groups are not equally represented as residents in the jurisdiction.*
- *Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.*
- *Racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.*

- *Racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.*

Although a good match in terms of driving quantity, driving quality, and driving location has been created between the population of drivers stopped by patrol officers and the population of drivers ticketed by the red light cameras, this method has not assessed racially biased policing for all types of stops in the entire target jurisdiction—only for certain moving violations that occur in tested areas (or in areas where the results can be reasonably generalized).

Benchmarking with Radar Data

Radar enforcement, like red-light-camera enforcement, can be “blind” to the racial/ethnic characteristics of traffic law violators.⁷ The same implementation procedures and requirements apply to both methods. For example, the nonradar police stops of vehicles included in the numerator data should be as equivalent as possible to the radar stops that comprise the denominator data. The numerator and denominator data are already equivalent with regard to type of traffic offense—both sets of data include speeders. The researcher still would need to produce equivalence or a match in terms of the geographic locations of the stops.

The strength of benchmarking with “blind” enforcement data (whether it be radar data or red-light-camera data) is its potential to develop a strong match between the benchmark population and the people at risk of being stopped by patrol officers. Again, to the extent that this match is maximized, the factors related to the four competing hypotheses are addressed. Both the red-light-camera

method and the radar method have an important limitation: the rigor of the match comes at a cost in terms of scope. Conclusions can be made about specific areas and about enforcement of certain traffic laws but not about the target jurisdiction as a whole or enforcement of all traffic laws.

Benchmarking with Data from Low-Discretion Stops

Racial/ethnic bias is more likely to manifest itself when officers have discretion in deciding whether to stop someone than when they have little choice in the matter. When a driver runs a red light in a busy intersection, most officers feel a strong need to respond. This is an example of a low-discretion stop. An officer is likely to respond to all violations of this kind, and any racial/ethnic biases an officer might have are not likely to enter into his or her decision to ticket the red light violator. On the other hand, officers have great discretion in deciding whether to stop someone who is going 5 miles per hour over the speed limit (a violation at the other end of the degree-of-discretion continuum). If an officer has biases, they are more likely to influence high-discretion decisions such as this one.

In an attempt to measure racial bias in a jurisdiction, some law enforcement agencies have compared high- and low-discretion traffic stops. They use the low-discretion stops (the denominator) as a benchmark for the high-discretion stops (the numerator). Agencies adopting this method compare the racial/ethnic profile of the people stopped in high-discretion situations to the racial/ethnic profile of those stopped in low-discretion situations to see if the latter produces higher proportions of minority drivers.

⁷ Radar enforcement is not always conducted in a manner that makes it “blind” to the race and ethnicity of drivers. Radar enforcement is not “blind” if the

officer targeting the radar at cars can determine from his/her vantage point the race or ethnicity of the drivers.

This method has a serious shortcoming. Unlike benchmarking with data from “blind” enforcement mechanisms, benchmarking with low-discretion stops cannot produce a good match in traffic law-violating behavior between the people at risk of being stopped and the people who are stopped. Recall that the comparison between technological enforcement using red light cameras and enforcement by patrol officers matched stops of red light violators (the denominator) to stops of red light violators (the numerator)—comparable offenses. But the use of low-discretion stops (the denominator) as a benchmark for high-discretion stops (the numerator) involves a comparison of drivers who commit very different violations. Benchmarking with data from “blind” enforcement mechanisms (for instance, red light cameras, radar) addresses the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*; it does this by making driving behavior equivalent in the two groups. Benchmarking with data from low-discretion stops does not achieve equivalence across groups in driving behavior.

This method, however, does address two other alternative hypotheses: *racial/ethnic groups are not equally represented as residents in the jurisdiction* and *racial/ethnic groups are not equally represented as drivers on jurisdiction roads*. A third hypothesis—*racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high*—is addressed if the researchers conduct analyses, as we recommend, within subareas of the jurisdiction.

Benchmarking with “Blind” Data from a Nontechnological Source

Researchers from The RAND Corporation studied vehicle stop data collected by the Oakland (CA) Police Department, and they developed a benchmarking method that is based on comparing “blind” and “not blind” stops (Ridgeway, G., K.J. Riley and J. Grogger 2004). The RAND researchers considered various methods that would allow them to compare the stops in which officers *could* discern race/ethnicity to stops in which officers were “blind” as to the race/ethnicity of the driver. They ultimately decided to use time of day to differentiate between stops where officers had greater and lesser visibility.

The researchers did not simply compare daytime and nighttime stops because the people at risk of being stopped by police during the day are different from the people at risk of being stopped by police at night. (The demographic makeup of drivers on the road in any jurisdiction or subarea can vary considerably across times of day.) Instead the RAND researchers compared stops during a limited time period (approximately 5 P.M. to 9 P.M.) on the assumption that the *visibility* of driver demographics changes during these hours, but the demographics of the driving population does not change then significantly. The numerator data covered stops that occurred between 5:19 and sunset (“daylight stops”) and the denominator covered stops that occurred between the end of civil twilight⁸ and 9:06 P.M. (“darkness stops”). Again, with these two groups they were able to compare stops that occurred when presumably officers could see the race/ethnicity of the drivers in the cars to stops that occurred when presumably the officers could not. The researchers conducted their analysis on a

⁸ “Technically, this occurs when the center of the sun is 6 degrees below the horizon, but practically it is when one can see the brightest stars and artificial light is

needed to perform most outside activities” (Ridgeway, G., K.J. Riley and J. Grogger 2004: 40).

subset of stops to increase the strength of their comparison; they conducted their analysis on only evening stops and not on stops that occurred at other times of day. Once again the greater rigor of the match came at the cost of the scope of the analysis.

Drawing Conclusions from the Results

Using data obtained from red light cameras and radar, law enforcement agencies can compare the racial/ethnic profile of technology-selected drivers to the racial/ethnic profile of human-selected drivers (drivers stopped by patrol officers). This comparison benchmarks the data on drivers stopped by enforcement methods that are devoid of discretion (the “blind” technology) against the data on drivers stopped by methods that involve the exercise of discretion (stops by patrol officers). If officers’ stopping decisions are made without racial/ethnic bias, then the racial/ethnic profile of the drivers they stop will match the racial/ethnic profile of the drivers stopped by the technology. In a variation of this method, a researcher could benchmark high discretion stops against low discretion stops or compare groups of stops that differ in terms of the prestop observability of a driver’s race/ethnicity.

When implemented in accordance with our recommendations, benchmarking with data from “blind” enforcement mechanisms such as red light cameras and radar 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.

Benchmarking with data from low-discretion stops or with data from stops where officers could not discern racial/ethnic characteristics has limitations as well. Because the types of stops represented in the numerators (high discretion stops, stops in

which characteristics are observable) are, or may be, dissimilar from the types of stops in the denominator (low discretion stops, stops in which characteristics are not observable), this method does not address the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*. Consequently, the bias hypothesis cannot be tested directly.

BENCHMARKING WITH DATA FOR MATCHED OFFICERS OR MATCHED GROUPS OF OFFICERS

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 the factors reflected in the alternative hypotheses described in Chapter 2. 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 would be warranted. This method has been referred to by Samuel Walker as “internal benchmarking” (2001, 2002, 2003).

To implement internal benchmarking, the agency must be able to link stop data to individual officers or to groups of officers. Comparing officers to each other is preferable to comparing groups of officers. Analysis at the individual level allows the agency to identify particular officers whose stopping activity is different from his or her colleagues’ stopping activity and intervene if appropriate. (Some agencies cannot match at the individual officer level because the agency’s stop data cannot be linked to individual officers.)

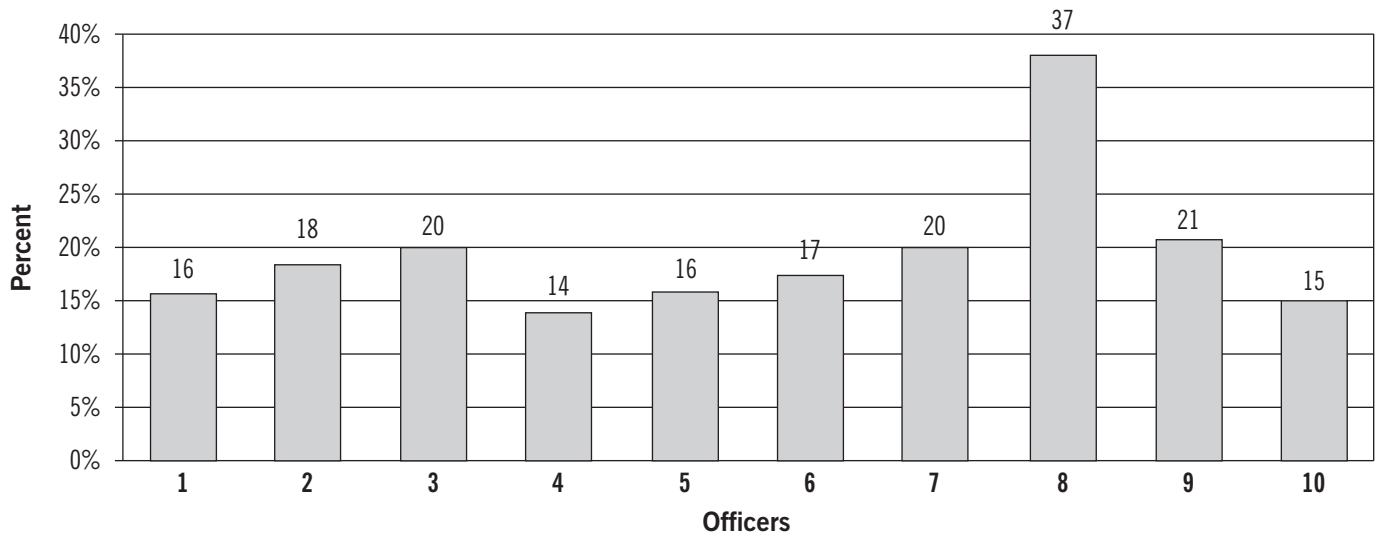


Figure 5.4. Matched Officers' Stops of Minority Drivers (Hypothetical Data)

The Matching Process

The strength of this method is directly linked to the quality of the match between the officers or groups of officers being compared. That is, the researcher wants to maximize the similarity among the officers being compared or among the groups being compared. To assess whether Officer A, for example, is making decisions to stop vehicles based on drivers' race or ethnicity, an agency can compare Officer A's stop data to the stop data of other officers who are policing essentially the same population in essentially the same way. The goal is to compare officers who are similar to one another in terms of the people at risk of being stopped by them. For instance, officers on the same shift, in the same geographic area, with the same assignment would be exposed to a similar population of drivers. Because the selected officers police similar populations, all of the factors related to the alternate hypotheses (driving quantity, driving quality, driving location) are held constant. The racial/ethnic profile of drivers on the road, as well as the racial/ethnic profile of law violators, are roughly equivalent for these matched officers. Since all of the

factors related to the alternative hypotheses are held constant in this comparison of individual officers, the racial/ethnic profile of the drivers they stop should be about the same unless one officer (or possibly several) is more inclined to stop drivers of particular racial/ethnic groups than are the others.

Figure 5.4 illustrates benchmarking with data for matched officers. For nine of the ten matched officers, the percentage of stops of minorities is in the range of 13 percent to 21 percent. (That is, for these officers, between 13 and 21 percent of the drivers they stop are minorities.) The percentage for Officer 8, however, is much higher—37 percent. This finding of disparate results does not *prove* that the officer is acting in a racially biased manner, but it should prompt a review of the policing activities of this officer.

An agency unable to link stop data to individual officers can still implement internal benchmarking if it can identify groups of officers that are similarly situated. That is, the unit of analysis would be the group not the individual. The numerator is the aggregate racial/ethnic profile of the drivers stopped by all of the officers in the group;

the denominator, or benchmark, is the racial/ethnic profile of the drivers stopped by the corresponding comparison groups. That is, the racial/ethnic profile of drivers stopped by Group A is compared to the racial/ethnic profiles of the drivers stopped by the officers in the matched Group B, matched Group C, and so forth.

Drawing Conclusions from the Results

Benchmarking with data for matched officers or matched groups of officers enables analysts to identify “outliers,” officers or groups of officers who stop racial/ethnic minorities at higher rates than do their matched counterparts. The degree of confidence analysts can have that policing by these officers is racially biased is entirely dependent upon the strength of the match. Perfect matches would fully account for the factors reflected in the alternative hypotheses and enable the analyst to test the bias hypothesis. 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 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.

In sum, definitive conclusions about racial profiling cannot be drawn from this benchmarking method because the match between individuals or between groups is imperfect: that is, the racial/ethnic profile of the drivers to which an officer or group of officers is exposed is not exactly the same as the racial/ethnic profile of the drivers to which the matched officers or matched group of officers are exposed. This internal benchmarking method can pinpoint outliers, but further review is essential to assess whether the disparity may be the result of bias.

There is another major drawback associated with this method: the relativity of the

findings. 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.

For example, it is clear from Figure 5.4 that Officer 8 is stopping minorities at a rate disproportionate to the rate of minority stops by his or her peers. But an analyst cannot conclude that the other nine officers in the match are stopping minorities in proportions that reflect legitimate stopping criteria: they, too, might be making decisions based on racial bias. Indeed, every officer in this matched group of ten officers could be practicing biased policing. In that case Officer 8 is only the officer whose stopping decisions appear to manifest bias most strongly. Similarly, in a group-level analysis, all of the groups in the comparison could be biased. From the analysis, however, the researcher cannot determine whether the matched groups are fair or biased in their policing. The analyst is able to identify only the officers or groups that stop the highest proportion of minorities.

To overcome this major obstacle (that is, the relativity of the findings), an agency could (if resources allowed) supplement internal benchmarking with other methods such as benchmarking with data from “blind” enforcement mechanisms or with observation data. In using internal benchmarking in conjunction with other methods, the researcher can take advantage of the great strengths of the internal benchmarking method and counter its greatest weakness as well.

Taking Appropriate Action against Officers or Groups

The data for outliers (whether individual officers or a group of officers) should not be considered proof of racially biased policing by them. The results of high-quality-match methods, however, do raise legitimate red flags that can and should prompt further investigation by the law enforcement agency. Specifically, the results justify a comprehensive inquiry into the officer's or group's stopping activity. The high rate of minority stops might have a legitimate explanation. For instance, the officer or group of officers might have a special assignment to a "hot spot" in the geographic area where minorities are present in numbers disproportionately higher than their representation in the rest of the geographic area.

Samuel Walker (2002, 86) advocates caution when law enforcement agencies interpret the results of this benchmarking method:

Where the data analysis identifies potential problem officers or supervisors, the [internal benchmarking] approach moves to the intervention stage. Intervention begins with a review of an officer's performance by supervisors. There may be extenuating circumstances that explain a particular pattern of traffic stops. The officer under review should enjoy a presumption of innocence until a full performance review is completed. The important point is that the *data represent a starting point*, the beginning of a departmental inquiry, and *are not in and of themselves conclusive*. Thus, no officer is automatically presumed guilty simply because he or she has made a high number of stops of minority drivers. A flexible system involving a command review of performance can accommodate officers who may be doing professional, proactive police work (emphasis in original).

