The Metropolitan Police Department's Biased Policing Project: A Review of the Vehicle/Pedestrian Stop Study and Recommendations for Next Steps

Submitted to the District of Columbia's Office of Police Complaints

By
Lorie Fridell, Ph.D.

January 30, 2007

The Metropolitan Police Department's Biased Policing Project: A Review of the Vehicle/Pedestrian Stop Study and Recommendations for Next Steps

Executive Summary

In early 2001, former Chief Charles Ramsey announced that the Metropolitan Police Department (MPD) of Washington, D.C. would address racially biased policing issues. As components of this initiative, the MPD developed a Biased Policing Task Force and commissioned a study of vehicle and pedestrian stops made by police in the District. The District of Columbia's Office of Police Complaints (OPC) was represented on the task force; Lamberth Consulting was selected to conduct the study and a final report for that study was issued by the MPD on December 29, 2006.

The goal of this report to OPC is to assess the study conducted by Lamberth Consulting (hereafter referred to as the "Lamberth Team") and make recommendations with regard to how the MPD in concert with the Biased Policing Task Force can and should proceed in their quest to respond to the national issues of racially biased policing and the perceptions of its practice.

Measuring Racially Biased Policing

The goal of vehicle and pedestrian stop studies, such as the one conducted by the Lamberth Team, is to measure racial bias in policing. As I explain in Section I of this report, social science cannot provide answers to all of the research questions we might want to pose. Relying on social science to determine whether or not police are inappropriately using race/ethnicity in making law enforcement decisions is a challenging endeavor. Social scientists can tell us if patterns exist, and if those patterns are tied to characteristics such as race, but cannot tell whether racial animus or bias are the reasons for disparate patterns.

As I report in Section II, various methods of "benchmarking" have been developed by social scientists who are conducting traffic/pedestrian stop studies. All of these have strengths and weaknesses. The observation technique, which was used by the Lamberth Team, is one of the stronger methods. However, even studies that use the higher quality benchmarks are constrained—by the limits of social science—to definitively identify bias.

In Section III, I comment on the quality of the data provided by MPD to the Lamberth Team and then discuss the Lamberth Team's implementation of the observation method. With regard to the former, the validity and credibility of any study are strongly linked to the quality of the data utilized in it. The MPD put procedures in place to

promote the submission of quality data by officers. The agency also conducted some assessments (or "audits") of data quality–something that is critically important to any study of police stops. Curiously and unfortunately, members of the task force were unable to obtain the results of those assessments/audits. This lack of transparency regarding the quality of the data is disturbing. Without that information, consumers of the report do not know the level or character of officer compliance with the data collection requirements. The bottom line is that we do not know whether the data set that was turned over to the Lamberth Team is of good quality and, indeed, there is some information calling the data quality into question. If the data are not of high quality, the entire study could be invalid.

The Lamberth Team selected a relatively strong benchmark method and the strengths of their implementation include:

- The forms of observation (stationary and mobile) used
- Their process for selecting locations for the observations
- The random selection of days of week and times of day for observations
- The training received by the surveyors
- The focus on proactive stops
- The various ways they "matched the numerator to the denominator" to strengthen their analyses.

One major concern is that the Lamberth Team used inappropriate reference groups in calculating their "odds ratios"—the statistical technique they selected to assess the existence of racial bias. The gist of this criticism is that, instead of comparing the stops of Black drivers to the stops of Caucasian drivers (to see if police treat Blacks differently than they do Caucasians), the Lamberth Team compared Black stops to the stops of Caucasians *and of other minority groups* (including non-Black Hispanics and other minority groups). Similarly, instead of comparing stops of Hispanics to stops of Caucasians, the Lamberth Team compared the stops of Hispanics to the stops of the combined groups of Caucasians, (non-Hispanic) Blacks and other minority groups. The reference groups used by the Lamberth Team are not only theoretically inappropriate, they could have the effect of reducing the likelihood of finding disproportionate stopping of discrete minority groups.

Section IV points out that even a solid implementation of a good benchmarking technique does not provide us with a valid measure of racial bias in policing. This occurs because of the constraints of social science. The MPD study is similar to many others conducted across the country in that it does not incorporate into its assessment key factors that might impact on police decisions to stop. For instance, even though we direct police to make stops based on driving quality/behavior, most studies (including the Lamberth Team's study) do not take driving quality into consideration when assessing the existence of bias. This is comparable to a researcher trying to assess whether teachers are biased in the math grades they give to males and females, without including math performance as a variable in the study.

It is because of the constraints of social science that the Lamberth Team's study, like many others conducted across the country:

- Only assesses disparity/bias as pertains to Blacks and Hispanics and not to other minority groups on the streets and sidewalks of Washington, D.C.;
- Does not assess disparity/bias as pertains to the many activities in which police engage other than pedestrian/vehicle stops;
- Only assesses disparity/bias at 20 discrete locations within the city and not in the whole of the jurisdiction that covers 68 square miles.

Next Steps for the MPD and the Biased Policing Task Force

A key point of sections I through IV is that vehicle/pedestrian stop data will never "prove" or "disprove" racially biased policing. That does not mean, however, that studies such as that conducted by the Lamberth Team cannot be of value. As set forth in Section V, 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.

As articulated by Chief John Timoney (2004) of the Miami Police Department, the reality is that "race is a factor in policing." The current issues involved in "racial profiling" and "racially biased policing" are not new—they are the latest manifestations of a long history of sometimes tense, even volatile, police-minority relations. I contend that *every* police executive needs to consider and address comprehensively the issues of racially biased policing and the perceptions of its practice.

In a comprehensive response to both racially biased policing and the perceptions of its practice the key areas are:

- Policies prohibiting racially biased policing,
- Education and training,
- Supervision and accountability,
- Recruitment and hiring,
- Outreach to diverse communities, and
- Data collection and analysis.

The MPD has already taken significant steps towards its "goal of ensuring effective, bias-free policing in the District of Columbia."[1] These efforts provide a strong basis for forward movement on the critical issues of racially biased policing and perceptions of it.

Importantly, the Biased Policing Task Force should continue to be utilized. This task force would be advisory to the Chief as the department implements state-of-the-art practices in fair and impartial policing. This group could review and discuss issues and concerns related to racially biased policing. They would, in this process, use information in the reports produced for MPD by the Police Foundation and the Lamberth Team.

With regard to a review of the Lamberth Team's results, the task force should discuss the "disparity" findings presented in the report. While data collection cannot prove or disprove racially biased policing it can "help pinpoint the decisions, geographic areas,

[1]This quote is from page 1 of the MPD website-based "Response to the Final Report and Recommendations of the Biased Policing Project" referring to the report completed by the Police Foundation at http://mpdc.dc.gov/mpdc/cwp/view,a,1246,q,555635.asp.

and procedures that should get priority attention" (Fridell 2004, p. 362). Consistent with the recommendations of the Lamberth Team, the task force should explore the factors that might produce the disparity results described below..

The Lamberth Team conducts four key analyses; they assess racial disparity in both vehicle and pedestrian stops and they assess ethnic disparity in both vehicle and pedestrian stops. The team reports that odds ratios between 1.0 and 1.5 are "benign." Odds ratios between 1.5 and 2.0 require "a review of stops in these locations." A value greater than 2.0, according to the Lamberth Team, indicates "potential targeting" of minorities requiring "further action…. from the agency."

In a number of locations studied in Washington, D.C., the Lamberth Team produced "no disparity" results. In the assessment of *racial disparity* in *vehicle* stops in Washington, D.C., five locations of 20 produced odds ratios in the 1.5 to 2.0 range (Lamberth 2006, p. 45). In the assessment of *ethnic disparity* in *vehicle* stops, three of seven locations produced odds ratios in this range (p. 49). In one of the two locations where viable results were produced for assessing *racial disparity* in *pedestrian* stops, the odds ratio was in the 1.5 to 2.0 range. Finally, in the single intersection at which *ethnic bias* was assessed for *pedestrian* stops, the odds ratio was 2.0.

The highest odds ratio was found in the assessment of racial bias in pedestrian stops. At Wisconsin Avenue and M Street NW, which is the main intersection in Georgetown, the odds ratio was 5.8. According to the Lamberth Team's odds ratio formulation, this means that African Americans walking in the vicinity of Wisconsin Avenue and M Street are almost six times as likely to be stopped as non-African Americans. This means that there was a strong finding of disproportionate stops of African Americans in Georgetown. The Lamberth Team refers to these results as "potential targeting." In reflecting upon these disparity results, the task force might ask whether African Americans are at risk for "race out of place" stops in this major tourist/shopping area of the city.

The most important job of the MPD and Biased Policing Task Force would be to facilitate the department's efforts to achieve state-of-the-art practices in fair and impartial policing that are tailored to the city.

My specific recommendations to the MPD are to:

- ✓ Release to interested stakeholders the data used by the Lamberth Team for its research.
- ✓ Release the results of the auditing analyses that the department conducted on these data to assess their quality.
- ✓ Maintain and expand the membership of the Biased Policing Task Force.
- ✓ Charge this task force with advising the chief on the department's continued efforts to ensure fair and impartial policing.

My recommendations to the MPD in concert with the Biased Policing Task Force are to:

- ✓ Review and discuss the findings presented in the Lamberth Team's report.
- ✓ Draw upon other data that pertains to racially biased policing and the perceptions of its practice.

- ✓ Identify what the department needs to do to achieve state-of-the-art practices within the realms of:
    - Policy prohibiting racially biased policing,
    - Education and training,
    - Supervision and accountability,
    - Recruitment and hiring,
    - Outreach to diverse communities, and
    - Data collection and analysis.
- ✓ With regard to data collection, the Chief and task force should decide jointly whether and/or what continued data collection efforts are worthwhile. I recommend as priorities internal benchmarking and hit rate analyses of searches.
- ✓ Implement state-of-the-art practices within the MPD to address racially biased policing and the perceptions of its practice.

The Metropolitan Police Department's Biased Policing Project: A Review of the Vehicle/Pedestrian Stop Study and Recommendations for Next Steps

Submitted to District of Columbia's Office of Police Complaints
By
Lorie Fridell, Ph.D.[2]

Introduction

Background

In March 2001, former Chief Charles Ramsey reported that the Metropolitan Police Department (MPD) in Washington, D.C. would address racially biased policing issues and collect data on traffic stops made by police. This announcement came on the heels of a series of reports in the local media that some MPD officers were exchanging e-mails on their patrol laptops regarding fellow officers and citizens that contained racist, sexist, and homophobic language. MPD formed an advisory group, the Biased Policing Task Force, to guide the department's review of these issues. The task force included representatives of several community and advocacy organizations in the city, as well as staff from the District of Columbia's police accountability agency, the Office of Police Complaints (OPC).