OBSERVATION BENCHMARKING

Using the observation method, 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 to be the observers. Researchers have four choices to make in observation benchmarking:

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

Each of these questions is discussed below.

Methods of Observation

Observations can be conducted from stationary or mobile positions. With stationary methods, the researcher places observers at locations beside roadways; with mobile methods (also called "rolling" or "carousel" methods), the observers are placed in vehicles that move with traffic.

Stationary methods have been used most frequently to observe the demographic characteristics of drivers on urban and suburban roads. For instance, the researcher places observers at carefully selected intersections, and the observers record the race/ethnicity of the drivers passing through those intersections. The demographic profile of the people passing through the intersections is compared to the demographic profile of people stopped by police in the same geographic

areas. Stationery methods have also been used to create benchmarks for highway stops (see for instance, Engel, Calnon, and Dutil 2003, and Lange, Blackman, and Johnson 2001).

Using mobile methods, Smith et al. (2003) conducted a comprehensive study of stops made by the North Carolina State Highway Patrol. The research team placed a driver and three observers inside two “observer vans” that moved along selected roadway segments at the speed limit. One observer recorded the demographic characteristics of the drivers of vehicles passing the van (along with information regarding the vehicle), and the other two observers measured the speed of these vehicles using stopwatches. (The two speed measures were averaged.) With this information the researchers were able to compare the demographic profile of speeders to the demographic profile of drivers stopped by police.

Focus of Observations

Researchers using observation benchmarking need to decide whether to compare the stop data (the numerator) against demographic data for all drivers regardless of driving quality and/or for traffic law-violating drivers. The former entails collecting data on the race/ethnicity of drivers on the roadways; the latter entails collecting data on the race/ethnicity of drivers who are violating specific traffic laws. If demographic data are collected on *all* drivers (without distinguishing between nonviolating and violating drivers), the agency has addressed the alternative hypothesis that *racial/ethnic groups are not equally represented as drivers on jurisdiction roads*. It has not, however, addressed the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior*.

Measuring Race/Ethnicity

The assessment of race/ethnicity for the benchmark data relies upon the perception of

the observers—and their perception, presumably, will be in error some unknown proportion of the time. Similarly, observation is the preferred way for officers to measure race/ethnicity for purposes of filling out their data collection forms; to the extent that officers make stopping decisions based on race/ethnicity, they do so based on their perceptions of race/ethnicity, not on the basis of, for instance, information on the driver's license. Since observation by officers is the preferred method for identifying race/ethnicity for the numerator data, observation by trained observers is equally viable as the method for obtaining the denominator data.

It is difficult for both police and observers to make fine distinctions between racial and ethnic groups. In the context of implementing the observation method, this difficulty has ramifications for the categories of race and ethnicity used for data collection. Particularly problematic is identifying *ethnicity* through observation. It also is difficult for observers—particularly stationary observers collecting data on fast-moving vehicles—to distinguish among, for instance, Middle Easterners, Hispanics, and Native Americans. The ability to discern race/ethnicity can be impeded by the time of day as well as by the speed of vehicles under observation. Because of the difficulty of perceiving accurately a driver's race/ethnicity, many of the researchers implementing observation benchmarking use two or three observers. Additionally, some researchers have addressed the problem of discerning the demographic characteristics of drivers by broadening categories of race/ethnicity to more closely match what observers can see (for instance, “Caucasian” and “not Caucasian” or “Black” and “non-Black”).

Location of Observations

For both stationary and mobile methods of observation benchmarking, researchers must determine the type of locations to be observed,

the number of locations, and their geographic area. The ability of the observers to discern the race/ethnicity of drivers is affected by the speed of the vehicles, lighting at the site, and where observers stand. In some urban or suburban locations, it may not be safe for observers to stand too near the traffic. Conducting observations at intersections with a stop sign or traffic light may be safer and provides better visibility of drivers because traffic is slowed and the lighting may be better at night. Like stationary methods of observation benchmarking, mobile methods must be structured to ensure visibility. Observations must take place on thoroughfares with at least two lanes in each direction so that cars can pass the observer vehicle and be passed by it. Lighting may also be a consideration during evenings and nights, but vehicle speed is less likely to be a factor since the observer vehicle is moving with the traffic. In making site selections, researchers also should consider the volume of activity. The selected sites must have sufficient numbers of both police stops (numerator data) and cars and/or violators passing by the sites (denominator data) to produce reliable results.

Timing of Observations

Observation benchmarking requires researchers to make decisions not only about the method, focus, and location of observations but also about their timing (the days of the week, the times of the day, and the length of the reference period). Decisions related to timing are important because the racial/ethnic composition of drivers on the roadways may vary considerably across days of week, times of day, or even seasons of the year. Choices related to the timing of observations (the denominator data) will affect time-related choices with regard to the stop (numerator) data.

In selecting days of the week for scheduling observations, researchers strive for “representativeness” in the nature and extent

of traffic behavior. Observations could cover all days of the week or, to be more efficient, researchers could develop “categories of days.” For example, a researcher might make the reasonable assumption that traffic on Mondays, Tuesdays, Wednesdays, and Thursdays is essentially similar in the area being studied but traffic on Fridays, Saturdays, and Sundays are each unique. Similarly, the times of day for collecting observation data should also reflect the goal of representativeness. For example, researchers would not conduct observations from 6 P.M. to midnight if they wanted to benchmark stops made during all times of the day.

In some jurisdictions traffic varies during different times of the year. (For example, a southwestern city may experience an influx of northern tourists during the winter months, and a university town with a popular football team may have more traffic during the fall.) This seasonal variation will affect the population of drivers on the roadways and thus the racial/ethnic profile of drivers. To account for seasonal variation in traffic, researchers can conduct observations at various points throughout a twelve-month period. Researchers who choose a reference period of less than one year (for example, six months) should include in the report a caveat that the results do not necessarily apply to the parts of the year for which data were not analyzed.

Conducting the Analysis

Researchers match the police stops (the numerator data) with the observation (or denominator) data at each site with regard to the types of violations observed (for instance, speeding), the geographic location of the stops, the time of day, and the reference period. In hypothetical City A, for example, observers collected demographic data at fifteen sites (intersections) for all drivers violating speeding, red light, and/or stop sign laws. The observation data were collected during

randomly selected time blocks between 7 A.M. and 7 P.M. on all days of the week for the period January through June. Around each site the researcher identified a perimeter within which the traffic resembled the traffic going through the intersection. Then, within each of these fifteen geographic areas, the researcher selected the police stops that occurred from January to June for speeding, red light, and/or stop sign violations between the hours of 7 A.M. and 7 P.M. For each of the fifteen sites, the researcher compared the demographic profile of the people stopped to the profile of the people observed.

Drawing Conclusions from the Results

The observation method, conducted in accordance with standard social science methods, can provide meaningful information for a jurisdiction exploring the existence of racially biased policing. The assessment, however, is limited because the researcher is only able to conduct “spot checks” of racially biased policing. For example, in hypothetical City A mentioned above, the researcher 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.

With observation benchmarking, like other benchmarking methods, the stronger the “match” between the numerator and denominator data, the greater the confidence the researcher can have in the results. However, this increased confidence comes at a cost in terms of the scope of the assessment. Another potential drawback of the matching process is that it may reduce the numbers of stops in the numerator and/or the number of observations in the denominator to the point that some analyses become unreliable.

The observation benchmarking method addresses the alternative hypotheses (explained in Chapter 2) that *racial/ethnic groups are not equally represented as residents in the jurisdiction* and that *racial/ethnic*

groups are not equally represented as drivers on jurisdiction roads. If analyses are conducted separately for specific geographic locations, the method addresses the hypothesis that *racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.* If observations are made of drivers rather than violators, the method does not address the alternative hypothesis that *racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.* If, however, the observations are made of drivers violating particular traffic laws, the method addresses this final hypothesis.

OTHER

BENCHMARKING METHODS

This chapter has explained how researchers can compare data on stopping activity by police to adjusted census data, Department of Motor Vehicle data, data from blind enforcement mechanisms (such as red light cameras and radar), peer officers or units, and observation data. In addition to these benchmarking methods, researchers can conduct

- crime data benchmarking,
- crash (auto accident) data benchmarking,
- transportation data benchmarking, and
- survey data benchmarking.

These four methods, although they have received less attention from researchers, are noteworthy. We briefly summarize them here in order to communicate the full spectrum of methodology options now being considered by jurisdictions nationwide.

Crime Data Benchmarking

In crime data benchmarking, unlike the other benchmarking methods we have described, the benchmark data are not compared to all “vehicle stops” (stops made by police of a person in a vehicle) or to “traffic stops”

(vehicle stops the stated purpose of which is to respond to a violation of traffic laws). Rather, the benchmark data are compared to a subset of vehicle stops: the “investigative stops” (police stops of people in vehicles when there is at least reasonable suspicion of criminal activity).^{9, 10}

Researchers conducting crime data benchmarking 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). In this method, the measures of crime (1) must be linked to the race/ethnicity of the suspect or perpetrator and (2) must reflect as closely as possible actual crime as opposed to crime responded to by police.

Arrest data from the Uniform Crime Reports (UCR) contain race/ethnicity information (satisfying the first criterion). But crime data that meet the second criterion are difficult for researchers to obtain. The demographic profile of people arrested in a particular jurisdiction (as reported in the UCR) reflects two factors: (1) who commits crime and (2) whom the police identify and target for arrest. The decisions made by police regarding whom to target for arrest could be affected by racial bias. If a law enforcement agency is racially biased in both its arrests and investigative stops of vehicles and if it uses arrest data as a benchmark for investigative stops, its bias will not be revealed in the results.

To lessen the problem of police bias skewing the results, some researchers have compared investigative stop data to carefully selected subsets of arrest data (for instance, arrests in which there is minimal police discretion).¹¹ The research team that conducted analyses in San Diego (Cordner, Williams, and Zuniga 2001; Cordner, Williams, and Velasco 2002) developed a racial/ethnic profile of criminals based on crime reports in which victims and witnesses provided descriptions of the race/ethnicity of the perpetrators. The San Diego team selected this measure of crime because of its strength with regard to the criterion stated earlier: the measure is minimally affected by police discretion.

Using viable measures of crime, a researcher can compare the racial/ethnic profile of drivers stopped by police to investigate crime (so-called investigative stops) to the racial/ethnic profile of suspected criminals within subareas of the jurisdiction.

Crash Data Benchmarking

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). Most researchers have used crash data to estimate “who is driving” rather than “who is driving poorly.” In the two major studies that used crash data, the North Carolina team (Smith et al. 2003) developed its benchmark using data on all people involved in crashes; the Miami-Dade team (The Alpert Group 2003) used data only on the drivers adjudged not to be at fault in

⁹ Using crime data to benchmark traffic stops would require one to make a tenuous assumption—namely, that the same people who commit traffic violations are the ones who commit crimes and vice versa.

¹⁰ Also refer back to the discussion in Chapter 4 regarding the analysis of subsets of data based on whether

the stops are for traffic violations or for suspicion of criminal activity.

¹¹ Serious crimes are examples of crimes where there is minimal police discretion to arrest. For instance, if police have probable cause to link a person to a murder or robbery, it is very likely an arrest will be made.

the crashes. The latter decision was based on transportation literature that conjectures that not-at-fault drivers in two-car crashes are a representative subset of drivers on the road.

If a jurisdiction decides to use crash data as a benchmark, it should make sure that

- (1) Race and/or ethnicity information on the drivers involved in the crashes is available.
- (2) The researchers have reasonable confidence that racial/ethnic groups in the jurisdiction report crashes at similar rates.
- (3) The researchers have reasonable confidence that the filing of accident reports by officers is systematic (that is, filed for all crashes reported to police or filed for some clearly defined subset of crashes).¹²
- (4) The crashes in the data set can be linked to their geographic locations within the jurisdiction so that researchers can conduct subarea analyses (see Chapter 4).

Working with crash data will be particularly challenging for researchers if the data are not computerized. Another problem relates to sample size. Although there may be numerous crashes within a jurisdiction as a whole, within particular subareas the number of crashes available for analysis may be too few to provide reliable assessments.

Transportation Data Benchmarking

The National Highway Traffic Safety Administration (NHTSA) collects demographic data—including race but not ethnicity—on drivers. While valuable for measuring the

demographics of drivers, these national data are not useful to individual jurisdictions analyzing their police–citizen contact data because it is not viable to assume that the demographic profile of drivers produced by a national study will mirror the demographic profile of drivers in a particular jurisdiction. State and local transportation data, however, can help researchers compare the racial/ethnic profile of drivers stopped by police to the racial/ethnic profile of drivers driving on jurisdiction roads (or violating traffic laws on jurisdiction roads).

The U.S. Department of Transportation gathers travel behavior information through surveys of randomly selected households. One such survey is the National Household Transportation Survey (NHTS) conducted every five years. Respondents to the NHTS report their race and ethnicity and other demographic characteristics. Unfortunately, most jurisdictions cannot use the data from the NHTS to benchmark their vehicle stop data. Again, national data may not accurately reflect the demographic profile of drivers in a particular jurisdiction. Moreover, there are generally not enough respondents within individual jurisdictions to produce a reliable profile.

The national data from the NHTS, however, *can* be produced at the local level; a jurisdiction can “purchase” a sufficient sample to produce valid jurisdiction-level results. That is, a jurisdiction (for instance, city, county, state) can request an NHTS “add on” that directs the U.S. Department of Transportation to sample more residents in the jurisdiction than it would have for purposes of the national survey. Sufficient numbers of residents are then surveyed to produce a

¹² Officers working in high-crime areas where they are kept very busy may be less likely to take reports on minor accidents than are officers working in areas where fewer problems requiring their attention arise.

If these two types of areas also vary by racial/ethnic composition, the accident reports for the jurisdiction will not be representative of people involved in accidents.

reliable assessment of the transportation behaviors of residents within those jurisdictions.

Some local jurisdictions and states conduct their own travel behavior surveys independent of the NHTS. The data produced by these surveys—if race/ethnicity data are included—can be used by a researcher to develop a benchmark for “who is driving.”

Survey Data Benchmarking

Some researchers have used survey data to try to assess whether policing in a particular jurisdiction is racially biased. These researchers have conducted surveys (written surveys, telephone interviews, or face-to-face interviews) of scientifically selected residents of the jurisdiction. The 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 researchers with valuable information on the various factors, referenced throughout this report, that can affect a driver’s risk of being stopped by police.

Survey data benchmarking has a unique advantage: the numerator data include people the police *don’t stop* as well as those they *do stop*. This is because the survey goes to a scientifically selected sample of residents, some of whom have been stopped by police during the designated reference period and some of whom have not. All of the other benchmarking methods discussed in this report rely upon the police-citizen contact data collected by police; that numerator data include information only on people who were stopped.

This benchmarking method also has several drawbacks. Respondents may not

accurately report the information that is solicited because their memories fail. In the case of vehicle stop information, they may forget some stops by police or provide faulty reports of driving quantity or quality based on their less-than-perfect memories. The faulty memories may lead to “telescoping,” a term to describe the reporting of an incident as occurring during the reference period when, in fact, it occurred before the start of the reference period. Some answers may not be fully accurate—not because of the faulty memories of the respondents—but because the respondents want to “look good” or “say the right thing.” This “social desirability effect” could be particularly applicable to the questions in a survey to measure racially biased policing. Some respondents may underreport stops by police because they are embarrassed about them or want others to think that their driving quality is better than it truly is. If faulty memories (including telescoping) and the social desirability effect do not manifest equally across racial/ethnic groups, the survey method will produce distorted assessments of racially biased policing.

Despite its shortcomings, this benchmarking method has been useful. With data from survey respondents regarding whether or not they were stopped by police, researchers can determine if there is disparity in the level of stops of various racial/ethnic groups in the target jurisdiction. A survey is valuable because it can link that disparity to causes by collecting from respondents information pertaining to the alternative hypotheses (that is, information regarding driving quantity, quality, and location).

CONCLUSION

This chapter has described methods that are being used around the country to benchmark stop data. Researchers can compare their data on stopping activity by police to adjusted census data, Department of Motor Vehicle data, data from blind enforcement

mechanisms, peer officers or units, observation data, crash data, and other transportation data. Researchers also can benchmark investigative stops against crime data and conduct surveys to collect “numerator” and “denominator” data simultaneously.

All of these methods have strengths and weaknesses; none of them can prove or disprove the existence of racially biased policing in a jurisdiction. A causal connection between drivers’ race/ethnicity and police stopping behavior cannot be proven, but this does not mean that data collection is for naught. By collecting police-citizen contact data, a law enforcement agency conveys an important message to residents: it shows that the agency is concerned about racially biased policing, is open to scrutiny, and is accountable to its constituency. Even if the results do not provide definitive conclusions regarding racial bias, they can serve as a basis for constructive discussions between police and community members regarding ways to reduce racial bias and/or perceptions of racial bias.

If an agency chooses to collect data, this effort should be only one component of its comprehensive response to the issues of biased policing and the perceptions of its practice. Reforms in the realms of supervision, policy, training, community outreach to minorities, and recruitment also should be considered (see Fridell et al. 2001). Further, if the agency chooses to gather information to measure racial bias, it might consider sources other than—or in addition to—stop data. It can hold police-citizen forums to learn about citizens’ concerns and perceptions, scrutinize complaints by the public, and organize meetings with supervisors to assess/discuss potential problems. Multiple responses to the issues of racial/ethnic bias are possible, and multiple sources of information are available to guide agency reforms. Before we explain how police-citizen contact data can be used in constructive ways (Chapter 8), we discuss actions police take after a driver has been stopped (Chapter 6) and how the results of benchmarking methods can be conveyed and interpreted (Chapter 7).

VI

Guidelines for Analyzing Poststop Activities by Police

Two questions have interested researchers analyzing data on vehicle stops:

- Does a driver's race/ethnicity have an impact on vehicle stopping behavior by police?
- Does a driver's race/ethnicity have an impact on police behaviors/activities during the stop?

The benchmarking methods discussed in Chapter 5 addressed the first question. Here we consider the second. Attention to poststop activities is important. Some stakeholders have expressed concern that poststop activities by police are more likely than stop decisions to be influenced by racial bias.

The poststop activities most commonly examined by jurisdictions are searches and stop dispositions (the officer's decision to arrest, ticket, warn, or provide no disposition). Other aspects of the stop (for example, length of stop and whether a person was asked to exit the vehicle) also have been examined by researchers interested in assessing whether policing in a jurisdiction is racially biased.

ANALYZING SEARCHES

Some agencies collect information on searches of vehicles and their occupants. Officers in these agencies record search information on the police-citizen contact forms. The search data collected on the forms can be analyzed by researchers in two ways: researchers can calculate the "percent searched" for each racial/ethnic group, and researchers can calculate "hit rates" (the percent of searches in which the officers find something) for each racial/ethnic group. Searches are intrusive behaviors by police, and search data can help researchers explore whether policing in a jurisdiction is biased. To analyze search data, however, requires certain resources.

Resources Required

For effective analysis of search data, jurisdictions must make sure that officers collect certain information on the forms they fill out. The form should include an item indicating whether or not a search was conducted. In addition, the form should solicit information on the legal authorization for the search.¹

¹ Example responses include probable cause, reasonable suspicion that a person is armed, incident to arrest,

"plain view," warrant, inventory, consent, and probation/parole waiver.

To examine “hit rates,” the agency must include on its form an item to indicate search results are either “positive” (something found) or “negative” (nothing found). To allow for a full examination of consent searches requires an item on the form that asks: “Did you request consent to search? Yes or No.”

Types of Search Data Analyses

“Percent Searched” Measures

“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, then the percent searched is 20 ($20/100 \times 100$). If 200 Caucasians were stopped in their vehicles and 35 of them were searched, then the percent searched is 17.5 ($35/200 \times 100$).

These percentages are often used erroneously to draw conclusions regarding racial bias. In many jurisdictions higher proportions of stopped minorities are searched than stopped Caucasians. Analysts, stakeholders, reporters, and even expert witnesses have mistakenly concluded 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 or, relatedly, whether or not the disparity is justified. Not every person who is detained is at equal risk of being searched by police; there are very legitimate reasons why some persons are at greater risk of being searched than other persons. Indeed, the public should not expect equal search proportions across stopped groups. Virtually all agencies report that stopped men are

searched in greater proportions than stopped women. Does this finding indicate police bias against men? Not necessarily. It could be that more men are at greater legitimate risk of being searched by police than women because men, more than women, manifest behaviors that provide legal grounds for a search.

Figure 6.1 provides data from a hypothetical jurisdiction showing “percent searched” data for racial/ethnic groups by gender. This figure indicates that 16 percent of the Caucasian males who were stopped by police were searched. Corresponding figures for African American males, Hispanic males, and “Other” males were 24, 21, and 15 percent, respectively. Similar information is provided for the females who were stopped by police. These data indicate disparity; detained minorities (particularly detained African American and Hispanic males) are searched more frequently than Caucasians. These results do not provide information regarding the cause or causes of that disparity.

Search “Hit Rates”

A hit rate is the percent of searches in which the officers find something upon the people being searched. Officers might find contraband (for instance, drugs, illegal weapons) or other evidence of a crime. 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.

A hypothetical example will help explain hit rates. If, during a specified reference period, police in an agency searched 100 of the stopped Caucasians, 80 of the stopped African Americans, and 60 of the stopped Hispanics and found evidence on 10 of the Caucasians, 4 of the African Americans, and 4 of the Hispanics. The hit rates would be 10 percent for Caucasians ($10/100 \times 100$), 5 percent

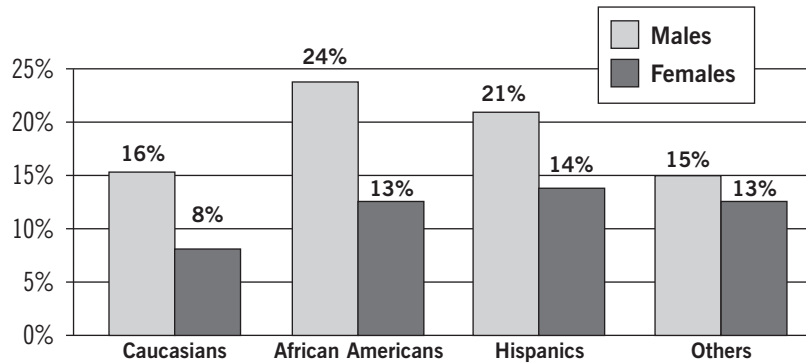


Figure 6.1. Searches as a Percentage of Vehicle Stops, by Race/Ethnicity and Gender of Detained Group, Hypothetical Jurisdiction

for African Americans ($4/80 \times 100$), and 7 percent for Hispanics ($4/60 \times 100$).

For all types of searches, hit rates provide descriptive information regarding whether or not there is disparity in “productivity.” If, for instance, 22 percent of the searches incident to arrest of African Americans produced hits compared to 30 percent of the searches incident to arrests of Caucasians, the Caucasian searches of this type are more productive. This is valuable information warranting exploration by the jurisdiction, but bias may not be the cause of this disparity. Legitimate factors may account for the differential hit rates.

Evidence-Based Searches. For “evidence-based searches,” however, researchers can say with reasonable confidence that any identified disparity is unjustified and likely caused by bias. For this subset of searches, search hit rates can rule out (not definitively but with an acceptable degree of confidence) the alternative hypotheses (hypotheses that factors other than bias influence police behavior). An economic theory called the “outcome test” will help us understand how.

The outcome test can be applied only when decision makers claim that their decisions are based on the probability of a particular outcome. First proposed by Nobel Prize-winning economist Gary S. Becker

(1993), the outcome test was applied by him to outcomes related to money lending. Assume a bank claims to make loan decisions based on the likelihood that the borrower will be able to pay the loan back. If the bank applies this criteria (probability of loan repayment) equitably across all racial/ethnic groups, then the default rates should be equal across groups. In other words, racial/ethnic groups should succeed in their loan repayment at the same rates. If, in fact, the minority borrowers default on their loans at a *lower rate* than their Caucasian counterparts, researchers can infer that the criteria for determining who would get a loan were not the same for Caucasians and for minorities; researchers can infer that minority borrowers were held to a higher standard by those deciding to make the loans.

The above example pertains to the differential allocation of benefits (that is, loans) across racial groups. The same test can be used to assess a decision maker’s allocation of detriments (for example, searches) across racial/ethnic groups. As Ian Ayres (2001, 406) explains, if police decisions to search minorities are “systematically less productive” than police decisions to search whites, one might infer that undeserving minorities are being subjected to searches. From these results one might infer that *different standards* were

utilized in selecting Caucasians and minorities for searches. Specifically, it appears that a lower standard of proof was applied to searches of minorities than to searches of Caucasians.

The outcome test does not focus on whether different proportions of minorities and Caucasians are searched. Different proportions of Caucasians and minorities might meet legitimate, unbiased criteria for a search. The outcome test focuses on the pool of people that the decision maker “deemed qualified” for a search (or for a loan, in the earlier example).

Another way to restate the example is by using a hypothetical construct, “units of evidence.” Imagine an officer who searches all minorities he detains for whom he has 50 units of evidence that they are carrying contraband or other evidence. He searches all Caucasians he detains for whom he has a corresponding 80 units of evidence. He has set a lower standard for searching minorities compared to Caucasians. The result will be that he is “wrong” more often with his minority searches; the officer is less likely to find evidence on the minorities, because he settled for a low level of evidence to initiate the search. He will have more “hits” in his searches of Caucasians because he didn’t search them unless he was highly confident that they were carrying contraband/evidence.² This produces a lower hit rate for minority searches. As Ayres explains (2002, 133), “A finding that minority searches are systematically less productive than white searches is accordingly evidence that police

require less [evidence] when searching minorities.”

As Ayres (2002, 134) explains, “The decision maker in an outcome test by her own decisions defines what she thinks the qualified pool is, and the outcome test then directly assesses whether the minorities and nonminorities so chosen are in fact equally qualified.” The bankers will claim that they make loan decisions based only on the probability of default. The corresponding circumstance for police is when they make searches based on the probability of finding contraband/evidence. This is true when the police conduct probable cause searches, frisks for weapons, searches based on “plain view” or drug odors, and, arguably, canine alert searches.³ These types of searches are “evidence-based searches.” The requirement of the outcome test (decisions must be based on the probability of a certain outcome) is not met with other types of searches, such as searches incident to a lawful arrest, inventory searches, or warrant searches.⁴

Table 6.1 provides sample results showing hit rates for evidence-based searches for groups defined by their race, age, and gender. These hypothetical data indicate that hit rates for evidence-based searches of young minority males are lower than for any other group.⁵ For young African American males and for young Hispanic males, the hit rates are 8 percent and 6 percent, respectively. All other groups have hit rates of at least 13 percent. Results such as these should prompt law enforcement agencies to examine their searches more closely and/or implement

² A “wrong” decision does not imply that the search was unjustified; similarly, a “hit” does not imply that the basis for the search was legitimate.

³ We discuss consent searches separately below.

⁴ This means that hit rate analyses will be conducted separately for different subsets of searches. The hit rate analysis conducted on the subset of evidence-based

searches can be interpreted in accordance with the outcome test.

⁵ Again, a small number of searches in a jurisdiction may preclude breakdowns of the data within categories such as race, gender, and type. Analyses with small numbers are unreliable.

Table 6.1. Evidence-Based Search "Hit Rates," by Race/Ethnicity, Gender, and Age, Hypothetical Jurisdiction

Race/Ethnicity	Female		Male	
	≤24	25+	≤24	25+
Caucasian	17%	14%	15%	16%
African American	13%	15%	8%	15%
Hispanic	15%	16%	6%	17%
Other	15%	13%	14%	15%

Source: Based on a table in Council on Crime and Justice and Institute on Race and Poverty (2003, 29).

interventions to reduce this apparent (albeit not proven) bias in searches. For instance, law enforcement agencies could expand collection of quantitative or qualitative data on searches to gather more information or implement interventions to eliminate or decrease potential bias in search decisions.⁶

The Special Case of Consent Searches.

With nonconsent evidence-based searches,⁷ the decision of the officer that is evaluated in the outcome test is the decision to conduct the search; in every instance the researcher will know whether or not the officer was right or wrong about whether the person was carrying contraband or other evidence. If the officer wants to conduct 100 nonconsent evidence-based searches, he will conduct 100 of them, and the researcher will know from the form filled out for each one whether or not there was a "hit."

With consent searches, however, the decision of the officer that is evaluated is the decision to *request consent* to search. The researcher wants to know if the officer, because of bias, requests consent to search from

minorities more than from Caucasians. The officer may want to conduct 100 consent searches but be able to conduct only 85 because consent is withheld by 15 people. To properly evaluate the officer's decision using the outcome test, the researcher would need to know for all 100 people from whom consent was requested who *was* and *was not* carrying contraband or other evidence. This information is known only for 85 of the 100. The researcher cannot assume that the 85 are representative of the 100. It is plausible that the 15 who refused to provide consent are carrying evidence/contraband at a different (likely higher) rate than the 85 who consented, and it is possible that the relationship between refusal and carrying differs across demographic groups.

If an agency has a large number of people who refuse to provide consent, the agency cannot include consent searches in the category of evidence-based searches for purposes of conducting hit rate analysis. There is no clear rule of thumb for when the level of missing "consent search" data is sufficiently

⁶ See Chapter 8 for actions agencies can take to reduce the potential for bias in consent searches and other high-discretion activities.

⁷ Again, these are probable cause searches, frisks for weapons, searches based on "plain view" or drug odors, and canine alert searches.