The MPD retained a law enforcement research organization, the Police Foundation, to conduct a "thorough look at the issues that might impact our commitment to provide equitable and unbiased police services to all residents, workers and visitors in the District of Columbia" (news release quoted in Police Foundation, 2004: 1).
In September 2004, the Police Foundation issued its report. Among other findings, the study revealed that a majority of respondents in each racial group—including 72% of the Black respondents and 68% of the Hispanic respondents—believed "police are more likely to stop non-whites for traffic violations" (p. 84). The report included 17 recommendations to address issues related to actual or perceived discrimination by MPD officers. Two of the 17 recommendations pertained to conducting research using police stop data (e.g., data on vehicle and pedestrian stops made by MPD officers) to

[2] My credentials are summarized in Appendix A. Included in that Appendix is a list of my authored and co-authored publications on this topic. Parts of this report draw verbatim from those prior writings, particularly from *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* and from a chapter, co-authored with Michael Scott, entitled "Law Enforcement Agency Responses to Racially Biased Policing and the Perceptions of its Practice." *By the Numbers,* published by the Police Executive Research Forum and funded by the U.S. Department of Justice, Office of Community Oriented Policing Services, is available for downloading from the PERF website at www.policeforum.org (see Free Doc Library and Racially Biased Policing). The contents of *By the Numbers* are summarized in *Understanding Race Data from Vehicle Stops: A Stakeholder's Guide* that is also available on the PERF website.

assess the existence of racially biased policing (RBP).[3] The Police Complaints Board (PCB), which oversees OPC, issued a report and a set of recommendations in January 2002 urging MPD to collect traffic stop data.

Consistent with the recommendations of both the Police Foundation and the PCB, MPD issued a request for proposals from individuals or teams that could conduct a traffic/pedestrian stop study.[4] Lamberth Consulting (herein referred to as the Lamberth Team) was the entity selected to conduct this assessment. The Lamberth Team's study was based on vehicle and pedestrian stop data collected between February 2005 and January 2006 and benchmarking data collected in November and December of 2005. The Lamberth Team submitted a draft report to the department in August 2006 and a final report to MPD in September 2006. The Lamberth Team's final report was released to the public on December 28, 2006.

Document Goals and Organization

There are two major goals of this document. The first goal is to discuss the study that was conducted by the Lamberth Team using the observation methodology. The second goal is to make recommendations to the MPD and the Biased Policing Task Force with regard to how they can and should proceed in their quest to respond to the issues of racially biased policing and the perceptions of its practice.

With regard to the first goal, I want to facilitate stakeholder understanding of the study, its methods and its conclusions. In achieving this objective, I do not merely focus on the report submitted by the Lamberth Team. I want to convey to stakeholders the "bigger picture" that includes understanding (1) the social science challenges associated with police stop studies and (2) how the benchmarking method selected by the Lamberth Team compares to alternative techniques. This broader conceptualization is important for assimilating the study, evaluating its strengths and weaknesses, and drawing conclusions from the results. In this manner, stakeholders (e.g., MPD and other city officials, interest groups, residents) can determine whether drawbacks/weaknesses are attributable to the observation methodology or the way the observation method was implemented by the Lamberth team, or are attributable instead to the social science challenges associated with attempts to measure racially biased policing.

To facilitate this more contextual review, the first three sections address the following topics:

- The challenges of using stop data to measure racially biased policing (Section I)
- Various benchmarking techniques (Section II)
- The Lamberth Team's implementation of the observation method in Washington D.C. (Section III)

[3] Specifically, the Police Foundation report recommended that the MPD "(e)stablish a stop data collection program for officers so that the department can monitor patterns of profiling among various units or by individual officers" and "incorporate data on stops into the Personnel Performance Management system in order to assess racial profiling of individuals or units" (2004: 117).

[4] For this study, MPD received funds from the Department of Justice Office of Community Oriented Policing Services.

In Section IV, I will draw upon the material in sections I through III to discuss the conclusions that can and cannot be drawn from the study conducted for MPD.

Section V will propose recommendations that will facilitate the MPD's objective of responding to the issues of racially biased policing and the perceptions of its practice.

Collecting traffic/pedestrian stop data can be a viable component of a comprehensive response to racially biased policing and, indeed, the MPD should be commended for voluntarily implementing the study conducted by the Lamberth team. A state-of-the-art response to the issues of RBP and the perceptions of its practice, however, does not end at data collection and, as I will argue, implementation of responses should not be dependent upon the results of a study. I will, in Part V, describe what I believe every law enforcement agency should do to address this historically challenging problem. These recommendations draw on and build upon the suggestions presented in the Police Foundation report of 2004 and the Lamberth Team's report.

Section I:
The Challenges of Using Police Stop Data to Measure Racially Biased Policing

Introduction

In response to widespread concerns regarding racial profiling (referred to here as "racially biased policing"), many agencies are collecting data on their vehicle stops; a smaller number of agencies are collecting pedestrian stop data. The agencies collecting data require officers to report information on designated stops. The information collected by officers includes the race/ethnicity of the driver and other information about the stop, such as the reasons for the stop, the disposition of the stop (a citation or warning, for example), whether a search was conducted, and the outcome of the search. Data collection is meant to help administrators and other stakeholders determine whether police decisions to stop drivers and/or pedestrians are influenced by racial bias. In other words, 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.

It is very important for stakeholders to understand what stop data can and cannot tell us.[5] As discussed in more detail below, social science cannot provide answers to all of the research questions we might want to pose. Relying on social science to determine whether or not police are inappropriately using race/ethnicity in making law enforcement decisions is a challenging endeavor. Social scientists can tell us if patterns exist, and if those patterns are tied to characteristics such as race, but cannot tell whether racial animus or bias are the reasons for disparate patterns.

In analyzing the data, researchers have attempted to develop comparison groups to produce "benchmarks" against which to measure 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 demographics of people stopped by police to the demographics of the residential population of the jurisdiction, to the demographics of residents with a driver's license, and to the demographics of people observed driving on jurisdiction roads—to name a few comparison groups.

The variation in quality across benchmarks is great. Findings based on a high-quality benchmark are more legitimate than findings based on a low-quality benchmark, although no benchmark method exists that can prove or disprove the existence of racially biased policing. (This inability to "prove" or "disprove" a relationship is not unique to analyzing driver/pedestrian stop data, but rather is true for social science research generally.) It is not difficult to measure whether there is "disparity" between racial/ethnic groups in stops made by police. All benchmarks can identify disparity. The

[5] I use stakeholders broadly to reference all people in a community who are concerned with the issue of racially biased policing. This would include (but is not limited to) members of the police department, jurisdiction policy makers, formal and informal leaders, and residents.

real difficulty comes in identifying the *causes* for any disparity. The higher quality benchmarks reduce the number of alternative causes that might explain disparity but, as will be conveyed in this report, even studies using strong benchmarks have drawbacks that limit a researcher's ability to measure police racial bias.

The Objective of 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 1 shows a model of the key research question as applied to traffic stop data: 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).

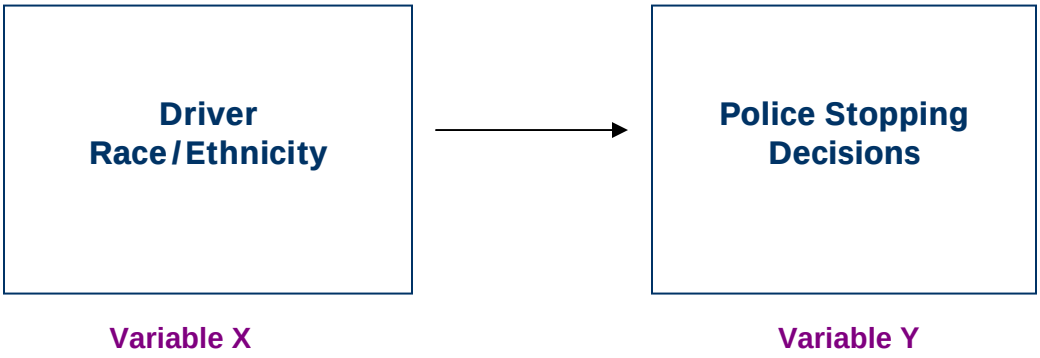

Figure 1. Model of the Research Question: Does Driver Race/Ethnicity Affect Vehicle Stopping Decisions Made by Police?

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). In attempting to test whether X causes Y, we need to rule out alternative explanations that either alone or together cause Y.

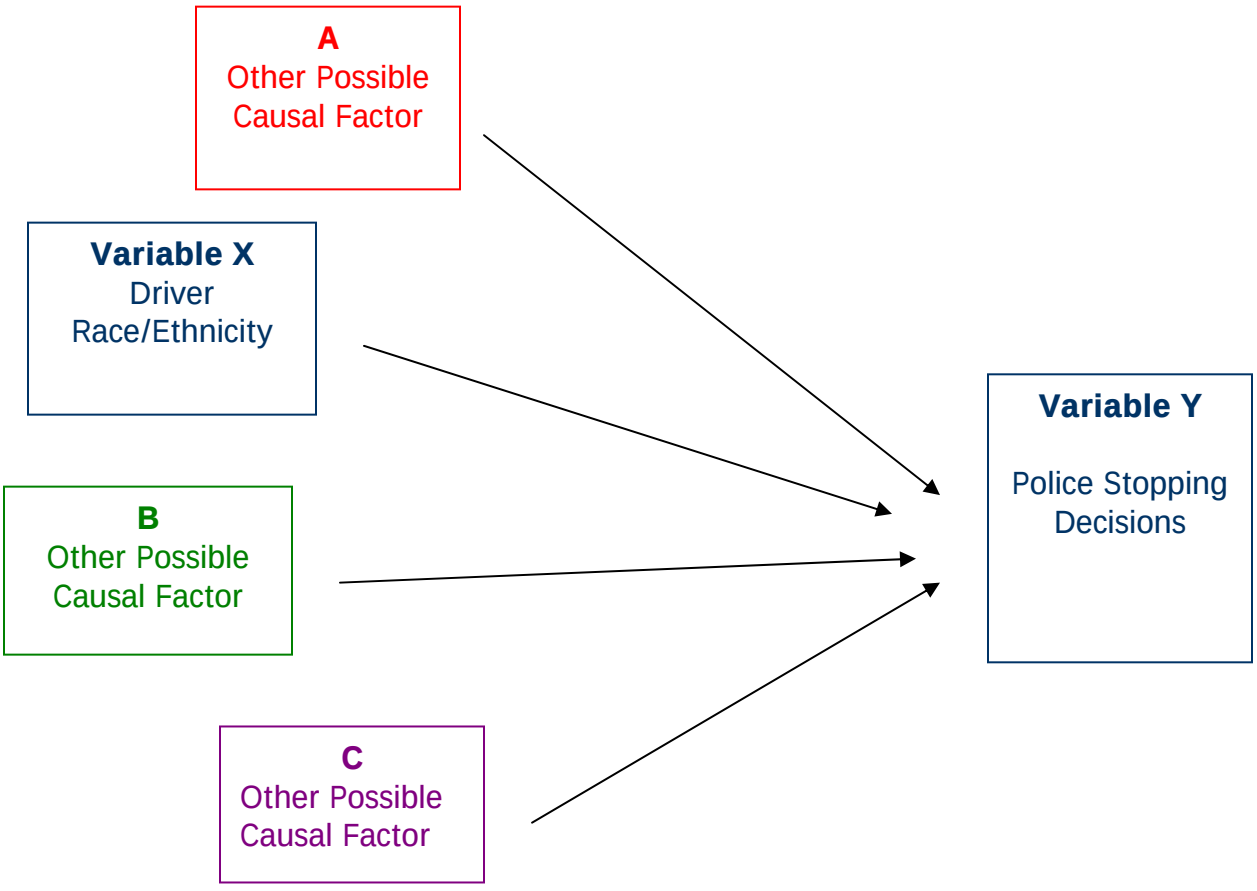

Figure 2. Model of Factors, Other than Bias, that Might Affect Stopping Decisions Made by Police

The following example clarifies why we seek to rule out rival causal factors 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 the same course and draw any conclusions regarding the teachers' gender bias. Instead, we must

consider other factors that could affect grading behavior. In this example, a key variable, of course, would be students' math performance. Our analyses must control for math performance (for example, using results on objectively scored tests) because it is possible that, in our example classroom, boys are objectively better at math than girls. To conduct a legitimate assessment, 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 the teacher.

Now let us return to the 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. Let us 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 seeing these statistics 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 explanations 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 the example above is to see if gender bias is at work. If we could determine the gender breakdown of people who *should be* more at risk of being stopped by police, we could compare it to the gender breakdown 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. This same logic can be applied to studies attempting to measure racial bias in police traffic stop decisions. It is important for researchers, where possible, to control for rival causal factors.

To summarize the key principles stated above, a researcher's goal is to develop a measure of the racial/ethnic demographics of 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. Driving quantity, quality and location are a few of the key factors that can impact on the risk of being stopped by

police. As previewed above and as discussed in more depth later in this report, even studies that use high quality benchmarks are constrained to definitively identify bias by the limits of social science.

Section II:
Various Benchmarking Methods

In this section, I describe briefly the various benchmarking methods that have been used for traffic stop studies (some have been applied to pedestrian stop studies, as well). (See also the discussion in the Lamberth report on pages 10-11.) This provides a backdrop for my discussion, in Part III, of the implementation of the observation methodology by the Lamberth team in Washington, D.C.

Major benchmarking methods include: [6]

- Benchmarking with Adjusted Census Data
- Benchmarking with Division of Motor Vehicle Data (DMV)
- Benchmarking with Data from "Blind" Enforcement Mechanisms
- Benchmarking with Data for Matched Officers or Matched Groups of Officers
- Observation Benchmarking

Benchmarking with Adjusted Census Data

The most commonly used benchmarking method is *unadjusted* census benchmarking. This weak method can be improved if law enforcement agencies modify or "adjust" census data to reflect factors that can legitimately influence police decisions to stop drivers. In traditional census benchmarking, law enforcement agencies compare the demographics of drivers stopped by police to the U.S. Census Bureau demographics of jurisdiction residents or of jurisdiction residents of driving age. A straight comparison between the demographics of these two groups is called "unadjusted" census benchmarking—a method that is not recommended. Various valuable adjustments can be made to produce "adjusted census benchmarking." For example, researchers may adjust the census data on the demographics of residents to take into consideration who, among those residents, owns a vehicle. This adjustment reflects the fact that not every resident owns a vehicle, and people without vehicles are clearly at less risk of being stopped in vehicles by police.

There is quite a bit of variation in the quality of the studies implemented using adjusted census benchmarking. Most of the benchmarks produced with this method, however, are quite a few steps removed from measuring "the people at risk of being stopped by police in an unbiased world." As an example, using residents who own vehicles as the benchmark group does not take into consideration the facts that police stop non-residents as well as residents, people with vehicles drive those vehicles in different amounts, some people violate traffic laws more than others, and so forth.

Benchmarking with Division of Motor Vehicle (DMV) Data

Some researchers have compared the racial/ethnic demographics of licensed drivers who reside in a jurisdiction (using DMV data) to the demographics of the drivers stopped by police. Like adjusting census data for vehicle ownership, this method produces an indirect measure of driving quantity. Benchmarking with DMV data, like benchmarking

[6] These methods are each discussed in depth in *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops* and in less depth in the summary companion guide to *By the Numbers* entitled *Understanding Race Data From Vehicle Stops: A Stakeholder's Guide*.

with adjusted census data that takes into account vehicle ownership, imperfectly assesses who is driving on roads in a jurisdiction.

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 and in different areas of the city.

Benchmarking with Data from "Blind" Enforcement Mechanisms

Law enforcement agencies can use "blind" enforcement mechanisms (red light cameras, radar, air patrols) to produce a benchmark against which they can compare their data on stops by patrol officers. With this method the racial/ethnic demographics of technology-selected drivers (for instance, drivers "caught" by a red light camera) are compared to the racial/ethnic demographics of human-selected drivers (that is, traffic law-violating drivers stopped by police). If officers are as "blind" to race/ethnicity as are the red light cameras, the demographics of the people stopped for red light violations by the officers should match the demographics 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.

When implemented in accordance with the recommendations outlined in *By the Numbers*, the researchers using this method have developed a benchmark group that is relatively close to representing the people at risk of being stopped by police absent bias. 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 for Matched Officers or Matched Groups of Officers

With "Internal Benchmarking" matched officers or matched groups of officers are compared to each other. Specifically, law enforcement agencies can compare stops by individual officers to stops by other peer officers, or they can compare stops by a group of officers to stops by other groups of officers. These comparisons must be made across "matched" sets of officers or groups of officers to control for legitimate rival factors (driving quantity, quality, and location) that increase the likelihood that a driver will be stopped. For instance, an agency might compare the racial/ethnic demographics of people stopped by individual patrol officers who work the same shift in the same precinct. If a particular officer stops proportionately more minority citizens than does his or her matched peers, further exploration of this officer's policing activities and decisions could be warranted.

While a strength of this method of analysis is that it can identify "outliers," or officers that are behaving statistically differently than their peers, it cannot determine whether or not all units used in the comparison (all officers in an officer-level analysis or all groups in a group-level analysis) are practicing biased policing because, in this method, the department is compared to itself. Therefore internal benchmarking methods may not detect bias when all or a majority of officers in an agency are engaged in biased policing.

Observation Benchmarking

The Lamberth Team used the observation methodology for its analysis of the MPD traffic/pedestrian stop data. In the observation method, members of the research team stand at the side of the road and make assessments of drivers' race/ethnicity as they drive by (or stand at a designated location and record pedestrians' race/ethnicity as they walk by). The researchers then compare the racial/ethnic demographics of drivers/pedestrians who are physically observed by researchers at selected sites within the city to the racial/ethnic demographics of drivers/pedestrians stopped by police in the same vicinity. The observation data is used as a benchmark for the stop data.

The observation benchmarking method, if implemented in accordance with solid methodological standards, is one of the strongest methods for controlling for the legitimate factors that affect stopping decisions by police (e.g., driving quantity and location). As discussed in more depth below, researchers give close attention to the following questions:

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

As above, the observation method, when conducted in accordance with standard principles of social science, can produce a strong benchmark representing the people at risk of being stopped by police absent bias. The method controls for driving quantity and driving location. One weakness of this method, however, is our imperfect ability to measure driving quality as part of these studies. As set forth above, people who drive poorly are more at risk of being stopped by police than those who don't. As discussed below, researchers handle this challenge differently; some try to measure selected law violating behaviors; others acknowledge that the potential law violating behaviors are too many to realistically encompass in a study. Another key drawback of the observation method is the fact that it can only produce a "spot check" of police decisions. The techniques researchers use to enhance the rigor of their analyses come at the cost of scope. The "rigorous" assessments produce results that pertain only to certain times and certain places.

Other Analysis Methods/Tools

Other benchmarking methods/resources (discussed in Chapter 10 of *By the Numbers*) include the following:

- Crime data benchmarking,
- Crash (auto accident) data benchmarking,
- Transportation data benchmarking,
- Survey data benchmarking,
- Geographic Information Systems (GIS) resources, and
- Other analytic tools.

By the Numbers also provides guidance for the analysis of poststop data. Whereas the methods described above attempt to detect bias in the decisions of officers to make a vehicle (or pedestrian) stop, poststop analyses attempt to detect bias in the decisions officers make once the stop is made. These decisions include whether or not to search,

whether or not to request consent to search, what the consequences of the stop will be (e.g., arrest, citation, no consequence), and so forth. Since poststop analyses were not a part of the MPD study, these methods are not presented here. I return, however, to the topic of poststop data collection and analysis in Section IV.

As indicated in Section I of this report. Social science faces considerable challenges in trying to measure racially biased policing. This section reports that some benchmarking methods are stronger than others and the observation method is considered among the group of stronger methods by many social scientists working in this arena. In the next section, I discuss the implementation of the observation method by the Lamberth Team in Washington D.C.

Section III:
The Lamberth Consulting Team's Implementation of the Observation Method in Washington D.C.

Data Quality

Before I comment on the Lamberth team's implementation of the observation method, I will discuss the quality of the data received by the team from the MPD. The validity and credibility of any study are strongly linked to the quality of the data utilized in it. Dr. Lamberth recognizes this in his MPD report when he writes "data that is complete, accurate and truthful is critical" (p. 8) and in his statement that "two primary components must be in place to determine whether racial profiling is occurring: benchmarks and *complete stop data*" (p. 10, italics added).

The MPD put procedures in place to promote the submission of quality data by officers. For instance, they trained officers on the data collection project and the forms to be submitted. Additionally, supervisors were instructed to review the forms completed by officers. These are both important ways to promote and facilitate the submission of quality data.

As stop forms start to come in, departments should review the incoming data to ensure that they are, in fact, getting the quality data they seek. In *By the Numbers* Chapter 4, I discuss how agencies can and should "audit" the incoming data from officers for quality. The purpose of these audits is to ascertain whether line personnel in the police department are (1) submitting data collection forms for each and every targeted stop and (2) filling out the forms fully and accurately.

To assess whether or not officers are in fact submitting forms for each and every targeted stop, the data collection forms can be cross-checked with other agency data—for example, citation data and/or computer-aided dispatch (CAD) data. For instance, an agency might compare (1) the total number of police-contact forms that indicated citations were issued to (2) other agency records regarding citations issued. Similarly, police-citizen contact form totals or subtotals can be matched to records of officers' calls to their communications or dispatch center regarding stops. If an agency can count (using, for instance, CAD data) the number of times officers reported they were making a vehicle stop, it can compare these totals to the total number of forms submitted for stops during the same period of time. Again, in these two examples, the researcher is comparing totals of some type of stop across data sets. If, for example, the number of traffic stops according to the CAD data is significantly larger than the number of traffic stop forms submitted by police, the researcher can conclude that forms are not being submitted by officers for all of required stop activities. In more sophisticated auditing systems, agencies attempt to match each citation issued to its corresponding stop form to assess full reporting by officers.