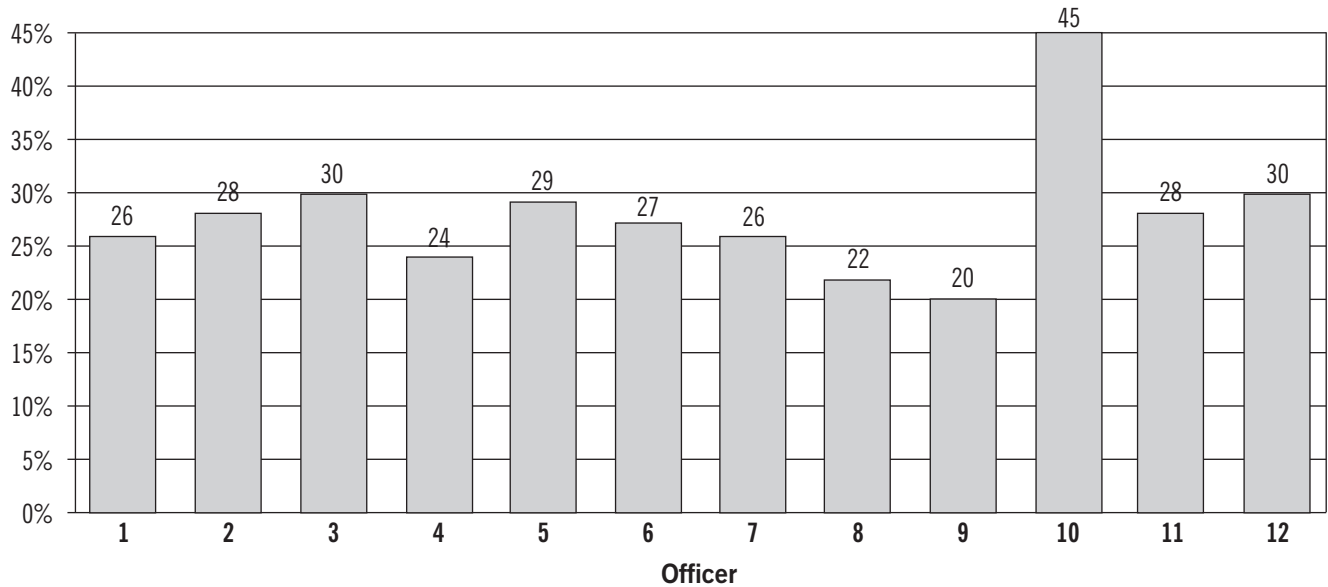


Figure 6.2. A Comparison of Twelve Officers Who Are Similarly Situated (“Matched”): Percent of Drivers Searched Who Are Minorities

low to determine unjustified disparate impact. However, we maintain that a researcher who has at least 95 percent agreement to the consent searches within *each* of the racial/ethnic groups can analyze the hit rates for consent searches and interpret them in accordance with the outcome test.⁸

Lower hit rates for minorities for evidence-based searches signal the possibility of racial bias. These findings are sufficient grounds for further exploration by a department, but they are not *conclusive evidence* of bias. Similarly, equal hit rates for minorities and Caucasians are not conclusive evidence that search decisions are bias-free. Legitimate factors unrelated to bias can produce lower hit rates for minorities, and biased police actions can produce equal hit rates in certain

circumstances (see Fridell 2004, chap. 11). Researchers should consider these factors or circumstances when interpreting hit rate results.

Other Ways to Examine Searches

In addition to analyzing “percent searched” data for racial/ethnic groups and “hit rates” for racial/ethnic groups, researchers can analyze searches in other ways. One way is to conduct “internal benchmarking” with search data.

Recall from Chapter 5 how the internal benchmarking method is implemented: To analyze stopping behavior by police, agencies compare stops by individual officers to stops by other similarly situated officers, or they compare stops by a group of officers to stops by other similarly situated groups of officers.⁹

⁸ To conduct this analysis, the law enforcement agency first must include in the data collection form an item regarding whether or not the person was asked for consent to search.

⁹ For instance, they compare officers who are assigned to the same geographic area, the same shift, and who have the same mission (such as patrol). These similarly situated officers are exposed to the same group of people at risk of being stopped by police.

Table 6.2. Stop Dispositions for Caucasians and Minorities

Race	Arrest	Citations	Warning	No Disposition	Total
Minorities	6%	59%	23%	12%	100%
Caucasians	5%	62%	25%	8%	100%

In the same way, agencies can compare similarly situated officers with regard to the percent of drivers searched who are minorities. Figure 6.2 (previous page) illustrates how this is done. In this hypothetical jurisdiction, between 20 and 30 percent of the drivers searched by officers (Officers 1 through 9 and Officers 11 and 12 in the figure) are minorities. In contrast, 45 percent of the drivers searched by Officer No. 10 are minorities. Officer No. 10 is an “outlier” (in social science terminology), and this officer’s decisions to search should be reviewed by the department to see if bias is influencing them.

ANALYZING STOP DISPOSITIONS

Does a driver’s race/ethnicity have an impact on what happens during a vehicle stop? To address this question, jurisdictions can analyze search data. They also can analyze data on stop dispositions (for instance, arrest, citation, warning, no action).

Jurisdictions do not agree on what disposition results indicate racial bias by police. Researchers for the Montgomery County (MD) Police Department (2001) have held that disproportionate representation of minorities among drivers given the *most serious* dispositions (arrests or citations) is an indication of bias. Other analysts have claimed racial bias is indicated by the disproportionate representation of minorities among those receiving the *least serious* dispositions such as warnings or no disposition. Such “low-level” outcomes are not viewed by them as a sign of police benevolence but as evidence that there may

have been no legitimate reasons for these stops in the first place. More low-level dispositions for minorities than for Caucasians is seen by this group as evidence of police “fishing” for evidence of crime among minorities.

These varied interpretations of disposition information reflect the challenge researchers face when analyzing this type of data. 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. The results in Table 6.2 for a hypothetical jurisdiction show that minorities are over-represented among drivers receiving “no disposition.” Like the “percent searched” data, disposition data can identify disparity in police actions but not the cause of that disparity.

Not all drivers are at equal risk of being searched; similarly, 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 team analyzing the vehicle stop data for the Washington State Patrol (WSP) (Lovrich et al. 2003) determined that the quantity and seriousness of the violations by the stopped driver appear to be key variables that influence police disposition decisions. Other variables that might influence the

Table 6.3. Dispositions for Moving Violations, by Race/Ethnicity, Hypothetical Jurisdiction A

Race	Number Detained	Percent of Detained	Disposition							
			Arrest		Citation		Written Warning		No Disposition	
African Americans	6,405	15.65%	402	6.28%	4,600	71.82%	801	12.51%	602	9.40%
Hispanics	1,700	4.15%	54	3.18%	1,267	74.53%	234	13.76%	145	8.53%
Other Minority	8,182	20.00%	402	4.91%	5,998	73.31%	1,035	12.65%	747	9.13%
Caucasians	24,629	60.19%	623	2.53%	18,772	76.22%	2,997	12.17%	2,237	9.08%
Total	40,916	100.00%	1,481	3.62%	30,637	74.88%	5,067	12.38%	3,731	9.12%

dispositions police choose include (but certainly are not limited to) the stopped driver's demeanor, the prior driving record of the stopped driver, and the geographic location of the stop. An example will illustrate the importance of stop location. An officer might consider speeding 10 miles per hour over the speed limit in a school zone as a more serious offense than 10 miles per hour over the speed limit on a highway.

All of these factors can legitimately influence the dispositions chosen by police, and researchers analyzing disposition data should attempt to take them into account. This type of analysis, however, cannot be performed unless the agency has certain resources available.

Resources Required

The form that officers fill out should include an item regarding the disposition of the stop. Common options are arrest, ticket/citation, verbal warning, written warning, and no action. Information related to the reasons for stopping the vehicle are relevant to analyzing the dispositions of those stops. Therefore, data collection forms should include a field for "reason for the stop." There is a lot of variation across agencies with regard to the specificity of the "reason" options; an agency may have as few as five or as many as twenty "reasons" for the stop. As noted

earlier, information on the stop form regarding the quantity and seriousness of violations can be very useful.

Analysis of Dispositions within Categories of Stops

One of the factors that legitimately influences the choice of dispositions is the seriousness of the offense. For this reason, researchers try to control for or isolate this factor. If researchers examined dispositions for data that included all possible offenses, they would not know, for instance, if a finding that African Americans received harsher dispositions than Caucasians was due to bias or to the possibility that they committed more serious driving violations. Instead of doing one analysis of dispositions for all violations combined, researchers are encouraged to look at dispositions across races within offense categories such as speeding violations, red light violations, failure to yield violations, and so forth.

Table 6.3 provides hypothetical disposition data for *moving violations* in Jurisdiction A by race and ethnicity. Relative to the other groups, African Americans are slightly under-represented among detained persons who receive a citation for moving violations (71.82 percent) and slightly over-represented among people who are arrested (6.28 percent). Even

if these differences were larger, conclusions about racial bias could not be drawn. This is because proportionately more African Americans than the other groups *could have* presented behaviors (or been linked to other bases for an arrest such as outstanding warrants) that legitimately led to the arrest disposition.

Within the broad “reason for a stop” category of “speeding,” a researcher could refine the analysis even further. The researcher could subdivide this category based on information on the stop form regarding how many miles per hour the person was speeding. For example, the researcher might produce a table similar to Table 6.3 for each of the following categories of speeding: less than 10 mph over the speed limit, 10 to 15 mph over the speed limit, 16 to 20 mph over the speed limit, and greater than 20 mph over the speed limit.

Other Ways to Examine Dispositions

Dispositions can be analyzed within categories of stops based on the seriousness of the offense as described above. Researchers also can analyze dispositions by matching drivers of one race to “similarly situated” drivers of another race. In an analysis of vehicle stop data for the Oakland Police Department, the drivers were considered similar if they matched on variables such as location of the stop, time of the stop, whether the driver was an Oakland resident, age of the driver, reason for the stop, and driver gender (Ridgeway, Riley, and Grogger 2004). By comparing these similarly situated drivers, the research team could assess differences in dispositions across race while eliminating the possibility that some variables besides race (for instance, stop location, driver age) might be the cause of observed differences.

Another way to examine dispositions is through multivariate analyses, a topic covered in the next chapter.

ANALYZING OTHER ASPECTS OF A STOP

In this chapter we have discussed analysis of search data and analysis of stop disposition data. Other aspects of a stop can be analyzed as well. Some jurisdictions, for instance, collect information on the duration of the search or the duration of the entire stop; they might collect information regarding whether the driver (or passengers) were asked to exit the vehicle, whether canines were brought to the scene, and whether firearms were drawn. An agency may decide to include one or more of these variables in its analysis to understand more fully what happens during traffic stops in its jurisdiction.

The general analysis concepts presented above, indeed throughout the book, apply to these and any other variables. Researchers will attempt to identify the factors other than racial bias that might account for disparity with respect to any of these variables and either control for them with their methods or reference their omission in interpreting results.

CONCLUSION

Researchers can analyze search data in various ways. They can calculate and compare “percent searched” for racial/ethnic groups to indicate whether disparity exists, but they cannot draw conclusions from the data about the existence or lack of racial bias in the jurisdiction. Similarly, hit rates for all types of searches can provide an indication of whether disparity exists. Hit rates that meet the assumptions of the outcome test can indicate the existence of unjustified disparate impact. The term “unjustified disparate impact” means that the disparity is not easily explained by legitimate (nonbias) factors. The searches that meet the assumptions of the outcome test are evidence-based searches (those where the decision to search is based on the probability of finding contraband/evidence). Lower hit rates for minorities than for Caucasians for evidence-based searches

signal that racial bias *may* be influencing police decisions, and the agency should consider additional assessments of searches or reform measures (see Chapter 8). Consent searches cannot be analyzed with the outcome test unless high proportions of subjects within each of the racial/ethnic groups acquiesced to requests to search.

From disposition data, agencies can identify disparities across races, but they cannot draw firm conclusions regarding bias by police because of the challenges associated with controlling for the key factors that might legitimately affect the selection of dispositions.

The analysis of poststop data is complicated, and most methods can indicate only whether disparity exists, not the cause.

Despite these constraints, researchers should analyze poststop data and report to the law enforcement agency and other stakeholders comprehensive information regarding what happens after stops are made. These poststop activities are vulnerable to racial bias by police, and they could have great negative consequences for the driver subject to them. It is important for police executives to know what is happening during vehicle stops since these incidents comprise the most frequent interaction between police and citizens. As we discuss in Chapter 8, a finding of disparity, even if the cause of the disparity cannot be identified, may provide impetus for constructive changes in law enforcement policies or practices.

VIII

Drawing Conclusions from the Results

Previous chapters have explained ways in which data on vehicle stops by police and data on poststop activity by police (for example, searches and dispositions) can be analyzed. Jurisdictions are trying to determine whether there is a cause-and-effect relationship between a driver's race/ethnicity and police behavior.

To examine police decisions to stop, for instance, researchers compare the racial/ethnic profile of the people identified in the police-citizen contact data and the racial/ethnic profile of a "benchmark population." This population might be composed of residents of the jurisdiction with access to vehicles; drivers with a license; drivers identified by red light cameras, radar, or air patrols; drivers stopped by "matched" officers or groups of officers; drivers observed on the road by researchers; or drivers identified through other benchmarking methods.

Figure 7.1 illustrates disparity in an analysis of stop data: minorities are over-represented among drivers stopped relative to their representation in the benchmark population. They represent 19.06 percent of the stopped drivers and 15.60 percent of the benchmark population. In this hypothetical jurisdiction, there is disparity between racial/ethnic groups in terms of stops made by police.

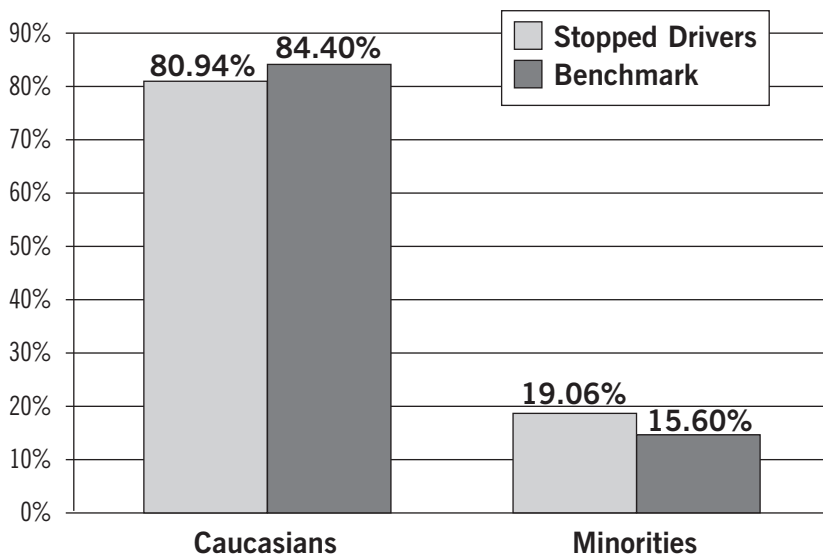


Figure 7.1. Disparity between Drivers Stopped by Police in Hypothetical Area A and the Benchmark Population for Area A, by Two Racial/Ethnic Groups

Disparity—such as that shown in Figure 7.1—can be conveyed in four ways: through absolute differences in percentages between those stopped by police and the benchmark population, relative differences in percentages, disparity indexes, and ratios of disparity. This chapter focuses on these four ways that disparity can be interpreted and conveyed to the public. In their analysis of stop, search, and disposition data, researchers can choose one or more of these measures of disparity. Two additional tools for assessing and conveying disparity—contingency analysis and multivariate analyses—are described as well.