The MPD completed some auditing comparisons to determine whether officers were reliably submitting forms (that is, submitting forms for each and every stop targeted by the data collection project).[7] Curiously and unfortunately, task force members were

[7] Chief Ramsey reports in a letter received by OPC on 11/9/2006 that the Research and Resource Development (RRD) Division "compared the total number of PD 76s, regardless of

unable to obtain the results of those comparisons (i.e., the results of the audit), notwithstanding requests for them and the fact that the auditing results were "reported regularly to the MPD chain of command at the Crime briefings" (Biased Policing Project, Task Force Minutes, 11/1/2005 at page 3). This lack of transparency regarding the quality of the data is disturbing. Without that information, consumers of the report do not know the level or character of officer compliance with the data collection requirements. Without the auditing information from the department, stakeholders cannot know whether or not many more stops were made by MPD police than forms submitted. And if there are a significant number of stops made for which no forms were submitted, one has to wonder (and be concerned about) which stops went unrecorded and how the study results are impacted. I cannot stress enough how important these data quality issues are. If MPD officers chose not to submit forms for a significant number of stops and did so in a manner that was not completely random, the entire study could be invalid.[8] In Section V, I recommend that the department release the methods and results of their auditing processes and make the vehicle/pedestrian stop data available to interested parties.[9]

Regarding the "full and accurate" completion of forms, we did learn that missing data was very low for variables including the all-important race/ethnicity variable. In a letter to the OPC received on 11/9/2006, Chief Ramsey reports that data for the variables Type of Stop, Address, and Race/Ethnicity had less than 2 percent missing data. This is a low amount of missing data.[10]

I turn next to my assessment of the Lamberth Team's implementation of the observation methodology

Most stop studies around the country have focused on only vehicle stops; it is commendable that the MPD and Lamberth Team studied pedestrian stops, as well as vehicle stops, in Washington, D.C.[11]

---

whether a ticket was issued to the entire universe of moving violations." If this process is reported correctly, the match is not a logical one. Arguably the comparison should have been between the total number of PD 76s *submitted for vehicle (not pedestrian) stops that resulted in a ticket* to the alternative data base representing tickets issued.

[8] Highlighting the issue of data quality is the fact that the number of stop forms increased two-fold over the course of the study reference period (that is during February 2005 and January 2006). One interpretation is that during the early months of the reference period only 50% of the stops made by police resulted in the submission of PD 76 forms. A legitimate question is "which 50%?" Because of their concern about the study and accusations of racial bias, were police more likely to turn in forms for Caucasian stops than minority stops? Dr. Lamberth, on page 42 of the report, provides some information to counter this suggestion. Stakeholder access to the data would allow for a more in-depth assessment.

[9] Lamberth, in footnote 12 on page 15, appears to support the release of data from stop studies. He criticizes another team's study because "the data underlying the study are not available to other researchers." (I acknowledge that the release of the data is within only MPD's power, not Dr. Lamberth's.)

[10] Lamberth provides additional information on the amount of missing data in footnote 23 of the report (p. 41).

[11] For lack of a better location, I commend the Lamberth team here for the analyses they present on pages 20 to 24. In this section, they provide data to highlight the weakness of using census data as a benchmark for vehicle or pedestrian stops. This information is important as we

As conveyed above, when implementing the observation method, researchers must address the following methodological issues:

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

I comment below on how the Lamberth Team answered each of these questions. Additionally, I comment on training observers and conducting the analyses.

How Should the Observations be Conducted?

Observations of drivers 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. The Lamberth Team, which has used both types of observation methodologies in previous studies, selected for the MPD study stationary methods for benchmarking traffic and, for all but one location, for benchmarking pedestrians. For the pedestrian benchmark location at the 2100 Block of Alabama Ave. S.E. "surveyors made their observations from a slowly moving vehicle" (p. 32).

What Should be Observed?

Two major issues relative to the question of "what should be observed" pertain to (1) whether the benchmark (produced by observation) should be comprised of drivers or traffic law violators and (2) how a research team should measure/categorize race/ethnicity. Each of these issues is discussed below.

Measuring "Who is Driving" versus "Who is Violating"

Important to the implementation of the observation methodology for studying traffic stops is deciding whether to gather (via observation) demographic (and other) information on people who *are driving* or on people who *are violating traffic laws*. Researchers in this realm have not formed a consensus on which focus is preferable. In *By the Numbers*, I advocate the observation of "who is violating" as opposed to observing merely "who is driving?"[12] Acknowledging valid differences of opinion, I had Dr. Lamberth and his co-authors argue the case for observing "who is driving" as opposed to "who is violating" in Appendix D of that book. I summarize key arguments on both sides below

Earlier in this document**,** I made the claim that driving behavior is an important component of any model that seeks to explain decisions by police to stop drivers. As stated in *By The Numbers* (2004: 17), "police are asked to make driving behavior a key part of these decisions, and therefore we must recognize this variable in our methodology unless we are quite confident that there are no differences across racial/ethnic groups." Excluding driving behavior from the model is equivalent, in my

---

advance and better understand the appropriate methods for analyzing race data from traffic or pedestrian stops.

[12]In Chapter 9 of *By the Numbers* I describe how some researchers have observed "who is violating?"

mind, to excluding math performance from the earlier analysis that tested gender bias in math teachers.[13] I cite research (see *By the Numbers*, pp. 17-22) that challenges the contention that there are "no differences" across racial groups in driving behavior and, indeed, the Lamberth team adds to these findings with the study of speeding behavior that they commendably included as part of the Washington, D.C. project.[14] Lamberth notes that a finding that "one racial/ethnic group were more egregiously violating one or more traffic laws" would be useful for a "determination of whether some group was being stopped more than would be expected." I will return to this important point in Section IV.

There are several arguments in support of observing "who is driving" as opposed to "who is violating." One argument is that all drivers violate traffic laws to some degree or another and, as such, violators and drivers are the same. Lamberth acknowledges this position, but also notes in his report that, while all drivers violate traffic laws, the "question remains as to whether one racial or ethnic group is more likely to violate traffic laws egregiously than another" (p. 14). This is consistent with my proposition above that people who drive poorly are more at risk of being stopped by police.

Another key argument in support of observing "who is driving" as opposed to "who is violating" is the fact that it is arguably impossible to measure comprehensively and with validity all violating behavior. As Dr. Lamberth and his co-authors write in Appendix D of *By the Numbers*, "An important argument against trying to measure 'who is violating' is the fact that there are literally hundreds of traffic violations for which a motorist can be legally stopped" (p. 414). Following the presentation of several examples of measurement challenges, they report, "in the realm of measurement challenges, detecting the vast majority of traffic violations for which a motorist can be stopped [through the observation method] is either not possible or prohibitively time consuming" (p. 415).[15]

Consistent with his prior, extensive work using this methodology, Dr. Lamberth and his team observed "who is driving" in their study for the MPD. To the team's great credit, they also tested the assumption that racial groups violate traffic laws, specifically speeding, at different rates. (See the section in the report entitled "Red Light and Photo Radar Benchmarking," starting on page 24).[16]

---

[13] Some have claimed that this stance is based on the assumption that minorities violate traffic laws more. I address this claim in *By the Numbers* (2004: 18): "Concerned stakeholders … have asked (me) whether the unstated implication (of my position) 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 the police.) If minorities do violate less, then it is important 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."

[14] As reported on pages 54 and 55, the Lamberth team found that Black motorists "are somewhat less likely to violate the speed laws [at the locations tested] by 11 mph and above than would be expected based upon their presence in the transient population" (p. 55). They draw the same conclusions from the Hispanic data, but also report that there may be too few observations to provide for proper analysis (p. 55).

[15] Lamberth highlights this measurement challenge in footnote 3 on page 5 of the MPD report.

[16] To clarify, the Lamberth Team's study-within-a-study of speeding behavior was distinct from the creation of their benchmark which was based on observations of "who is driving," not "who is violating."

Because "reasonable social scientist minds" can and do differ on this issue, I do not criticize the Lamberth team for the decision to observe "who is driving" instead of "who is violating." I do, however, return to this issue in Section IV where I discuss how the constraints of social science (in this instance, our inability to measure driving quality) limit our ability to measure racial bias.

Measuring Demographics

Appropriately, the Lamberth team relied on their observers' perceptions to assess the race/ethnicity of drivers who were passing through the intersections under study. 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 the perceptions of officers is the preferred method for identifying race/ethnicity for the police stop data, the perceptions of trained observers is equally viable as the method for obtaining the benchmark data.

It is important to recognize, however, how difficult it is 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. It can be particularly problematic, report researchers (e.g., Alpert Group 2003; Smith et al. 2003) to identify *ethnicity* through observation. It is also difficult for observers to distinguish among, for instance, Middle Easterners, Hispanics, and Native Americans.

Because of the difficulty of perceiving accurately a driver's race/ethnicity, some of the researchers implementing observation benchmarking use at least two, and sometimes three observers. The pair (or group) of observers records demographic data for the same drivers; the researchers track and report the inter-rater reliability of the observers' findings. The Lamberth Team relied on the observations of a single observer for producing their intersection benchmarking data. While teams of two were assigned to the same intersection at the same time, these individuals observed different lanes (p. 31). The observers or "surveyors" were assessed during training for inter-rater reliability and the coefficient was greater than .80 (footnote 20 on page 31).[17]

Because of this problem of discerning the demographic characteristics of drivers, researchers have broadened categories of race/ethnicity for purposes of observation to more closely match what observers can see. Thus, instead of asking observers to distinguish between, say, seven categories of race/ethnicity, often researchers will have them distinguish among broad categories such as "black and non-black" (Alpert Group 2003) or Caucasian, Black, Other.[18]

[17] A "greater than" .80 rating means that the two observers being compared could report different races and/or ethnicities in up to 20 percent of the observations.

[18] Small numbers within racial/ethnic categories is another reason for collapsing or reducing these categories for the research. See *By the Numbers* page 176-177.

Presumably, this is the rationale for the Lamberth Team analyzing the Washington D.C. data to assess disparity for stops of Blacks and Hispanics and not other racial groups.[19] Again, the constraints of social science support this decision, but again we see the constraints of social science that reduce our ability to measure comprehensively the existence of racial bias. I return to this issue in Section IV where I discuss the social science challenges associated with measuring racial bias.

What Locations Should be Selected for the Assessment

Researchers must determine the number and geographic locations of the sites within a jurisdiction that will be assessed for racial bias. With regard to the latter (geographic location), researchers must decide whether to select the locations in a random or purposive manner.

A key consideration in selecting sites is the ability of the observers to discern the race/ethnicity of drivers. Relevant factors include the speed of the vehicles and lighting at the site. Lamberth has advocated the use of controlled intersections (intersections with a stop sign or light) because they have advantages relevant to these variables. Traffic is generally slower at controlled intersections than at noncontrolled intersections and nonintersections (straight-aways). Intersections are also more likely than straight-aways to have light sources and safe locations for observer placement. For these reasons, the Lamberth Team legitimately chose to use controlled intersections for observation sites in Washington, D.C.

In making site selections, the researcher should also consider volume of activity. That is, the researcher should strive to select sites that have sufficient numbers of both police stops and cars/pedestrians passing by the sites to be observed to produce reliable results. Lamberth selected the Washington, D.C. sites in part on the basis of volume. The considerable attention by the Lamberth Team to site selection is reported on page 29.