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. We evaluate these opinions and explain useful tools that researchers can use to interpret data. This chapter explains 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 this book is that researchers can measure disparity easily, but identifying the cause of disparity presents a challenge. That theme continues through this chapter. 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 (nonbias) 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.

FOUR MEASURES OF DISPARITY

If benchmarking analysis reveals a disparity between the racial/ethnic profile of stopped drivers and the racial/ethnic profile of the benchmark population, researchers have a choice of four ways to measure and convey that disparity: *absolute differences* in percentages between those stopped by police and the benchmark population, relative differences in percentages, disparity indexes, and ratios of disparity. Table 7.1 explains how each of those measures can be calculated. The table is based on the disparity shown in Figure 7.1 between stopped drivers and the benchmark population for Area A in the hypothetical jurisdiction. To simplify the explanation, citizens are separated into just two groups: Caucasians and minorities.

Column A presents the number of stops of minorities and Caucasians across the reference period (for instance, one year). Researchers who are calculating measures of disparity should include in their tables the number of stops so that the discerning reader can assess whether this number is sufficient to produce reliable results.¹

¹ Analyses with small numbers of stops are less reliable than those with larger numbers of stops.

Table 7.1. Four Disparity Measures to Describe Stops in Hypothetical Area A, for Two Racial/Ethnic Groups

	A	B	C	D	E	F	G
	Number of Stops	Percent of Stops	Percent of Benchmark	Absolute % Difference	Relative % Difference	Disparity Index	Ratio of Disparity
Equation		$[A(m \text{ or } c)/t] \times 100$		B-C	$[(B-C)/C] \times 100$	B/C	F(m)/F(c)
Minorities (m)	15,492	19.06%	15.60%	3.46%	22.18%	1.22	1.27
Caucasians (c)	65,789	80.94%	84.40%	-3.46%	-4.10%	0.96	
Total (t)	81,281	100.00%	100.00%				

Note: These data produced the summary results presented in Figure 7.1.

Column B presents the percentage of the stops by police that were of minorities and of Caucasians (summing to 100 percent). Thus, for instance, the percentage of stops that were of minorities is 19.06 $[(15,492/81,281) \times (100)]$. Column C presents from Figure 7.1 the percentage of minorities and Caucasians in the benchmark population. If the jurisdiction were implementing benchmarking with adjusted census data (for instance, adjusted for access to vehicles), Column C would indicate that 15.60 percent of the jurisdiction's residential population with access to vehicles were minorities, 84.40 percent were Caucasians. If the benchmark represented people observed violating speeding laws (as opposed to jurisdiction residents), Column C would indicate that 15.60 percent of the people speeding on the jurisdiction's roads were minorities and 84.40 percent were Caucasians.

For this example, to calculate *absolute differences* in percentages between those stopped by police and the benchmark population, subtract Column C (representation of the group among the benchmark population) from Column B (representation of the group among the drivers stopped by police). For the minority group, the absolute percentage difference is 3.46 percent (19.06% – 15.60%). This result can be conveyed in the following

language: “there are 3.46 percent more minorities among the people who are stopped than are represented in the benchmark group.”

A second way that researchers can convey disparity is through *relative differences* in percentages between those stopped by police and the benchmark population. For the minority group in Table 7.1, the relative percentage difference is 22.18 percent or $[(19.06 - 15.60)/15.60] \times 100$. In other words, 19.06 percent is 22.18 percent greater than 15.60 percent. This difference could be expressed as follows: “there are 22.18 percent more minorities among the people who are stopped than are represented in the benchmark group.” Or, “minorities are over-represented among people stopped by 22.18 percent relative to their representation among the benchmark group. Similarly, whites are under-represented among people stopped by 4.10 percent relative to their representation among the benchmark group.” The wording used to describe absolute and relative differences in percentages is the same. There is no particular language for conveying the results that distinguishes the figures that are absolute percentage differences and relative percentage differences. Researchers should convey the meaning of the disparity by

describing in the report the equation used: either $B - C$ or $[B - C/C] \times 100$ (see Table 7.1).

A third way to convey disparity is a “disparity index.” For the minority group in Table 7.1, the disparity index is 1.22, which is calculated by dividing Column B (group percentage among drivers stopped) by Column C (group percentage among benchmark population). A value of 1 would indicate no disparity; that value would be obtained in our example if 19.06 percent of the stops were of minorities, and minorities comprised 19.06 percent of the benchmark population. A value greater than 1 indicates over-representation among drivers stopped relative to the benchmark, and a value less than 1 indicates under-representation among drivers stopped relative to the benchmark. The results in Table 7.1 indicate an over-representation of minorities among stops relative to their representation in the benchmarked group.²

A “ratio of disparity” (referred to by some researchers as an “odds ratio”) is the fourth way a finding of disparity can be conveyed. The disparity index for one group is divided by the disparity index for another group. The group in the denominator is the “reference group” to which the other group is compared. In our example, we use the disparity index to gauge how minorities (the numerator in the equation) fare relative to Caucasians (the denominator in the equation).

For the minority group in Table 7.1, the ratio of disparity is 1.27 ($1.22/0.96$). The disparity index for minorities is divided by the disparity index for Caucasians to produce a single number. A number greater than 1 indicates over-representation, and a number less than 1 indicates under-representation. Researchers could explain the ratio of disparity

shown in Table 7.1 in any of the following ways:

- “Minorities are stopped 1.27 times more than Caucasians.”
- “If you are a minority, you are 1.27 times more likely to be stopped by police than if you are Caucasian.”
- “For every Caucasian stopped, 1.27 minorities are stopped.”

Table 7.2 shows how to calculate ratios of disparity when there are more than two racial/ethnic groups. Because Hispanics comprised 8.24 percent of the stops and a very similar percent of the benchmark population (8.20 percent), the disparity index for Hispanics is 1.00 ($8.24/8.20$), indicating no disparity. The disparity indexes for African Americans and Caucasians show over-representation of African Americans relative to the benchmark (1.46) and under-representation of Caucasians (0.96). Recall that to produce the ratio of disparity for the two groups in Table 7.1, we divided the disparity index for minorities by the disparity index for Caucasians ($1.22/0.96 = 1.27$). To calculate the ratio of disparity with three racial/ethnic groups, researchers again must identify which of the three groups is the “reference group.” The disparity index for this chosen reference group becomes the denominator for the ratio of disparity calculations for the other two. We suggest that the relevant group in any calculation of a ratio of disparity for vehicle stop analysis be the Caucasian group. This is because the main question we are trying to answer is as follows: “Are minority residents treated differently from Caucasian residents because of their racial/ethnic status?”

² Consistent with our caveat that small sample sizes produce unreliable results, note that all of these measures are unstable when sample sizes are small.

Table 7.2. Disparity Indexes and Ratios of Disparity to Describe Stops in Hypothetical Area A, for Three Racial/Ethnic Groups

	A	B	C	F	G	
	Number of Stops	Percent of Stops	Percent of Benchmark	Disparity Index (B/C)	Ratio of Disparity	
					Formula	Result
African Americans (a)	8,798	10.82%	7.40%	1.46	F(a)/F(w)	1.53
Hispanics (h)	6,694	8.24%	8.20%	1.00	F(h)/F(w)	1.05
Caucasians (w)	65,789	80.94%	84.40%	0.96		
	81,281	100.00%	100.00%			

THE CHALLENGE OF SELECTING MEASURES OF DISPARITY

Above we explained four different ways that researchers can convey disparity: absolute percentage difference, relative percentage difference, disparity index, and ratio of disparity. These measures can be used to describe disparity in stops, dispositions, searches and other aspects of vehicle stops. We turn now to a new question: Which measure or measures of disparity should researchers select to present their data?

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. Multiple measures, they say, might lead the various stakeholders with different concerns or agendas to pick and choose the figures in the report that confirm their views or preconceived expectations regarding the results.

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. Care must be exercised in the interpretation of the findings. 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.

DIFFERENT MEASURES OF DISPARITY: DIFFERENT INTERPRETATIONS

Table 7.3 shows how the four measures of disparity explained in the first section of this chapter can convey very different results. The table presents four measures of disparity for three hypothetical police departments: A, B, and C. Which department has the most

Table 7.3. Various Measures of Disparity for Hypothetical Departments A, B, and C

Department	Representation of African Americans Among Stops	Representation of African Americans Among Benchmark	Absolute % Difference	Relative % Difference	Disparity Index	Ratio of Disparity
A	14.0%	9.0%	5.0%	56.0%	1.6	1.6
B	1.3%	0.6%	0.7%	117.0%	2.2	2.2
C	67.0%	54.0%	13.0%	24.0%	1.2	1.7

Source: Farrell 2004

disparity? Well, the answer depends on the measure of disparity we consider.

In terms of the absolute percentage difference, Department C has the most disparity: African Americans are over-represented in the stop data relative to the benchmark data by 13.0 percent. In terms of the other three measures of disparity, Department B has the most disparity. Although Department B has an absolute percentage difference of only 0.7, it has a relative percentage difference of 117. The disparity index and ratio of disparity for Department B are both 2.2. Department A has the second highest disparity when disparity is calculated as the relative percentage difference (56 percent) or disparity index (1.6); Department C has the second highest disparity (1.7) when calculated as ratios of disparity. Clearly, the measure chosen makes a difference in terms of the level of disparity indicated.

Recall the important point made earlier: if the population of stopped drivers or the population of the benchmark population has very high or very low percentages of minorities (or of Caucasians), the researcher's selection of one measure over another could make a big

difference in the interpretation of results. Table 7.3 shows that the percentage of minorities in both the stopped driver population and the benchmark population is low for Department B; as a result, the variation between two of the measures of disparity (the absolute percentage difference and the relative percentage difference) is extreme.³ Minorities represent only 1.3 percent of the persons stopped and only 0.6 percent of the benchmark population; the absolute percentage difference is tiny (0.7 percent), but the relative percentage difference is large (117 percent).

This extreme variation is even more evident in Table 7.4. In order to highlight the effects of low levels of minorities in the stop and benchmark populations on the four measures of disparity, we arbitrarily set the absolute percentage difference at 2 percent for thirty-five hypothetical departments. For low levels of minority representation (the top of Table 7.4), the relative percentage difference can be very high—misleadingly high—even when the absolute percentage difference is low (in these cases, 2 percent). For Department 2, minorities comprise 3 percent of the drivers stopped and 1 percent of the

³ Here we focus on the situation when the percentage of minorities is low in the stop and/or benchmark populations. The same problems would occur if

Caucasians were the group with low percentage representation.

**Table 7.4 Disparity Measures for Multiple Departments
When Absolute Percentage Difference is Set at Two**

Dept.	Percent of Stops		Percent of Benchmark		Percentage Difference		Disparity Index		Ratio of Disparity
	Caucasians	Minorities	Caucasians	Minorities	Absolute	Relative	Minority	Caucasian	
1	98	2	100	0	2.0	NA*	NA*	0.98	1.02
2	97	3	99	1	2.0	200.00%	3.00	0.98	3.06
3	96	4	98	2	2.0	100.00%	2.00	0.98	2.04
4	95	5	97	3	2.0	66.67%	1.67	0.98	1.70
5	94	6	96	4	2.0	50.00%	1.50	0.98	1.53
6	93	7	95	5	2.0	40.00%	1.40	0.98	1.43
7	92	8	94	6	2.0	33.33%	1.33	0.98	1.36
8	91	9	93	7	2.0	28.57%	1.29	0.98	1.31
9	90	10	92	8	2.0	25.00%	1.25	0.98	1.28
10	89	11	91	9	2.0	22.22%	1.22	0.98	1.25
11	88	12	90	10	2.0	20.00%	1.20	0.98	1.23
12	83	17	85	15	2.0	13.33%	1.13	0.98	1.16
13	78	22	80	20	2.0	10.00%	1.10	0.98	1.13
14	73	27	75	25	2.0	8.00%	1.08	0.97	1.11
15	68	32	70	30	2.0	6.67%	1.07	0.97	1.10
16	63	37	65	35	2.0	5.71%	1.06	0.97	1.09
17	58	42	60	40	2.0	5.00%	1.05	0.97	1.09
18	53	47	55	45	2.0	4.44%	1.04	0.96	1.08
19	48	52	50	50	2.0	4.00%	1.04	0.96	1.08
20	43	57	45	55	2.0	3.64%	1.04	0.96	1.08
21	38	62	40	60	2.0	3.33%	1.03	0.95	1.09
22	33	67	35	65	2.0	3.08%	1.03	0.94	1.09
23	28	72	30	70	2.0	2.86%	1.03	0.93	1.10
24	23	77	25	75	2.0	2.67%	1.03	0.92	1.12
25	18	82	20	80	2.0	2.50%	1.03	0.90	1.14
26	13	87	15	85	2.0	2.35%	1.02	0.87	1.18
27	8	92	10	90	2.0	2.22%	1.02	0.80	1.28
28	7	93	9	91	2.0	2.20%	1.02	0.78	1.31
29	6	94	8	92	2.0	2.17%	1.02	0.75	1.36
30	5	95	7	93	2.0	2.15%	1.02	0.71	1.43
31	4	96	6	94	2.0	2.13%	1.02	0.67	1.53
32	3	97	5	95	2.0	2.11%	1.02	0.60	1.70
33	2	98	4	96	2.0	2.08%	1.02	0.50	2.04
34	1	99	3	97	2.0	2.06%	1.02	0.33	3.06
35	0	100	2	98	2.0	2.04%	1.02	NA*	NA*

*Not applicable because formula places a zero in the denominator of the equation.

benchmark population; the absolute percentage difference of 2 percent is paired with a relative percentage difference of 200 percent. Similarly, the disparity index for minorities and ratio of disparity are very high at 3.0 and 3.06, respectively.

Stakeholders need to understand that different measures of disparity can, in some circumstances, lead to different interpretations of the results. With this knowledge they can engage in a discussion with the researcher prior to the production of results regarding some key decisions such as how many measures of disparity will be produced and which one or ones will be selected. Even if stakeholders are not involved in those decisions, this knowledge will help them understand results that include multiple measures of disparity that produce various interpretations. If stakeholders are presented with results conveyed with a single measure they will understand that the results might have been different if the researcher had made an alternative selection.

USING CONTINGENCY TABLES TO IDENTIFY DISPARITY

The relationship, if any, between the race/ethnicity of drivers and various actions by police (such as stops, searches, and dispositions) can be assessed using contingency tables. These tables have a consistent format: the independent variable defines the columns, and the dependent variable defines the rows. Table 7.5 portrays hypothetical search data in contingency table format. The independent variable is the race/ethnicity of the driver, and the dependent variable is

whether or not a search was conducted.⁴ Column percentages sum to 100 percent, and the table is read across. Searches were conducted of 17.61 percent of the stopped African Americans, 11.58 percent of the stopped Hispanics, and 7.00 percent of the stopped Caucasians.

These results indicate that African Americans were more likely than Hispanics and Caucasians to be searched. But what does this finding mean? It means only that a disparity exists. Researchers cannot conclude that bias influenced search decisions because other factors could have caused the disparity.

Researchers can use statistical programs to assess the strength of a relationship that is indicated by the data in a contingency table. Importantly, measures of association (and tests of statistical significance) provide information regarding *disparity*, not bias. For instance, if we had found a strong association indicating that African Americans were disproportionately represented among drivers searched, we would know only that a disparity exists, not *why* it exists. We cannot conclude that bias influenced search decisions because other factors could have caused that disparity.

USING MULTIVARIATE ANALYSIS TO IDENTIFY DISPARITY

Multivariate analysis examines the impact of multiple factors (independent variables) on an outcome (the dependent variable).⁵ It can provide a more thorough and accurate interpretation of vehicle stop data than bivariate analysis.

⁴ A dependent variable is the outcome variable, or the subject of the analysis. An independent variable is the predictor variable that is hypothesized to cause changes in the dependent variable. In this example, we are testing whether race/ethnicity of a driver (the independent variable) will impact on whether or not a search is conducted (the dependent variable).

⁵ In bivariate analysis, researchers look at the relationship between two variables. In multivariate analysis, multiple variables are taken into consideration, and the strength of the relationship between each independent variable and the dependent variable is determined while controlling for the impact of the other variables in the equation.

Table 7.5. Contingency Table to Assess Relationship Between Driver Race/Ethnicity and Police Searches

Driver Race/Ethnicity				
Search Activity	African Americans	Hispanics	Caucasians	Total
No Search	7,249 82.39%	5,919 88.42%	61,184 93.00%	74,352 91.48%
Search	1,549 17.61%	775 11.58%	4,605 7.00%	6,929 8.52%
Total	8,798 100.00%	6,694 100.00%	65,789 100.00%	81,281 100.00%

Note: The Contingency Coefficient is 0.121.

For example, Smith et al. (2003) and Tomaskovic-Devey, Wright, and Czaja (2003) analyzed information from a survey of drivers in North Carolina, including information on the extent to which the drivers were stopped by police. These researchers wanted to find out whether the driver race/ethnicity affected the extent to which people were stopped. The frequency of being stopped during the reference period was the dependent variable. A bivariate analysis with these data would look at the relationship between the race/ethnicity of the survey respondents and the number of stops by police they reported. Researchers would not know from this bivariate analysis, however, if variables like driving quantity, quality, or location had affected the stopping decisions by police. Researchers could show whether disparity existed (for instance, they might find that minorities were stopped more than Caucasians), but they would not know if race—or alternative, legitimate factors—produced that disparity. If a survey data set on stopped drivers in a jurisdiction included information on driving quantity, quality, and location (and the North Carolina survey did), researchers conducting multivariate analysis could look at the effect of race on the frequency of being stopped, controlling for these other factors.

The Key Limitation of Multivariate Analysis

Multivariate analysis is an important tool for social science and can have value for an examination of racial bias in policing. It does not, however, overcome the challenges associated with analyzing vehicle stop data—particularly those challenges associated with identifying and measuring the alternative legitimate factors that can influence police decision making. Multivariate analysis is based on certain assumptions, and a key one is “no specification error.” This is a fancy phrase used by statisticians to reference a key theme of this book: for a method to be most effective it must take into consideration all of the alternative legitimate factors that might have an impact on police behavior. For multivariate analysis to be effective in determining whether driver race/ethnicity has a causal impact on police behavior, it must include independent variables that reflect the alternative legitimate factors that affect police behavior.

A researcher might find a significant relationship between independent variable X and dependent variable Y that would disappear if the researcher had included variable C in the model. A simple example illustrates this point. Let us imagine that a researcher finds

a significant positive relationship between the consumption of high-grade coffee and the square footage of homes. Subjects who drink high-grade coffee, the researcher finds, are more likely to live in large houses. Clearly, drinking high-grade coffee does not *cause* a person to have a large house. The “omitted variable” C, which is wealth, leads to both the drinking of high-grade coffee and the purchase of large houses. Without the independent variable C in the model, the results are misleading: the results indicate a direct relationship where none exists. With wealth in the model, the multivariate methods would indicate a relationship between wealth (not high-grade coffee) and large houses.

Applied to vehicle stops, multivariate analysis can similarly identify a misleading relationship between the dependent variable and the independent variable. It is misleading because the inclusion of a previously omitted variable can make the relationship or correlation disappear. For example, multivariate analysis might find a relationship between race/ethnicity and police dispositions that would have disappeared (as it did in the analysis of the Washington State Patrol data) if the researcher had included number of violations or seriousness of offense(s) as independent variables.

Not including key variables in a multivariate equation can also serve to “mask” racial bias. A researcher may, for instance, find no indication of racial disparity in search decisions—where, in fact, it exists—because the researcher fails to include in the equation crucial independent variables.

For multivariate analysis to be most effective in determining whether driver race/ethnicity has a causal impact on police behavior, it would include all of the independent variables that reflect the alternative legitimate factors that affect police behavior. Quite frequently, however, social scientists cannot identify or measure all of the factors

that they should or would like to include as independent variables. This is not unique to the analysis of vehicle stop data.

Researchers should not lead stakeholders to believe that the use of fancy multivariate statistical techniques, however beneficial, overcomes all the challenges associated with analyzing vehicle stop data. The researcher should make explicit reference to the potentially relevant variables that were not included in the equation and report that these omissions could have had an impact on the results.

WHEN DOES DISPARITY MEAN BIAS?

Isolating the causes of disparity presents a formidable challenge for researchers. An identified “amount” of disparity in stopping behavior by police could be caused by any of the following: bias on the part of police; demographic variations in the quantity, quality, and location of driving; demographic variations in other legitimate factors that have an impact on police behavior; and/or other measurement error. Researchers don’t know what proportion of the disparity comes from what source. With strong benchmarking methods, researchers can *reduce* the number of plausible causes, but only in a perfect world where they can control for *all* alternative, legitimate factors and achieve perfect measurement could they equate a disparity measure or measures with police bias. For this reason, there is no agreed upon “bright line” that researchers can set whereby disparity levels above it indicate racial bias and disparity levels below it indicate lack of bias.

Note also that disparity does not indicate bias just because the results are “statistically significant.” Researchers can use tests of statistical significance in their analysis of vehicle stop data for descriptive purposes to show that a finding is robust.⁶ For instance, a researcher might report that the difference

between the representation of minorities in the stop population and among the benchmark population is “statistically significant.” This shows that the numerical differences are worthy of notice. Whether this disparity is caused by bias cannot be discerned by this test.

THE BRIGHT LINE CONTROVERSY

Some researchers argue that conclusions about the existence or absence of racial bias can and should be drawn from disparity measure calculations in order to provide the clarity needed to guide jurisdiction policy and practice. Other researchers claim that any cut-off point is arbitrary—providing a false sense of clarity where none exists. They note that even large amounts of disparity could be wholly explained by nonbias factors.

Important for stakeholders to understand is that there is no mathematical formula that can produce a legitimate cut-off point above which one can say with confidence that disparity is equal to bias. A researcher may have detected disparity between the racial/ethnic profile of drivers stopped by police and the racial/ethnic profile of the benchmark population, but the researcher has no way of knowing how much of this disparity is due to measurement error and unmeasured variables that influence police behavior. The researchers who advocate cut-off points are, in effect, arguing that if the disparity is particularly large, then chances are, the alternative factors cannot explain all of it. Certainly, it is probably safe to say that the larger disparities are more likely than the smaller disparities to encompass many causes, including bias. It is important to note, however, another possibility: a large disparity could be produced

entirely by alternative legitimate factors, and a small disparity could be entirely produced by bias.

For the researchers who choose to select a cut-off point, we suggest they select the cut-off point before analyzing the results *if feasible* and set the cut-off point in conjunction with a police-resident advisory board after educating that board about the challenges of drawing conclusions about police bias from calculations of measures of disparity. Most importantly, researchers need to convey to the consumers of their reports the constraints associated with setting cut-off points.

A researcher might reasonably choose not to select a cut-off point, believing it unwise to select a point above which “a problem” is indicated. The quote below expresses the reasons why one group of researchers decided not to set a cut-off point:

As with other studies, we faced a problem of establishing a “bright line” above which the conclusion is that all departments are engaged in disparate citation practices that constitute racial profiling and below which all departments are not engaged in disparate citation practices. . . . In studies of disparity, regardless of topic area, it is generally inappropriate to conclude that any difference between the studied population and the comparative population automatically constitutes a meaningful disparity or racial bias. Such differences may be the result of real differences or may be a product of sampling or measurement error (Farrell et al. 2004, 15).

These researchers conclude, “How much disparity is acceptable to a community is

6 More often tests of statistical significance are used in research to make inferences about whether the results from a sample can be generalized to the population from which that sample was randomly drawn.

However, most data that are studied to assess the existence of racial bias represent information (gleaned from forms) on *all* police stops made in a jurisdiction, not a random sample.

fundamentally a question that should be addressed by stakeholders and policy makers in each jurisdiction" (Farrell et al. 2004, 16).

THE GOOD NEWS

When grappling with the question of "how much [disparity] is too much," researchers can avail themselves of two important tools. First, they can compare disparities. Such comparisons will help them interpret their data and provide stakeholders with meaningful feedback and guidance. As an example, an agency using internal benchmarking identifies the officers (or units of officers) with the "most disparity" and initiates a review that will determine whether there are explanations other than bias for the disparity (see Chapter 5). Similarly, other benchmarking methods can identify geographic areas within a jurisdiction, agencies within a state, and/or units within a department with the highest levels of disparity. The researcher could rank those subareas, agencies, and departmental units based on measures of disparity. A *descriptive* cut-off point could then be selected. A unit above the designated level is considered only to have high disparity but no interpretation is made as to the cause of that disparity. Such an identification can serve to help policy makers identify the high priority targets for additional review or for change efforts as discussed more fully in the next chapter.

A second tool to help researchers interpret and report findings of disparity is what we call a "qualitative review of quantitative data." By meeting with law enforcement agency personnel and with other stakeholders, researchers can gain insight and perspective on the quantitative results. These reviews can help ensure that jurisdiction data are correctly and responsibly interpreted. Two reviews are advisable: (1) a review and discussion of the results by researchers and law enforcement agencies, and (2) a review

and discussion of the results by law enforcement personnel and resident stakeholders.

The independent researcher or researcher employed by the law enforcement agency should discuss the results of vehicle stop data analysis with sworn personnel before publishing them. The purpose of this discussion is to gather information from a "street perspective" regarding what the data mean. The purpose is not to "explain away" any disparity that may have been identified but to better understand what factors—legitimate or otherwise—might be producing the results.

The Northeastern University team, in both its Massachusetts (Farrell et al. 2004) and Rhode Island (Farrell et al. 2003) reports, indicates that the ultimate interpretation of the results comes during discussions between police and citizens. One benefit of including residents in discussions of results is the fresh and helpful perspective they bring to understanding what the data mean. Like the police, residents have information about the jurisdiction that can add perspective and context to the numbers produced by the researcher. But discussions between police and residents are about more than how to interpret data. The issue of racially biased policing has, in many communities, exacerbated the "divide" between police and residents, particularly residents who are racial/ethnic minorities. Data collection has the potential to help heal the divide and provide direction for joint reform efforts by police and community members. Police-resident discussions of data become a part of the change process. We discuss in the next chapter how police and residents can come together to use these data for the purposes of reform.

CONCLUSION

In this chapter we discussed four ways to present the results of vehicle stop analyses: absolute percentage differences, relative percentage differences, disparity indexes, and

ratios of disparity. These measures can be used to present results on stops, searches, dispositions, and other types of vehicle stop data. Researchers can choose one or more of these measures to convey the results of their benchmarking analysis to the public. The challenge is not in producing these measures of disparity but in deciding which one or ones to use and present. Some social scientists use just one measure of disparity in their reports to reduce ambiguity and avoid multiple interpretations of results. Others prefer to report multiple measures of disparity. Under some circumstances, the interpretations drawn from one measure might be very different from those drawn from another. Researchers might also develop contingency tables to convey results or use multivariate methods to analyze the data.

This chapter also explained why definitive conclusions about bias cannot be drawn from calculations of measures of disparity. Stakeholders must evaluate the extent to which nonbias factors (factors related to driving quantity, quality, and location) have been addressed by the jurisdiction's benchmarking method. Conclusions about racial/ethnic bias as the *cause* of disparity are suspect because every benchmarking method imperfectly addresses the alternatives to the bias hypothesis.

In the next chapter we describe how police and stakeholders can come together, reflect upon the vehicle stop data analyzed by social science researchers, and identify methods for improving policing practices and the relationships police have with local residents.

VIII

Using the Results for Reform

Vehicle stop data have benefits and constraints as a means of measuring whether policing in a jurisdiction is racially biased. The limits of social science preclude researchers from drawing definitive conclusions from the data regarding the existence or lack of racial bias. Faced with this fact, explained at length in previous chapters, the reader well might ask: of what value are the results if researchers cannot report, with confidence, the existence or lack of racial bias in the jurisdiction?

The answer is that the results of benchmarking analysis 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.¹ In its report on traffic stop data for the state of Rhode Island, the Northeastern University team wrote: “We do not view this analysis as an end of the discussion about the existence and extent of racial profiling in

Rhode Island, but rather it will provide . . . information to begin an important dialogue. . . . [A] well conceived and implemented study of racial disparities in traffic stops can serve as a very useful springboard for community level conversations about the issues of racial profiling” (Farrell et al. 2003, 6).

Below we describe various ways that police and resident stakeholders² can come together to reflect on the results of data collection. The ultimate aim of these meetings is mutual understanding and reform. Specifically, we describe in this chapter

- 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

¹ This should not be construed as an endorsement of mandatory data collection. As indicated in the first PERF publication on the subject (Fridell et al. 2001), there are pros and cons of data collection that a local jurisdiction or state should consider before making a decision regarding whether to collect data.

² In this chapter the term “resident stakeholders” refers to citizens, journalists, advocacy group members, government officials, and others who reside in the community and have a particular stake in the outcome of researchers’ race data analysis.

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, vehicle stop data collection and analysis should never be viewed—either by police or resident stakeholders—as a “pass-fail test” (Farrell 2004). Instead, it should be viewed as a diagnostic tool to help the agency, in concert with concerned residents, set priorities for addressing the problem or perception of racial profiling. The collection and analysis of vehicle stop data can pinpoint geographic sub-areas of a jurisdiction or particular policing procedures that warrant further study. In order to make full use of researchers’ analysis of vehicle stop data, jurisdictions are encouraged to convene a local task force on racial profiling.

THE TASK FORCE AND ITS MEMBERSHIP

In Chapter 3, “Getting Started,” we recommended that jurisdictions create a local racial profiling task force to guide police departments in the development of their data collection system.³ This task force, composed of fifteen to twenty-five people, could plan how data would be collected and analyzed. The task force would bring credibility to the data collection system, and its members would understand both the limits and the potential of vehicle stop data analysis. We recommend including people in the community who are most concerned about racial bias and police personnel representing all departmental levels, particularly patrol.

It is preferable, but not essential, that the task force be convened before data collection begins. If it is formulated after data collection has started, however, it still has an important mission—engaging in constructive dialogue to identify where change is needed. This group should meet and begin its work *before* the report of findings on the vehicle stop data analysis is publicly released.