As above, a researcher consideration is whether to select sites randomly or purposively. In random selection of sites for observation benchmarking, the researcher identifies the locations in a jurisdiction that meet certain criteria and then randomly selects from among those locations. In purposive selection, the researcher will rely completely on specific criteria to select the sites and will not use random selection during this process. The selection criteria are devised to meet particular objectives, such as visibility on the part of observers, sufficient volume of activity, geographic diversity, heterogeneity across racial makeup, heterogeneity across traffic types, and so forth.

The advantage of random selection is that it allows the researcher to generalize. In other words, the results from the selected sites can be applied to other locations in the city that meet the criteria for inclusion in the pool from which the sites were selected. Sometimes, however, a researcher is unable to use random selection or it is inadvisable to do so. Purposive selection is quite acceptable so long as the researcher does not, in conveying his/her results, generalize beyond the specific sites studied. The Lamberth team used purposive selection. Appropriately, the Team does not attempt to generalize

[19] The PD 76 used to collect data includes the following race/ethnicity categories: American/Indian/Alaskan Native, Asian, Arab American, Black/African American, Hispanic/Latino, White, and Other.

their results beyond the 20 intersections studied. In Section IV, I return to the constraint mentioned above that implementing with rigor the observation methodology comes at the cost of the scope of the assessment. In this case, the assessment is of 20 intersections in the city of Washington, D.C. and the results will not allow stakeholders to draw conclusions about police stops generally throughout the city.

When Should the Observations be Conducted?

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 regarding 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.

In selecting days of the week for scheduling observations, researchers strive for representativeness in the nature and extent of traffic behavior. Randomly selecting days, as the Lamberth team did, is a good way to achieve this. Similarly, the times of day for collecting observations should reflect the goal of representativeness. Again, the team's random selection of times of day achieves this goal. They conducted observations around the clock (i.e., including during the night).[20]

The "reference period" refers to the time period during which the observation data are collected. In some jurisdictions the nature and extent of traffic vary during different times of the year. This seasonable variation will affect the population of drivers on the roadways and thus the racial/ethnic profile of drivers will vary by seasons. Other jurisdictions may have no seasonal variation in their traffic. Researchers can handle suspected seasonal variation in several ways. One option is to conduct the analyses using a full year of data; that is, the researcher would conduct observations at various points throughout a twelve-month period. A twelve-month reference period, however, may not be economically feasible or politically viable.

There are several ways a research can justify or handle a reference period of less than 12 months. A research team could include in their analyses only the police stop data that occurred during the same months that the observation data were collected. If observations for benchmarking are collected during January through March of a particular year, the analyses would include only the police stops that occurred during the same months. The researcher would include a caveat in the report indicating that the results of the study pertain only to police activity during those months and do not necessarily generalize to the entire year. Researchers might also make the case that traffic in the jurisdiction under study has no seasonal variation in driving.

The Lamberth Team collected observation data during November and December of 2005 and used this information to benchmark stops made during February 2005 through January 2006. More specifically, they collected traffic observation data during 14 days in November (page 38 of the Lamberth Team's report) and pedestrian observation data during 14 days in December (p. 39). Intersections were observed for 30 to 60 minutes total (pp. 34-35). What this means is that 60 minutes of traffic observations made during several days in November at a particular location were used to benchmark 12 months of

[20] Some researchers who use this method do not advocate nighttime observations because of the difficulty in discerning the demographics of drivers.

traffic stops at that same location. It seems quite unlikely that the vehicles traveling through that intersection during those 60 minutes are representative of the vehicles that travel through that intersection year-round. At the very least, the Lamberth Team should have conveyed to the consumers of their report why they feel justified generalizing the observation data collected during these 60 minutes to the rest of the year.

Training Observers

Researchers who conduct observation benchmarking must train the individuals who will be making the observations in order to ensure that they collect data that are as accurate as possible (enhancing validity) and do so in a manner similar to each other (enhancing reliability). The Lamberth Team describes the classroom and field training on pages 30 and 31 of the report. The field training is particularly important. Under supervision, the observers conduct pilot observations—recording their findings on data sheets. The trainers provide feedback to observers about their performance throughout the training course.

Conducting the Analysis

Once the police stop data are accessed from the department and the observation data are collected in the field, researchers conduct the analyses. Below I discuss why researchers might analyze subsets of data, the concept of "matching the numerator to the denominator," and calculations for measuring disparity. I also address the question, "When does disparity mean bias?" And I address the types of conclusions that you can legitimately draw from such traffic stop analysis procedures.

Conducting Analyses on Subsets of Data

There are several reasons why a researcher might conduct his/her analyses on subsets of their police-citizen contact data. Some of the legitimate subset selection decisions made by the Lamberth Team were based on the distinction between proactive and reactive stops. In analyzing police-citizen contact data, researchers are attempting 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 might, and indeed should, affect their law enforcement behavior. Therefore, analyses of "who is stopped" should legitimately focus on those incidents where officers have discretion in making this decision (proactive stops) and exclude stops where officers have little choice in the matter (reactive stops). Consistent with this concept, the Lamberth Team excluded from their analyses, the PD 76 forms for stops made at vehicle checkpoints at which officers do not have discretion regarding whom to stop (p. 12).

Matching the Numerator and the Denominator

Also important to various aspects of the analyses is the concept of "matching the numerator and the denominator." Social scientists analyzing police-contact data use "numerator" in reference to the data collected on stops by the police and "denominator" in reference to the data collected to produce the benchmark. To "match the numerator and the denominator" means the researcher should adjust the stop data to correspond to any limiting parameters of the benchmark and vice versa.

For example, a researcher might benchmark vehicle stops in a jurisdiction against census data for jurisdiction residents who own a vehicle. That is, this researcher would be conducting "adjusted census benchmarking" (Chapter 5 of *By the Numbers*) and, in so doing, would compare the racial/ethnic demographics of the people stopped by police to the racial/ethnic demographics of people who live in the jurisdiction who have access to vehicles as measured by 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 *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.

In the above example, there was an inherent limitation on the denominator. In this next census benchmarking example, the inherent limitation is on the numerator. Only people of driving age will be included in the numerator. With very few exceptions (we hope so few we can reasonably ignore them), the drivers stopped by police will be of legal driving age (in many jurisdictions it is 15 or 16 years of age and older). Because only people of driving age will be represented in the numerator, the researcher must also limit the denominator data to people of driving age. Thus, in the example of census benchmarking, the researcher will not calculate race/ethnicity of all residents of the jurisdiction, but only of those residents who are of driving age (for example, age 15 and older).

"Matching the numerator and denominator" applies to other benchmarks including observation benchmarking. As with other methods, matching reduces the scope of the analysis but increases its strength.

Researchers using the observation method must match the stop data and observation data geographically. It is inappropriate to compare stops that occur in one side of the city with observations made on the other side of the city. To match data geographically, the researcher conceives of a radius around each observation site and uses for the analysis of that site only the police stops that occurred within the radius. There is no fixed, standard size that is appropriate in all jurisdictions. The researcher should select a radius around the site that she or he believes contains the same type of traffic—in terms of density, nature (residential or commercial), and driver (or pedestrian) demographics—as that going directly by the site under observation.

The Lamberth Team explains that they used for their analyses the police stops "within a quarter of a square mile in each direction (generally about three blocks) with the named intersection being the center of the polygon" (p. 45). Impressively, the Lamberth Team reduced the polygon sizes at two locations (the intersections at 17th Street NW and Euclid St. NW and at Wisconsin Ave. NW and M St NW) for purposes of analyzing the pedestrian stops because of changes in the nature of the areas. Thus, for instance, the team reports that they revised the analysis perimeter at Wisconsin and M "because the area usage changes from commercial to residential within a block or two from that intersection" (p. 52).

As above, the Lamberth Team did not match the stops and observations across time of the year; they used observations from (two sets of)13 days in winter to benchmark stops

conducted year round. Since their observations for purposes of benchmarking the vehicle stops were conducted around the clock, no matching was required for time of day (that is, they appropriately used stops made round the clock as well). They limited their observations of pedestrians to the period 5 p.m. to 1 a.m. because "the majority of stops of pedestrians occurred in the late afternoon until the early morning hours" and so appropriately included in their analyses only the stops made by police during that same time of day.

Calculating Measures of Disparity

As discussed in depth in *By the Numbers* Chapter 12, disparity can be conveyed numerically in various ways. Disparity can be conveyed through absolute percentage differences, relative percentage differences, disparity indexes, and/or ratios of disparity (also called "odds ratios"). The challenge for researchers is 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. This is an important decision because, as shown numerically in Chapter 12 of *By the Numbers*, the conclusions drawn from one measure might be very different from those drawn from another.

The Lamberth Team chose to use "odds ratios" to assess disparity. I convey here how an "odds ratio" is calculated so that I can make an important point below about the particular equations used by the Lamberth Team. To produce the odds ratio one first calculates a "disparity index" for each racial group. As an example of this calculation, for African Americans, the researcher would divide the African American percentage among drivers stopped by the African American percentage among the benchmark population. Thus, for instance, from Table 7 of the Lamberth Team's report, to produce the disparity index for African Americans at the first location (4th St E and Chesapeake SE), the researcher would divide 94.8 by 95.6. Using this formula, we produce a disparity index of .99 for African Americans at this intersection. Next the researcher calculates the disparity index for the other racial group or groups for that same location. Thus, for instance, the researcher would use the same calculations to produce a disparity index for Caucasians at the same location. Finally, to produce the odds ratios, the disparity index for one group (e.g., African Americans) is divided by the disparity index for another group (e.g., Caucasians). The group in the denominator is the "reference group" to which the first group is compared. So the researcher could divide the disparity index of African Americans by the disparity index for Caucasians to gauge how African Americans fare (in traffic and/or pedestrian stops) relative to Caucasians.

The selection of the reference group is very important. In *By the Numbers*, I argue (pp. 316-318) that the most relevant reference group (albeit among imperfect choices) in any calculation of an odds ratio for vehicle (or pedestrian) stop analysis is 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?"

The Lamberth Team did not use Caucasians as their reference group in assessing bias against African Americans and Hispanics in Washington, D.C. Instead, for assessing racial bias—first for Blacks and then for Hispanics—the Lamberth Team used "everyone

else" as the reference group as opposed to Caucasians.[21] For assessing disparity as pertains to Blacks, the reference group is non-Blacks. For assessing disparity as pertains to Hispanics, the reference group is non-Hispanics. This means that instead of comparing the stops of Blacks to the stops of Caucasians, the team is comparing the stops of Blacks to the combined stops of Caucasians, non-Black Hispanics, Arab Americans and so forth.

The key point here is that the reference group used by the Lamberth Team included other minorities who may themselves be victims of police racial bias. If a jurisdiction is practicing racially biased policing in their stops of non-Black Hispanics and Arab Americans, that fact will reduce the amount of disparity measured when Black stops are compared to a group that includes these Arab Americans and non-Black Hispanics. In using the reference group selected by the Lamberth Team, one ends up assessing whether minorities are treated differently from the combined group of Caucasians and other minorities. In fact, in Table 8 (page 49) of the report it is conceivable that the negative percentage differences (found at several locations) between the percent Hispanics in the benchmark and the percent Hispanics in the stops is due to disproportionate stopping of (non-Hispanic) Blacks at those locations. The higher the percentage of (non-Hispanic) Blacks stopped at that location, the lower the percentage of Hispanics stops.

In my opinion, the Lamberth Team's choice of a reference group for their calculations of odds ratios was wrong.

When Does Disparity Mean Bias?

A theme of *By the Numbers* that also runs through this report is that we can measure disparity easily but identifying the cause of disparity presents a challenge. As the researchers working in this field well know, fancy calculations do not overcome this challenge. The calculations that researchers present in their report represent different ways to measure disparity; these are not necessarily measures of bias. The key question then is "when does disparity mean bias?"

An identified amount of disparity in vehicle 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. The quandary for researchers is that they 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" researchers can set whereby disparity levels above it indicate racial bias and disparity levels below it indicate none.

[21] Their reference group is indicated by the team's articulation of how to understand the odds ratio. For instance, on page 46 the reader learns, "The odds ratio is best understood by filling in the blank in the following sentence: "If you are a Black motorist/pedestrian, you are ___ times more likely to be stopped as if you are not a Black motorist/pedestrian."

Some researchers, including Lamberth, have set cut-off points. As an example, in the MPD and other reports the Lamberth Team has provided a scheme for interpreting the different values of the odds ratios (e.g., "Ratios between 1.5 and 2.0 provide an indication that a review of stops in these locations should be conducted by the MPD"). This is as legitimate a practice as choosing *not* to set cut-off points. Researchers who set cut-off points are, in effect, arguing that if the disparity is particularly large then, chances are, alternative factors cannot explain all of it. Certainly, it is probably safe to say that larger disparities are more likely than smaller disparities to encompass many causes, including bias.

Other researchers choose not to set cut-off points. These researchers believe it is unwise to select a point above which "a problem" is indicated or a "next step" is advocated. They note that setting a cut-off point is rather arbitrary; a large disparity could be produced entirely by alternative, legitimate factors, and a small disparity could be entirely produced by bias. In the same vein, the finding of no disparity does not prove lack of racial bias.

As indicated above, researchers disagree on whether it is appropriate and advisable to set a cut-off point. Both practices are used. The current cut-off categories articulated by the Lamberth Team are appropriately cautious; that is, they do not convey that "racially biased policing exists above Level X." The team explains that, "odds ratios between 1.0 and 1.5 generally are seen as benign. Ratios between 1.5 and 2.0 provide an indication that a review of stops in these locations should be conducted by the MPD. Ratios above 2.0 point to the *potential* targeting of minority motorists, and further action may be required from the agency" (emphasis added). [22]

The Lamberth team conveys similar cautions when it aggregates the odds ratios across observation sites (that is, in producing a single odds ratio that encompasses the results from all of the individual intersections). The report reads, "we provide this aggregate odds ratio, but caution that it is for descriptive, not analytical, purposes, as not all statistical assumptions for further statistical tests can be met with these aggregated data" (p. 47).[23] This caution, unfortunately, is not conveyed to the stakeholder who reads only the executive summary and/or conclusion. In those sections of the report, for the traffic stop assessment, only the aggregate odds ratios are presented in conveying the results. Aggregate measures can be misleading and, if they are calculated and shared at all, they should be conveyed carefully in the context of the results obtained for each of the multiple, individual intersections. Several examples show the importance of presenting discrete results alongside aggregate results, if the researcher chooses to report aggregate results at all. In San Antonio, the aggregate odds ratio for the assessment of Black stops produced by the Lamberth Team was "benign" (1.3) even though odds ratios at individual intersections were as high as 3.3 (Lamberth, 2003). In the Grand Rapids study, too, the aggregate odds ratio for the assessment of Black stops was "benign" (1.4) although odds ratios at particular intersections were as high as 3.7 (Lamberth, 2004b).

---

[22] In prior reports (e.g., Lamberth, 2004a, p. 25), Lamberth describes the 1.5 to 2.0 odds ratios as "(suggesting) that in the absence of other explanations, targeting of [minorities] may be occurring."

[23] I had raised a question about the method the team used to produce the aggregated odds ratios and find the response to my inquiry on pages 47 and 48 of the Lamberth Team's report to be quite reasonable.

Summary and Conclusions

The observation method is a strong method for analyzing stop data. Dr. Lamberth and his team at Lamberth Consulting pioneered this method and their implementation in Washington, D.C. has some strengths. These strengths include:

- The form of observation (stationary or mobile) used
- Their process for selecting locations for the observations
- The random selection of days of week and times of day for observations
- The training received by the surveyors
- The focus on proactive stops
- The various ways they "matched the numerator to the denominator."

One major criticism is that the Lamberth Team used inappropriate reference groups in calculating their odds ratios, which they then used to assess the existence of racial bias. Another major criticism is their use of 60 minutes of observation data at intersections to benchmark stops made at those intersections year-round.

In the next section, I report on the constraints of social science that limit our ability to draw conclusions about police racial bias from vehicle/pedestrian stop studies.

Section IV

Drawing Conclusions from the Results

In Section I, I argued that in reviewing a vehicle/pedestrian stop study designed to measure police racial bias, one must consider (1) the strength and weaknesses of the benchmarking method selected, (2) the strength and weaknesses of the implementation of that method by the researchers, and (3) the constraints of social science as pertains to the study of police racial bias.

In Section II, I described the strengths and weaknesses of the observation method. In Section III, I examined the implementation of this method by the Lamberth Team finding strengths and weaknesses. In this section, I review some of the constraints imposed by social science on efforts to measure police racial bias and then comment on our ability to draw conclusions about the existence or lack of racial bias by police during vehicle/pedestrian stops in Washington D.C.

In Part A of this section, I summarize what is not measured in this study and others and how these various information deficiencies impact on our ability to draw conclusions. In large part due to the constraints of social science (combined with resource issues), this study does not:

- Incorporate differences in driver/pedestrian behavior (e.g. speeding, running red lights, etc) into its benchmark
- Measure disparity as pertains to racial groups other than African Americans and Hispanics
- Measure disparity in police activities other than traffic/pedestrian stops
- Measure disparity in stops that occur in areas of the city not studied

In Part B of this section, I explain that the constraints of social science limit, not just our ability to determine when racial bias *exists*, but also our ability to determine when it *does not exist*. I link this key point to the conclusions drawn by the Lamberth Team in the report for MPD.

Part A: What is Not Measured

In section II I explained that in order to measure whether there is a causal relationship between driver/pedestrian race/ethnicity and police stopping behavior, we must rule out—through our social science methods—explanations other than bias. Further, to be fully comprehensive in such a study, it would be ideal to test this relationship in all areas of a jurisdiction and to study police activities other than stops.

Driver/Pedestrian Behavior

Not every person on the road or on the sidewalk is at equal risk of being stopped by police even in an unbiased world. Behavior is one factor that might legitimately increase or decrease the likelihood that a person would be stopped in a vehicle or on foot by police. Yet in this study as in most others, we don't have a measurement of driver or pedestrian behavior. This fact returns us to the previous discussion regarding whether a benchmark should represent "who is driving" or "who is violating."

Recall, that the Lamberth Team benchmarked the MPD stops against observations of "who is driving." Separate from this analysis, they conducted a study-in-a-study to assess whether African Americans, Hispanics and others exceeded the speed limit at selected intersections at the same or different rates. They found differences across racial/ethnic groups. Specifically, they found that, in the limited geographic areas studied and for the single violating behavior of speeding, African Americans are underrepresented among speeders relative to their representation among drivers. (The Lamberth Team reports the same finding for Hispanics, although the data for this assessment were not sufficient for purposes of drawing conclusions.)

As discussed previously, social scientists have offered some viable reasons for not measuring driving behavior including the argument that it would be incredibly complex to reliably and validly measure all the possible violating behaviors that might legitimately lead to a stop. That said, an increasing body of research shows that there are demographic differences in driving behavior by race; in the present study the Lamberth Team reports some differences in violating behavior across racial groups in Washington, D.C. In explaining their inclusion of this important component of the study, the Lamberth Team reports that "a better understanding of violations by racial/ethnic groups would help sharpen our analysis of whether one or another group was being stopped more than would be expected. The logic here is that if one racial/ethnic group were more egregiously violating one or more traffic laws, it would allow us to use this information in our final determination of whether some group was being targeted" (p. 54).[24]

Some hypothetical research results will help to convey this important point.[25] Let's say that a jurisdiction uses observations of "who is driving" to create a benchmark and finds that the demographics of "who is driving" matches perfectly the demographics of people stopped by police in their vehicles. The researchers, on the basis of these findings, might report "no evidence of targeting." However, because we have not measured driving behavior, it is still possible that policing in the jurisdiction is racially biased. If minorities are violating traffic laws at proportions *less than their representation as drivers*, the fact that they are stopped *proportionate to their representation as drivers* indicates disparity, and it may indicate racially biased policing. In this example, the existence of racially biased policing may be "masked" by weaknesses of the benchmark.

The hypothetical data on the representation of minorities and nonminorities among drivers, traffic violators, and drivers stopped by police are presented in Figure 3. It shows that 25 percent of the drivers are racial/ethnic minorities as are 25 percent of the people who are stopped by police for traffic violations. If researchers were comparing just these two proportions, they would report "no disparity" or "no evidence of targeting."

The figure, however, also shows (the middle bar for each group) the proportion of minorities and nonminorities who are traffic violators (information that would <u>not</u> be

[24] Lamberth seems to contradict this sentence about the "logic of assessing differences in driving behavior across racial groups" elsewhere in the report. On page 54 he argues, using the quote just referenced, that it's important to gather this information to interpret the results. However, in a footnote added to the final draft he argues that it is not appropriate to use these findings to interpret the results because "speeding is but one violation among hundreds for which motorists can be stopped" (footnote #3 on page 5).

[25] This example is based on a similar presentation in *By The Numbers* pp. 31-32.

available to the researcher who measured only "who is driving"), and this information indicates that minorities are over-represented among the drivers who are stopped. 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*, disparity is indicated. The key here is that the researcher conducting observation benchmarking for "who is driving" would not have the information on violating behavior. S/he might conclude "no disparity" when, in fact, disparity exists.

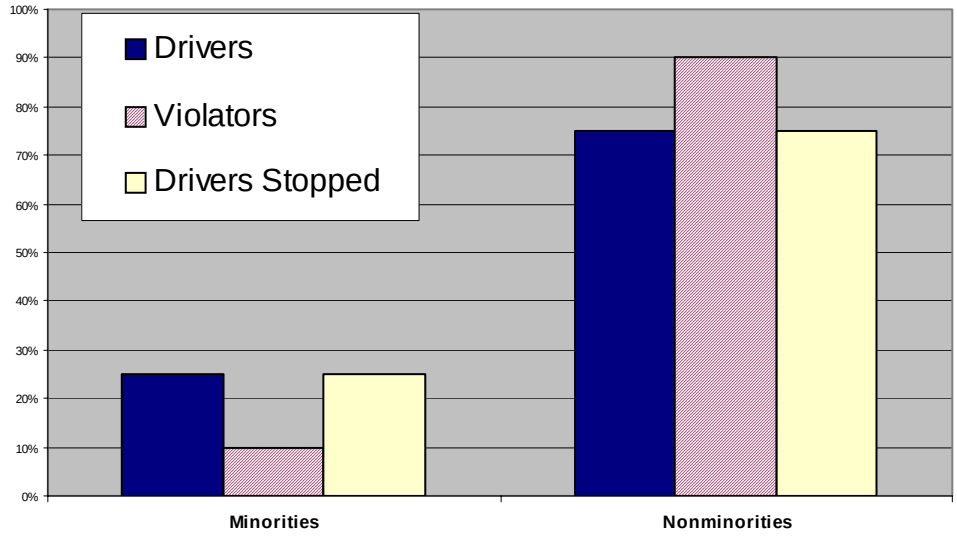

Figure 3. Racially Biased Policing "Masked" When Violators are not Measured

This hypothetical example reflects the constraints on the interpretation of the MPD study by virtue of the (understandable) decision by the Lamberth Team to use observations of "who is driving" for benchmarking the MPD stops. For drawing conclusions about the existence of racial bias, the Lamberth Team only had information on "drivers" and "drivers stopped" (the left and right bars for each group) and not on "violators" (the middle bar). This weakness is not unique to the Lamberth Team study.