A group with equal representation of law enforcement personnel and resident stakeholders should review and discuss the data. Nonresident stakeholders also could be included. They could be representatives from state or national groups, such as the American Civil Liberties Union (ACLU), the National Association for the Advancement of Colored People (NAACP), and the Urban League; nonresident commuters to the jurisdiction; and nonresident owners of businesses located in the jurisdiction.

It is usually appropriate for the agency executive to call for and develop this task force. It then serves in an advisory capacity to the executive and makes recommendations that he or she will consider adopting. The agency executive should not be a member of the group since it has been convened to provide him or her with advice on what actions to take. We recommend, however, that the executive attend the task force meetings. By attending the meetings, the executive can convey to task force members, the executive’s staff, and the wider community the importance of the issue. There may be circumstances when another official or group develops the task force rather than the law enforcement agency executive. For instance, a mayor or city council might call for a task force for a jurisdiction or a governor might convene a statewide task force. The executive

³ Because data collection was organized at the state level, the Northeastern University team had a state-level

task force advising it. The team, however, advocates that discussions of the data occur at the local level.

should be a member of the task force if it was not set up and overseen by the executive; the members would make recommendations to the person or organization that developed their group.

The local racial profiling task force should meet on an ongoing basis. For some of the early discussions described below (for instance, on trust-building and on general issues and concerns related to racially biased policing), we advise the use of a trained, neutral (nonpolice, nonstakeholder) facilitator. This facilitator should have experience working with groups on issues that provoke emotions and passions and have knowledge of the topic of racially biased policing. This facilitator might be retained to oversee the long-term work of the task force or, after the early sessions, turn over meeting facilitation to a task force chair or to co-chairs. For the co-chair model, the group may elect, or have appointed, one co-chair who is an internal stakeholder (that is, affiliated with the law enforcement agency) and another who is an external stakeholder. This group may have a finite tenure, or it may become a permanent fixture in the jurisdiction.⁴

THE AGENDA OF THE POLICE-STAKEHOLDER TASK FORCE

The first few sessions of the task force (sessions led by a neutral facilitator, as explained above) should be devoted to developing trust between police and resident members. The task force then would

- discuss general concerns related to racially biased policing,
- review the vehicle stop data,

- review other sources of information about racial bias and perceptions of racial bias, and
- consider possible reforms.

Developing Trust

In a rare situation, a stakeholder group may be able to begin its discussions of racially biased policing at the first meeting; most groups, however, will be well served by engaging in some exercises and discussions on topics other than racial bias before delving into the volatile topic that brings them together. A group in Lowell, Massachusetts—not a task force but a group formed for a one-time discussion—began immediately talking about racial bias. After some finger-pointing, raised voices, accusations by citizens against police, and defensiveness on the part of police, the group turned its attention to developing ways to resolve the particular problems it had identified. On their own, without prompting from the facilitator, the group members agreed that they needed to meet regularly to continue the process of sharing, listening, and resolving problems. Ed Davis, Superintendent of the Lowell Police Department, continued the group as the “Race Relations Council,” which the mayor later described as “the best thing that has happened in Lowell in a long time.”

Although this particular group during a single session was able to move from the heated and angry exchanges at the beginning of the meeting on the controversial issue of race to a sober and rational discussion of a constructive plan of action, most groups cannot. We recommend that task forces engage in activities that will develop trust

⁴ For various reasons, a jurisdiction may be unable (or unwilling) to convene a task force of police and stakeholders. In such circumstances, the department should convene personnel to discuss key topics outlined below,

including general issues related to racially biased policing, the vehicle stop results, other sources of information, and needed reforms.

among members before tackling the challenging topics that define their existence. This trust-building may require a number of meetings.

The Chicago Forums

One trust-building model comes from Chicago. The former superintendent of the Chicago Police Department, Terry Hillard, sponsored a series of forums for police and minority residents of the community. Community activists were recruited to aid the police department in its search for solutions to racial tensions. Department staff of all ranks were also invited to participate. Before the first forum was convened, participants were surveyed for their opinions about racially biased policing and the department's strengths and weaknesses regarding minority outreach. In the survey, respondents also were asked for their ideas on how to improve relations between police and minorities and for their thoughts on how to resolve issues. A facilitator moderated the initial sessions.

During the morning session of the first forum, community members were asked to talk about strengths and weaknesses of their interactions with the police, and police staff were asked to listen and hold their responses until later in the day. Lunch was structured as a mixer, with informal discussions. In the afternoon, police staff shared their thoughts and reactions to the morning session, and residents were instructed to listen and not respond. Then there was an opportunity for discussion. While the issue of racially biased policing was raised by both groups during this first meeting, it was just one of many issues raised. Race issues became a more central focus in subsequent forums and, during those gatherings, the group identified specific actions to be taken by both police and community members to address them. Superintendent Phil Cline who succeeded Hillard has continued these forums.

The Lamberth Workshops

John Lamberth's consulting team uses a two-session workshop to "enhance the trust between law enforcement and the local community" and to develop "collaborative community-based racial profiling solutions" (Clayton 2004). For the first gathering, the Lamberth team holds separate sessions with the police and resident stakeholder participants. The purpose of these separate discussions is to "enhance participants' understanding of the issues" surrounding racially biased policing.

Discussions within these separate groups address the definition of racial profiling, differing perceptions of the issue on the part of law enforcement and resident stakeholders, and the expectations and responsibilities of police and drivers during vehicle stops. By the end of the first session, each group has identified

- safety issues that concern police,
- concerns or fears that drivers might have when stopped by police,
- ways racial profiling harms police-community relations, and
- the group's expectations when making contact with the other group.

During the second session of the workshop, the group composed of police and the group composed of resident stakeholders are brought together for small- and large-group discussions and activities. The police group and the resident stakeholder group review their separate discussions from session one and identify the areas where their expectations and perceptions are shared and where they are different. Together, the first session and first half of the second session serve to initiate constructive dialogue, develop trust between participating police and resident stakeholders, and identify common concerns and expectations. These sessions set the stage for the rest of the workshop during

which the participants develop a plan of action for addressing issues related to racially biased policing and the perceptions of racially biased policing.

Reviewing General Concerns Related to Racially Biased Policing

We have discussed the first item on the agenda of a local racial profiling task force: developing trust. We have described two methods for developing trust and enhancing communication during non-stress times: forums convened in Chicago for police and minority residents and two-session workshops developed by John Lamberth's consulting team. Following trust-building gatherings similar to those we have described, members of the task force should review general concerns and perceptions related to racially biased policing.

To provide structure to the potentially heated conversation, the facilitator might invite resident stakeholders to share their concerns—allowing them to voice their perspective without defensive responses by the police. While the police might feel inclined to “explain away” all the concerns voiced by citizens (and, indeed, there will be incidents described by residents where the police feel strongly—and maybe correctly—that there is a race-neutral explanation), it will ultimately be more valuable for the police to just listen to the residents' concerns. Residents need to be heard on this issue and taken seriously. This discussion also can highlight for police how important it is to deal with *perceptions* that police in the jurisdiction are racially biased. Then the facilitator could ask police on the task force to share their concerns related to accusations or

perceptions that bias is influencing their policing decisions.⁵

Reviewing the Vehicle Stop Results

After a general airing of concerns, the task force should be ready to conduct a qualitative (that is, nonempirical) review of the quantitative data on vehicle stops (see Chapter 7). This review is a continuation of the data analysis. During the researcher's earlier empirical examination of the stop data, all of the factors that might have influenced stopping decisions by police could not have been considered. A “qualitative” review allows for a constructive assessment of the factors, other than bias, that might account in whole or in part for findings of disparity (or lack thereof) between the racial/ethnic profile of the population of stopped drivers and the racial/ethnic profile of the benchmark population. The police and residents who have been brought together on the task force have an important contribution to make. They have valuable knowledge researchers don't have about law enforcement activities and geographic areas in the jurisdiction. Therefore, they can provide a unique and helpful perspective for understanding the empirical results obtained by the researchers.

The goal of the qualitative review of quantitative data is not to determine whether the agency “passed” or “failed” a racial profiling test. As stated earlier, the goal is to identify geographic areas, procedures, and decisions that should get the highest priority when the police department initiates efforts to address community concerns. Even though the quantitative data cannot provide the whole picture or a perfect picture, the data, if carefully interpreted, can direct the task force toward

⁵ The task force should include police leaders at all ranks who are open to exploring the issue of police racial bias and committed to identifying ways of doing

business that can reduce or prevent the problem and perceptions of the problem. These people should be problem solvers and consensus builders.

particular reform targets such as stops of minorities for equipment violations, consent searches of young African American males, or vehicle stops on the “south side” of the city.

Before reviewing the data, members of the task force should become informed about what can and cannot be understood from the analysis of vehicle stop data. They could be encouraged to read this book, which was written to clarify these issues for stakeholders, or they could be educated some other way (perhaps by the researcher) about the meaning of key terms, such as “benchmarking,” “disparity,” and the difference between disparity and bias. Once all members of the group have a good preliminary understanding of vehicle stop analysis, they can review the stop and poststop data. The following questions can help guide this discussion of the data:

- Are there indications of disparity in the stop or poststop results?
- Are there reasons, other than racial bias, that might have led to these disparities?
- For what activities (for example, stops, searches, choice of disposition) is racial bias a possible or probable cause?
- Regardless of whether or not bias is a cause, what is the impact of particular disparities on residents and on relations between police and residents of the jurisdiction?
- Do the costs of certain policing practices that produce disparities—practices that may be race-neutral—outweigh the law enforcement benefits?

Each of these questions will now be examined in greater detail.

An appropriate first question to launch the discussion of the vehicle stop data is “Are there indications of disparity?” The group must keep in mind that indications of “disparity” not “bias” are being discussed.⁶ This conversation about disparity may be shorter than later conversations about the other questions listed. The key is to summarize what disparities were identified by the empirical analyses.

The more interesting, challenging, and longer discussion will focus on the reasons, other than racial bias, that might have led to these disparities. This conversation might start by focusing on each specific finding of disparity (for instance, disparity in stops across racial groups in Area A). Participants might reflect on how the methods used to produce the measure did or did not capture certain important factors. For instance, a resident participant might point out that racial disparity in stops around a stadium that was identified using census benchmarking might reflect the high volume of nonresident, multi-racial/ethnic traffic on game days. An officer might report that the high level of stops in a particular minority area is the result, at least in part, of requests from residents in that area for strong enforcement of the speed limit.

Indeed, the purpose of the discussion of these on-the-ground realities is not to “explain away” disparities but to examine legitimate factors that might account, at least in part, for them. The task force will also consider the possibility that certain identified disparities could be the result of biased decisions by police. The group should consider the possibility that bias has caused disparity

⁶ The group should also be reminded that the same methodological challenges that keep researchers from equating disparity with bias can produce results

showing no disparity when racial bias does, in fact, exist (see Myth 1 in Chapter 2).

if it cannot identify alternative, legitimate explanations for findings of disparity; if there is an accumulation of disparity findings or very large levels of disparity; and/or if a particular police activity is highly discretionary and thus vulnerable to bias.

The results—whether quantitative or qualitative or both—will never lead to definitive conclusions, a key point repeated often in the preceding chapters. Despite these inevitable constraints, the conversation between police and stakeholders and researchers is worthwhile and should continue. The task force is not looking for “proof” of racial bias. (If it is, it will not find it.) Instead the task force is trying to identify priorities for its initial change efforts.

Even if task force members do not view racial bias as the cause for particular identified disparities in vehicle stop data, their deliberations may reveal the need for some changes in police procedures. Law enforcement activities may not be influenced by bias, but they may be detrimental nonetheless. It is constructive for the group to discuss the potential negative impact on the jurisdiction of even (potentially) race-neutral disparities and target efforts to change them.

For instance, data on poststop activity by police may indicate that African Americans are much more likely than Caucasians to be asked to consent to a search; the data also may show that these consent searches are very unproductive (as measured by hit rates). Although it may not be possible to determine whether bias produced this disparity (see Chapter 7), the group may decide to recommend some changes nonetheless. Such a recommendation may make sense if minorities

in the community *perceive* racial bias in these requests by police. This disparity in searches—regardless of whether it is caused by bias—may be too costly in terms of relations between police and minorities. The frustration and anger of minorities may be too high a price to pay for whatever crime control value is derived.

Reviewing Other Sources of Information about Racial Bias and Perceptions of Racial Bias

In addition to vehicle stop data, there are other sources of information that task forces should consider when trying to identify positive steps the jurisdiction can take to address racially biased policing and perceptions of its practice. These alternative sources could include conventional wisdom regarding the types of law enforcement activities that might be most vulnerable to officer biases, surveys of jurisdiction residents to assess their perceptions of policing, and results of focus groups held around the jurisdiction.⁷ The task force also might want to review other sources of data within the department (for example, aggregate data on official complaints against officers, data on the use of force, and arrest data).⁸ Selected tapes from in-car video cameras might be another valuable source of information.

Considering Possible Reforms

The discussions outlined above can strengthen the police-community relationship and promote trust, as well as highlight areas of concern to guide reform efforts. These benefits, however, can be lost if the move from discussing results to discussing

⁷ In some jurisdictions, focus groups of residents might be supplemented by focus groups of nonresidents (for example, business owners and commuters) with a stake in the professional performance of police.

⁸ The department researcher within the Las Vegas Metropolitan Police Department examined force reports. In one analysis, he looked at the race and ethnicity of subjects who were cuffed during a stop and then released with no arrest.

reform is predicated on a forced “confession of guilt” on the part of the law enforcement department.

Following a discussion of the vehicle stop data by the task force, resident stakeholders in the group (including government leaders) may demand a confession of guilt. This is a mistake. A confession of guilt should not be a criterion for moving the discussions forward because vehicle stop data collection/analysis is not a pass-fail test. As conveyed throughout this book, a jurisdiction will not have “proof” of racial bias (or the lack thereof). Moreover, “proof” of racial bias is not a prerequisite for decisions that reforms are worthwhile. All agencies can move closer to the ideal of bias-free policing. Perhaps most importantly, exploring reform without a forced confession of guilt is the most constructive and effective way to proceed.

Police-stakeholder discussions of “racial profiling” that involve finger-pointing by residents and defensiveness by police are not helpful. Discussions when resident stakeholders accuse police of “widespread racism” and of frequently “stopping people *solely* on the basis of race” are not constructive. These types of accusations inevitably lead to defensive responses on the part of police. In a more constructive dialogue, the stakeholders would acknowledge how racial/ethnic bias still is pervasive in their community and how even well-meaning people (including, but not limited to, police officers) might make decisions that manifest bias. The police would acknowledge the concerns of the community and express a willingness to engage concerned citizens in discussions about how to move forward.

Without making a confession, a chief can still acknowledge the need to address the concerns of residents, local officials, policy makers, and other stakeholders. The chiefs might say that, while they cannot prove whether or not their agencies have a problem with racially biased policing, they do know

that some residents have very real concerns and perceptions of a problem that must be taken seriously. The chief could acknowledge that these concerns and perceptions harm the relationship between the police and the racial/ethnic minorities in the community and could welcome a dialogue that leads to positive change.