Although we don't have for Washington, D.C. the (middle) bars in Figure 3 showing "who is violating" for each racial/ethnic group for each of the intersections chosen for the assessment, we have some limited information provided by the Lamberth Team that implies that the "violators" bar for minorities might be lower than the "drivers" bar. At least in the limited geographic areas studied and for the single violating behavior of speeding, the Lamberth team found that African Americans are underrepresented among speeders relative to their representation among drivers.[26]

Let me summarize the issues pertaining to the measurement of driving behavior as it pertains to the Lamberth Team's study.

[26]Again, there is not sufficient data to draw conclusions for Hispanics.

- In my view, it is acceptable that the team decided to benchmark the MPD stops against "who is driving" instead of "who is violating." Social scientists make this decision because it is arguably impossible to measure with precision "who is violating."
- It is commendable that the Lamberth Team decided to conduct a "study within a study" to assess differences in speeding behavior across demographic groups. They found differences in speeding behavior across racial/ethnic groups.
- The cause for concern, however, is that the Team does not explain the drawbacks of the inability to benchmark with measures of "who is violating" versus "who is driving." Relatedly, it is inappropriate that the Team did not acknowledge the ramifications of their study-within-a-study results for interpreting their findings regarding racial/ethnic disparity.

The above example that pertains only to the drawbacks of not measuring violating behavior, has application to the other behaviors that might impact on police stopping activity. Dr. Lamberth lists several additional unmeasured behaviors/factors that might produce differential police stopping of racial ethnic groups. He points out, for instance, that minority motorists might be stopped more frequently than non-minority motorist because minorities may be overrepresented among the drivers who are driving vehicles with equipment malfunctions. Specifically, he notes that minorities are "overrepresented in the lower socioeconomic strata of the country," "poor motorists generally find it more challenging to keep their vehicles in good repair than do more affluent motorists," and motorists driving vehicles with equipment malfunctions are at more risk of being pulled over by police than motorists driving vehicles without such malfunctions (pp. 56-57). He also points out that "studies have shown that Black motorists, particularly young Black and Hispanic motorists, wear seat belts less frequently than do other groups." This is another *unmeasured* behavior that could impact on the frequency of minority stops.

To help us understand how to evaluate the issue of unmeasured behavior, I return to the example of math grades discussed in Section II to reiterate how unmeasured factors limit our ability to draw conclusions. In that example, the unmeasured factor was math performance. A study attempting to examine whether teachers were gender-biased in their allocation of grades in a math course would need to consider the actual (objectively measured) math performance of the students. Without such a control, valid conclusions could not be drawn about the types of grades given to male and female students.

Disparity as Pertains to Other Racial/Ethnic Groups

As reported in Section III, some researchers collapse or reduce racial categories in their studies because of the difficulties of measuring (e.g., through observation) which people belong to which of the various different racial/ethnic groups. While this is an acceptable decision based on the constraints of social science, the result is our inability to detect racial bias for all groups that might be targeted by police. Because of measurement issues, the present study in Washington, D.C. only measures disparity in police stops for African Americans and Hispanics. We cannot draw any conclusions about disparities in stopping Asian, Middle Eastern, Native American or other racial or ethnic groups that walk or drive in Washington, D.C.

Disparity in Police Activities Other than Traffic/Pedestrian Stops

The MPD study focused only on police decisions to pull over vehicles and to detain pedestrians. Of course, there are lots of other activities in which police engage that may produce biased policing including arrests, evidence-based searches, requests for consent to search, attempts to seek warrants, and the initiation of investigations, to name a few.

One realm of activities that are sometimes the subject of jurisdiction studies on police racial bias are post-stop activities such as issuing a citation or searching a vehicle. In fact, stakeholders have expressed the concern that poststop activities may be at even greater risk of racially biased policing and may have greater negative consequences than stop activities. Police can exercise considerable discretion in making poststop decisions (for instance, whether to request consent to search, what disposition to give) and, as explained in *By the Numbers*, high-discretion decisions are at greater risk than low-discretion decisions for racial bias. Poststop decisions may result in significant costs for motorists. For instance, searches intrude on motorists' liberty and privacy, they produce fear and even embarrassment, and they "mark" the person as a suspect. A decision to give someone a ticket rather than a warning will have primary costs in the form of fines and potential secondary costs in the form of insurance rate increases. The Lamberth Team was not asked by MPD to conduct poststop analyses or to analyze other police activities. As such, the MPD study looks at only two police activities–stops of drivers and stops of pedestrians. These activities do not encompass all of the police activities in which racial bias might manifest.

Disparity in Stops that Occur in Areas of the City Not Studied

The observation method is strengthened when the researchers "match the numerator to the denominator" in the various ways referenced in Section III. As pointed out, however, these techniques that enhance the rigor of the analyses reduce the scope of the study. What this means is that, in the city of Washington, D.C. that encompasses over 60 square miles, we have an assessment of racially biased vehicle stops of African Americans at only 20 intersections and of Hispanics at seven intersections. We have an assessment of racially biased pedestrian stops of African Americans at two intersections and of Hispanics at a single intersection.

We most certainly do not have a comprehensive assessment of bias across all areas of Washington, D.C. We have a "spot check" of disparity at a few locations.

Part B: Bi-directional Nature of the Constraints on Research

A key point made in *By the Numbers* is that the constraints on measurement pertain to our ability to measure both whether (1) racially biased policing exists or (2) racially biased policing does **not** exist. Specifically, I refute the "myth" that "no racial/ethnic disparity means no racially biased policing" (p. 30). The results produced in a study, regardless of whether they show under-representation, over-representation, or proportionate representation of minorities among the persons stopped by police, do not enable researchers to draw conclusions about the existence of racially biased policing. This important caution has not been heeded by authors of a few reports. In some

instances, authors of these reports acknowledge that their benchmarking method cannot produce conclusions regarding the *existence* of racially biased policing, but they argue 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 researchers who put forth the argument that a showing of no disparity means no racially biased policing fail to recognize that an imperfect benchmark can "mask" (or hide) disparity. An example of "masking" was provided above in the discussion of the ramifications of not measuring driving quality (not measuring "who is violating") in a study of traffic stops. We presented hypothetical results in Figure 3 from a study that looked only at "who is driving." In that hypothetical study, the researchers found proportionate stops of minorities relative to their presence in the driving population but, because of limitations on their methods (most notably, not measuring "who is violating"), they did not detect the possibility that minorities were over-stopped relative to their representation in the population of traffic law violators.

The Lamberth Team's results conveyed as "no evidence of targeting" is merely a finding of "no disparity." Based on the constraints associated with the study, I contend that the statement that "MPD is to be congratulated for their evenhandedness in enforcing the traffic laws at the 20 locations" goes beyond what has been found or indeed what *could be found* in such a study.

Summary

The point of this section is to convey that even a strong study using strong methods is subject to the constraints of social science that limit our ability to draw conclusions about police racial bias. The MPD study (like many others conducted across the country) does not, in assessing disparity/bias, incorporate or produce information on:

- Driver/pedestrian behavior
- Disparity as pertains to racial groups other than African Americans and Hispanics
- Disparity in police activities other than traffic/pedestrian stops
- Disparity in stops that occur in areas of the city not studied.

As I wrote in the last chapter *of By the Numbers*: "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" (p. 371). Despite the sincerity of most people posing the questions, answers that are definitive cannot be offered.

Faced with this fact, one well might ask: of what value are these results if researchers cannot report, with confidence, the existence or lack of racial bias in the jurisdiction? The answer is that results like those from the Lamberth Team's study *can* be of 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. 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).

In the next chapter, I discuss how these data can be used in the context of making some recommendations directed to the new leadership at the MPD and the Biased Policing Task Force as they continue in their quest to address the issues of racially biased policing and the perceptions of its practice.

Section V
Next Steps

Because the data will never "prove" or "disprove" racially biased policing, I contend that vehicle/pedestrian stop data collection can never be viewed—either by police or resident stakeholders—as a "pass-fail test" (Farrell 2004). Studies—such as that conducted by the Lamberth Team in Washington, D.C.—can, however, be viewed as diagnostic tools to help pinpoint the decisions, geographic areas, and procedures that might get priority attention when the agency, in concert with concerned residents, identifies its next steps for addressing the problem or perception of racial profiling.

As articulated by Chief John Timoney (2004) of the Miami Police Department, the reality is that "race is a factor in policing." The current issues involved in "racial profiling" and "racially biased policing" are not new—they are the latest manifestations of a long history of sometimes tense, even volatile, police-minority relations. The longstanding nature of the issue 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 the past few decades, there has been a revolution in the quality and quantity of police training, the standards for hiring officers, the procedures and accountability regarding police activity, and the widespread adoption of community policing.

The MPD is moving toward state-of-the-art practices in a number of realms including homeland security and use-of-force investigations. In this section I provide a guide for MPD achievement of state-of-the-art practices in fair and impartial policing.

I contend that *every* police executive needs to consider and address the issues of racially biased policing and the perceptions of its practice. *Every* agency should have meaningful policy that tells officers when they can and cannot use race and ethnicity to make law enforcement decisions. *Every* agency needs to have effective academy and in-service training directed to the ways that racial bias might manifest in even the best departments. *Every* agency should have strong accountability mechanisms in place to promote professional behavior and those mechanisms ought to be directed, as well, to promoting fair and impartial policing. *Every* agency should attempt to hire diverse personnel and people who can police in an unbiased fashion. *Every* agency should have ongoing efforts to strengthen the relationships between the department and the diverse communities that it serves.

Below I describe the potential manifestations of racially biased policing. In the section that follows, I describe the elements of a comprehensive response to racially biased policing and the perceptions of its practice. Finally, I set forth the steps that the MPD can take to achieve these state-of-the-art policies/practices in partnership with the Biased Policing Task Force. I conclude by summarizing my recommendations.