No agency executive should declare his or her agency “innocent of” or “immune from” racial bias. The many caveats in this book regarding vehicle stop data make clear why such a declaration is unwise. The results of vehicle stop data analysis will never support such a strong statement of innocence and, besides, it’s very unlikely that any agency is without room for improvement on this issue. A statement of innocence would anger constituencies that have strong concerns and perceptions of police bias, and it could significantly undermine police relations with minorities. Furthermore, this chief could never implement reform measures with any degree of acceptance from agency personnel since he or she has previously declared publicly that there is no problem to address.

CHARTING CHANGE INITIATIVES

Having agreed to move forward without a public declaration of guilt or innocence by the law enforcement agency, the local racial profiling task force can begin outlining specific change initiatives. In this endeavor it can use as a guide its discussions of general concerns regarding racial bias, vehicle stop data that may indicate bias and/or deleterious disparity, and other sources of information. The interventions the task force identifies might be specific to a particular “finding,” or they might be of a general nature.

First, we will consider examples of specific findings that could lead to reforms. The task force might find in the data a large number of consent searches of minorities that are unproductive (no contraband or other

evidence is found) or a curiously large proportion of minority stops with unproductive searches and “no action” dispositions. To address the specific problem of many consent searches of minorities that are unproductive, the task force might suggest that the chief adopt an agency policy requiring citizens to sign a consent form before being searched. This consent form would inform residents of their right to refuse. Alternatively or additionally, the task force might suggest that the chief implement a minimum “level of proof” for consent searches, such as reasonable suspicion.⁹ In response to the finding of a large proportion of minority stops with unproductive searches and “no action” dispositions, the task force might suggest that the agency executive revise policies or retrain officers to ensure that stops are made only for legitimate reasons. The chief could establish means for commending officers whose searches are the most productive (as measured by their hit rates).¹⁰ To reduce questionable stops, the task force might suggest that the agency adopt a policy that prohibits pretext stops.

These are a few specific changes that could be recommended. Broader initiatives are outlined in *Racially Biased Policing: A Principled Response* (Fridell et al. 2001). In that book the authors argue that all agencies—whether they have collected vehicle stop data or not—should consider reforms in the following areas:

- Supervision/accountability,
- Policies,
- Recruitment and hiring,

- Education and training, and
- Minority community outreach.

Community members should be full partners in implementing the solutions. For instance, residents could help develop the agency’s policy on antibiased policing, assist with efforts to recruit minority officers, participate in the development of a recruit or in-service training curriculum, support agency outreach efforts to racially diverse communities, or identify external funds for the purchase of equipment and software that might promote good policing practices and greater transparency of police decision making.

Specific findings or general conclusions based on the vehicle stop data might prompt the task force to recommend the collection of more information by the jurisdiction. For example, an agency that conducted analyses of the jurisdiction as a whole might choose to conduct subarea analyses to determine whether there are particular geographic areas where disparities are very high. An agency that used a relatively weak benchmark and found areas with large disparities might implement a stronger benchmark in the identified areas. An agency that compared its vehicle stop data to an “external benchmark” (for instance, agencies using observation benchmarking, benchmarking with adjusted census data, or benchmarking with blind versus not-blind enforcement mechanisms) might choose to implement internal benchmarking (see Chapter 5). The agency then could identify the particular officers who

⁹ The reforms in this example were implemented by Chief Stanley Knee in Austin after vehicle stop data showed that greater proportions of minorities than Caucasians were subject to consent searches. The consent searches of minorities were not very productive, and resident stakeholders perceived that racial bias was the cause of this identified disparity. The chief implemented a consent form and a policy requiring reasonable

suspicion on the part of the officer prior to requesting consent to search. He set a goal of decreasing consent searches by 40 percent over two years; within one year he reported a 63 percent decrease (2,141 consent searches in 2003; 804 in 2004).

¹⁰ Hit rates should not be examined in isolation, but rather within the context of other performance or productivity measures.

produced the disparity so that their policing decisions could be subject to further review. Alternatively or additionally, an agency might decide that additional data elements need to be included on its forms for recording police-citizen contacts. With more information it then could further explore a potential problem area (for instance, consent searches).

To better understand some aspect of the data, an agency might choose to conduct focus groups of officers, a community survey of perceptions of racially biased policing, or a consumer survey (for instance, a survey of drivers stopped by police). All of these initiatives would help the agency to obtain positive and negative feedback regarding community members' interactions with officers. However, the police and resident stakeholders on the task force should not emphasize data collection and measurement to such an extent that the most important work—implementing change—is neglected or postponed.

CONCLUSION

This book has set forth both the benefits and the limits associated with the use of vehicle stop data to measure whether policing in a jurisdiction is racially biased. “Benchmarking” is the method of analysis used to make this measurement, and, as noted in Chapter 2, benchmarking presents a real challenge for researchers because they must consider the following four alternatives to the bias hypothesis when analyzing data on drivers stopped by police:

- *Racial/ethnic groups are not equally represented as residents in the jurisdiction.*
- *Racial/ethnic groups are not equally represented as drivers on jurisdiction roads.*
- *Racial/ethnic groups are not equivalent in the nature and extent of their traffic law-violating behavior.*

- *Racial/ethnic groups are not equally represented as drivers on roads where stopping activity by police is high.*

Researchers must similarly consider alternatives to the bias hypotheses when analyzing search, disposition, and other poststop data. Identifying and ruling out the “alternative, legitimate factors” that can influence police decisions concerning stops, searches, or dispositions is a complex and painstaking task. Nevertheless, many departments have taken on the challenge of data collection and analysis. *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* (Fridell 2004) was written to guide researchers—inside and outside of departments—in this endeavor. This book summarizes that information for the non-researcher stakeholder.

We expect that some frustration will be generated by our message that data collection cannot provide unequivocal answers to questions about the existence of racial bias by police in a jurisdiction. Despite the sincerity of most people posing the questions, answers that are definitive cannot be offered. Data analysis is not as easy as comparing stop data to jurisdiction-level census data, although police departments and concerned residents may well wish it were. We hope, however, that the frustrations that may be experienced are offset somewhat by concrete and useful advice. This book (and its companion, *By the Numbers*) provides previously lacking information concerning how data can be analyzed and the results reported responsibly. We also hope frustrations are offset by the knowledge that even equivocal data can provide guidance for useful changes in a jurisdiction. A key value of these data is their potential to bring police and residents of the community together around a table to identify what might be done to make progress in the jurisdiction on the issues of racially biased policing and the perceptions of its practice.

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.
- Ayres, Ian. 2001. *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination*. Chicago: University of Chicago Press.
- . 2002. Outcome Tests of Racial Disparities in Police Practices. *Justice Research and Policy*, 4: 131–142.
- Becker, Gary S. 1993. *The Economics of Discrimination*. Chicago: University of Chicago Press.
- Clayton, Jerry. 2004. *Enhancing Law Enforcement and Community Trust*. PowerPoint presentation at the conference “By the Numbers: How to Analyze Race Data from Vehicle Stops,” sponsored by PERF and the Office of Community Oriented Policing Services, Las Vegas, NV, July 13–14.
- Cordner, Gary, Brian Williams, and Alfredo Velasco. 2002. *Vehicle Stops in San Diego: 2001*. Report submitted to the San Diego Police Department.
- Cordner, Gary, Brian Williams, and Maria Zuniga. 2001. *Vehicle Stops for the Year 2000: Annual Report*. Report submitted to the San Diego Police Department.
- Council on Crime and Justice and Institute on Race and Poverty. 2003. *The Minnesota Racial Profiling Study*. A report submitted to the Minnesota Legislature. Available on the PERF website at <http://www.policeforum.org>. (On the PERF website, enter as a guest, click “Areas of Interest,” “Racially Biased Policing,” “Supplemental Resources,” “Collecting Data,” and “Jurisdiction Reports.”)
- Edwards, Terry D., Elizabeth L. Grossi, Genaro F. Vito, and Angela D. West. 2002a. *Traffic Stop Practices of the Louisville Police Department: January 15 – December 31, 2001*. Report submitted to the Louisville Division of Police.
- . 2002b. *Traffic Stop Practices of the Iowa City Police Department: April 1 – December 31, 2001*. Report submitted to the Iowa City Police Department.
- Engel, Robin Shepard, Jennifer M. Calnon, and Joshua R. Dutil. 2003. *Project on Police-Citizen Contacts, Six-Month Report*. Report prepared for the Office of the Commissioner of the Pennsylvania State Police by the Population Research Institute of the Pennsylvania State University.
- Engel, Robin, Jennifer M. Calnon, Lin Liu, and Richard Johnson. 2004. *Project on Police-Citizen Contacts, Year 1 Final Report*. Report prepared for the Office of the Commissioner of the Pennsylvania State Police.
- Farrell, Amy. 2004. *Drawing Conclusions from Results*. PowerPoint presentation at the conference “By the Numbers: How to Analyze Race Data from Vehicle Stops,” sponsored by PERF and the Office of Community Oriented Policing Services, Las Vegas, NV, July 13–14.
- Farrell, Amy, Jack McDevitt, Lisa Bailey, Carsten Andresen, and Erica Pierce. 2004. *Massachusetts Racial and Gender Profiling Study, Final Report*. Submitted to the Massachusetts Department of Public Safety. Available on the PERF website at <http://www.policeforum.org>.

- Farrell, Amy, Jack McDevitt, and Michael E. Buerger. 2002. Moving Police and Community Dialogues Forward through Data Collection Task Forces. *Police Quarterly*, 5(3): 359–379.
- Farrell, Amy, Jack McDevitt, Shea Cronin, and Erica Pierce. 2003. *Rhode Island Traffic Stop Statistics Act: Final Report*. Report submitted to the State of Rhode Island by the Northeastern University Institute on Race and Justice, June 30.
- Feest, J. 1968. Compliance with Legal Regulations: Observation of Stop Sign Behavior. *Law and Society Review*, II: 447–461.
- Fridell, Lorie. 2004. *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops*. Washington, D.C.: Police Executive Research Forum.
- Fridell, Lorie, Robert Lunney, Drew Diamond, and Bruce Kubu. 2001. *Racially Biased Policing: A Principled Response*. Washington, D.C.: Police Executive Research Forum.
- Glassbrenner, Donna. 2003. *Safety Belt Use in 2002: Demographic Characteristics* (National Highway Traffic Safety Administration Research Note [DOT HS 809 557]). Washington, D.C.: U.S. Department of Transportation, National Highway Traffic Safety Administration, March.
- Greenwald, Howard P. 2001. *Final Report: Police Vehicle Stops in Sacramento California*. Report to the Sacramento Police Department, October 31. Available at http://www.cityofsacramento.org/spdata/pdf/data_collection_report_2001.pdf.
- Lamberth, John. 1996a. *Revised Statistical Analysis of the Incidence of Police Stops and Arrests of Black Drivers/Travelers on the New Jersey Turnpike between Interchanges 1 and 3 from the Years 1988 through 1991*. Report of defendant's expert in *State v. Pedro Soto*, 734 A.2d 350 (N.J. Super. Ct. Law. Div. 1996).
- . 1996b. Report of plaintiff's expert in *Wilkins v. Maryland State Police et al.*, Civil No. MJG-93-468 (D. Md. 1996).
- Lange, James E., Kenneth O. Blackman, and Mark B. Johnson. 2001. *Speed Violation Survey of the New Jersey Turnpike: Final Report*. Report submitted by the Pacific Institute for Research and Evaluation to the Office of the Attorney General, State of New Jersey, December 13.
- Lovrich, Nicholas, Michael Gaffney, Clay Mosher, Mitchell Pickerill, and Michael R. Smith. 2003. *Washington State Patrol Stop Data Analysis Project: Data Analysis Project Report, June 1, 2003*. A report submitted to the Washington State Patrol by Washington State University. Pullman, WA: The Division of Governmental Studies and Services.
- MVA and Joel Miller. 2000. *Profiling Populations Available for Stops and Searches* (Police Research Series Paper 131). London: Home Office.
- 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>.
- Montgomery County (MD) Department of Police. 2001. *Traffic Stop Data Collection Analysis* (2nd report).
- Novak, Kenneth. 2004. Disparity and Racial Profiling in Traffic Enforcement. *Police Quarterly*, 7(1): 65–96.
- Pisarski, Alan E. 1996. *Commuting in America II: The Second National Report on Commuting Patterns and Trends*. Lansdowne, VA: Eno Transportation Foundation, Inc.

- Ramirez, D., J. McDevitt, and A. Farrell. 2000. *A Resource Guide on Racial Profiling Data Collection Systems: Promising Practices and Lessons Learned*. Washington, D.C.: U.S. Department of Justice.
- Ridgeway, G., K.J. Riley, and J. Grogger. 2004. Analysis of Oakland's Stop and Search Data, in *Promoting Cooperative Strategies to Reduce Racial Profiling: A Technical Guide* (Chapter 9), Oakland Police Department.
- Scales, Robert. 2001. *Racial Profiling: Seattle's Community Involved Approach*. Presentation at the 2001 International Problem Oriented Policing Conference sponsored by the Police Executive Research Forum, San Diego, CA, December.
- Smith, Terry. 2003. *Review, Critique and Recommendations for Improving Racial Profiling Studies*. Document developed by the Service Improvement Analyst of the City of Eugene Police Department.
- 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.
- Thomas, Deborah. 2001. *Preliminary Summary Report: Denver Police Department Contact Card Data, June 1, 2001 through August 31, 2001*. Report provided to the Denver Police Department. Available at <http://admin.denvergov.org/admin/template3/forms/DPDPreliminaryReport3monthNov2001.pdf>.
- . 2002. *First Annual Report, Denver Police Department Contact Card Data Analysis, June 1, 2001 through May 31, 2002*. Report provided to the Denver Police Department, October. Available at <http://admin.denvergov.org/admin/template3/forms/DPDContactCardAnnualReport102902.pdf>.
- 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.
- Tomaskovic-Devey, Donald, Cynthia Pfaff Wright, and Ronald Czaja. 2003. *Self-Reports of Police Speeding Stops by Race: Results from the North Carolina Reverse Record Check Survey*. Unpublished manuscript, Department of Sociology and Anthropology, North Carolina State University.
- U.S. Department of Transportation, Federal Highway Administration. 1995. *Nationwide Personal Transportation Survey* (Microdata Files CD-ROM). Washington, D.C.: Author.
- Walker, Samuel. 2001. Searching for the Denominator: Problems with Police Traffic Stop Data and an Early Warning System Solution. *Justice Research and Policy*, 3(2): 63–95.
- . 2002. *The Citizen's Guide to Interpreting Traffic Stop Data: Unraveling the Racial Profiling Controversy*. Unpublished manuscript.
- . 2003. *Internal Benchmarking for Traffic Stop Data: An Early Intervention System Approach*. Discussion paper available at <http://www.policeforum.org> (On the PERF website, enter as a guest, click "Areas of Interest," "Racially Biased Policing," "Supplemental Resources," "Collecting Data," and "Articles and Commentary.")

Resources

Fridell, Lorie. 2004. *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops*. Washington, D.C.: Police Executive Research Forum. Available online at www.policeforum.org.

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 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 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.

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.