Ways that Racially Biased Policing Might Manifest

Executives need to think about and respond to the various ways that actual and perceived racially biased policing might manifest. Specifically, the executive needs to think about (1) how to identify and deal with "bad apples," (2) how to guide well-meaning officers, and (3) how to identify and fix institutional practices that contribute to the problems. That is, an executive needs to attend to the possibility of current or future problems of biased policing that result from the actions of a few "bad apples" among

his/her personnel, the unintentional biased activities on the part of well-meaning officers, and/or department policies and practices that inadvertently (we hope, anyway) reflect bias or result in biased enforcement decisions.

The “bad apples” are the small minority of racist officers who act on their biases with impunity. These personnel are a great challenge to executives. Policy and training are not likely to impact on these officers; for the most part, their actions are already contrary to the existing policies of the agency and the training they have received. The greatest hopes for changing the behavior of these practitioners are close and effective supervision, an early warning system to identify problem officers, and accountability through discipline or dismissal. There may be additional problem officers that, while they may not be conducting racially biased policing, treat citizens in such a negative and disrespectful manner as to give rise to the perceptions of biased behavior. Again, this calls upon measures in the form of effective supervision, early warning, and accountability.

In stark contrast to these bad apples, the vast majority of police personnel are well-meaning individuals who are dedicated to serving all citizens with fairness and dignity. Despite their good intentions, however, their behaviors may still manifest racially biased policing or give rise to the perceptions of it. It is likely that many of these officers, like humans in every profession, are not fully cognizant of the extent to which race/ethnicity are used in their decision making or fully cognizant of the behaviors that may give rise to citizen perceptions of bias. These officers need policy to provide them with guidance on the circumstances in which race/ethnicity are and are not appropriate factors in the decisions they make and training that conveys that anti-biased policing policy and that facilitates their analytical understanding of racially biased policing. They need to be informed about their actions that citizens perceive as biased. Also important are effective supervision and an agency reward structure that reinforces behaviors consistent with the executive’s commitment to impartial law enforcement and the dignified treatment of all citizens.

An executive needs to consider the possibility that racial bias is not the result of either intentional or unintentional deviation from agency policy or standards, but instead is the result of adherence thereto. Executives need to review policies, enforcement strategies, deployment, reward structures, and other operational administrative practices to ensure that they do not reflect biases, and communicate openly and constructively with residents who express concern about the disparate impact of police policies.

The Elements of a Comprehensive Response to Racially Biased Policing

In this section, I describe briefly the key elements of a comprehensive response to racially biased policing. More extensive coverage is contained in the book, *Racially Biased Policing: A Principled Response* (Fridell et al., 2001) and in a chapter entitled “Law Enforcement Agency Responses to Racially Biased Policing and the Perceptions of its Practice” (Fridell and Scott, 2005).

The issue of racially biased policing actually breaks down into two challenges for the executives of law enforcement agencies: (1) personnel’s racially biased policing conduct, and (2) residents’ perceptions of racially biased policing. We are unable in some contexts or within individual incidents to determine whether racially biased policing is real or perceived. The executive should commit to address both even if they cannot be

fully disentangled. While racially biased policing is the misuses of race/ethnicity to make law enforcement decisions, the counterpart is the perception on the part of citizens that race/ethnicity is being used inappropriately in police decision-making.

In a comprehensive response to both racially biased policing and the perceptions of its practice the key areas are:

- Policies prohibiting racially biased policing,
- Education and training,
- Supervision and accountability,
- Recruitment and hiring,
- Outreach to diverse communities, and
- Data collection and analysis.

Anti-Biased Policing Policy

Racially biased policing occurs when law enforcement inappropriately considers race or ethnicity in deciding with whom and how to intervene in an enforcement capacity (Fridell et al., 2001). There are significant differences of opinions as to when it is and is not appropriate to consider race or ethnicity and these views are reflected in the various policies that have been adopted around the nation. It is critically important for an agency executive to be sure that his/her personnel are provided with guidance as to the appropriate use of race/ethnicity in making decisions. A key way to convey this guidance is in policy supported by training.

Executives shouldn't assume that all of their personnel use race/ethnicity the same way and should be concerned that their use may be broader than what the executive (and the agency's constituencies) believes is just. In focus groups held around the country, it became clear that practitioners at all levels–line officers, command staff and executives–have very different perceptions regarding the circumstances in which officers can consider race/ethnicity. Participants discussed when officers can use race/ethnicity as one factor in the "totality of the circumstances" to establish reasonable suspicion or probable cause. We found many differences of opinion among line officers and command staff, even *within agencies*, on this point.

Education and Training

Education and training can play critical roles in reducing actual and perceived racial bias in policing. It is my belief (there have been no empirical assessments), that most agencies across the country assume they are providing comprehensive and effective training in racially biased policing based on their use of traditional diversity training (i.e., conveying to officers how to most effectively interact with people of varying races, ethnicities, traditions) and professional traffic stop training. While these topics are important, they are not effective at addressing the needs of the well-meaning officers, described above, who may not be fully cognizant of the extent to which bias impacts on their decision-making. Training can not rid humans of their biases; instead training should make *unconscious* bias *conscious*. That is, effective law enforcement training will convey to recruits and/or in-service personnel the biases that many humans have, how these biases could manifest in decision-making, and how to ensure that bias in decision-making does not, in fact, occur. There are some academy and in-service curriculum models that have this focus.

Quality first-line supervisor training on racial bias topics is an important component of a comprehensive response. Such training would include guidance in how to identify and respond to officers who may be acting in a racially biased manner or acting in a manner that produces perceptions of racially biased policing. Such training is particularly important because of the critical role of first-line supervisors in promoting professional law enforcement behavior. It is also important because dealing with race issues in any employment setting is fraught with difficulties. These supervisors should receive guidance for dealing with these very challenging issues.

Supervision and Accountability

Police accountability and supervision are important for reducing or eliminating bias in policing. The importance of first-line supervisors and the need for targeted training was discussed above. Equally important are the development and application to fair and impartial policing of strong accountability measures. Every department has some accountability mechanisms that it uses to promote quality performance at all levels. These include Early Intervention Systems, in-car cameras, COMPSTAT, complaint receipt/review systems, and so forth. With regard to racially biased policing, these systems (1) need to be strong, and (2) need to be used to promote fair and impartial policing.

Recruitment and Hiring

Recruiting and hiring practices have the potential to reduce racially biased policing and citizen perceptions that an agency is biased in two basic ways: (1) by establishing a police workforce that reflects the racial demographics of the community that the agency serves, and (2) by hiring officers who can police in an unbiased manner and in a manner that reduces perceptions of racial bias.

Much has been written to guide agencies in their efforts to bring diverse applicants into the pool. It's also important for a department to examine the process for screening applicants to determine if minorities disproportionately fail or "no show" at particular stages. Knowing this information, the department can assess whether, *while maintaining the high screening standards of the agency*, efforts can be implemented to reduce this loss.

Another aspect of recruitment and hiring that pertains to racially biased policing is hiring people who can police in an unbiased manner. It is no easy task to identify these people. Two screening mechanisms, the personal interview and background investigation, however, have potential for identifying an applicants' strengths and weaknesses as pertains to biased behaviors.

Outreach to Diverse Communities

Both the incidents and the perceptions of racially biased policing lead to mistrust of police. Relying as they do on resident input, support, and compliance, the police cannot function effectively where tensions are prevalent. Outreach to all residents, but particularly to minority communities is an important component of any departmental strategy to respond to racially biased policing and the perceptions of its practice. Departments should (1) reach out to minority communities on the specific topic of racially

biased policing, and (2) institute methods for building and sustaining, at a more general level, mutually respectful and trusting relationships.

With regard to outreach on the topic of racially biased policing, police practitioners should be willing to discuss racially biased policing and perceptions thereof with community residents. Constructive dialogue between the police and citizens can lead to an agreement that racially biased policing likely occurs to some unknown degree within the jurisdiction, but perceptions may not always reflect the scope and nature of the problem. Additionally, continued dialogue about enforcement activities (e.g., traffic and pedestrian stops, drug enforcement, aggressive violence abatement) can help stakeholders understand how police make decisions about which activities to use and where. With this understanding in place, police and citizens can begin to collaborate to develop ways to address problems in their community without the use of bias. A jurisdiction-level task force comprised of police and citizens can take the lead—analyzing the problems and formulating interventions.

Data Collection and Analysis

As this report conveys, there are great challenges associated with efforts to measure racially biased policing. Data collection efforts, however, *can* be a viable part of a comprehensive response so long as stakeholders understand fully what the effort can and cannot produce and so long as resources exist to both (attempt to) *measure* racially biased policing and *respond* to it.

Next Steps

The MPD has already taken significant steps towards its "goal of ensuring effective, bias-free policing in the District of Columbia."[27] The department commissioned the report of the Police Foundation and, pursuant to recommendations therein, commissioned the Lamberth Team to implement a vehicle/pedestrian stop study. The MPD developed the Community-Police Task Force the members of which were involved in selecting the research team and meeting periodically with the team and MPD representatives as the study progressed. The department also had the Police Executive Research Forum conduct focus groups around the city with residents.[28]

These efforts provide a strong basis for forward movement on the critical issues of racially biased policing and perceptions of its practice. In this section, I pull from Chapter 13 of *By the Number*s to describe how the MPD could continue its implementation of state-of-the-art practices in concert with the Biased Policing Task Force.[29]

[27] This quote is from page 1 of the MPD website-based "Response to the Final Report and Recommendations of the Biased Policing Project" referring to the report completed by the Police Foundation at http://mpdc.dc.gov/mpdc/cwp/view,a,1246,q,555635.asp.
[28] This report has not yet been released.
[29] The Lamberth Team, too, recommends that the MPD continue work with the Biased Policing Task Force.

The Police-Community Task Force: Membership and Role

It is significant and important that the MPD developed the Biased Policing Task Force as part of its Racially Biased Policing Project.[30] The MPD should continue to utilize this group. As described in more depth below, this task force would be advisory to the Chief as the department implements state-of-the-art practices in fair and impartial policing. The current group seems to represent well the community groups in Washington, D.C. most committed to fair and impartial policing. I recommend, however, that more and different representatives from the police department be added. It was arguably appropriate during the course of the Lamberth Team's study to have disproportionate representation on the task force of department personnel involved in research. For the next steps, the task force should include police personnel representing all departmental levels, particularly patrol. (The chief should not be a "member" of the group, since the group is convened to provide her with advice on what actions to take.)

In Chapter 13 of *By the Numbers*, I discuss methods that might be used during the first few sessions of the newly constituted task force to develop trust between police and resident members. The task force would then, over the long term, discuss:

- General issues and concerns related to racially biased policing,
- What the group can learn from the Lamberth Team vehicle/pedestrian stop study, and
- What other sources of information indicate about racial bias and perceptions of racial bias.

With regard to a review of the Lamberth Team's results, the task force should review the "disparity" findings presented in the report. As above, while data collection cannot prove or disprove racially biased policing it can "help pinpoint the decisions, geographic areas, and procedures that should get priority attention" (Fridell 2004, p. 362). Consistent with the recommendations of the Lamberth Team, the Task Force should explore the factors that might produce the disparity results that are reviewed below.

The Lamberth Team conducted four key analyses; they assessed racial disparity in both vehicle and pedestrian stops and they assessed ethnic disparity in both vehicle and pedestrian stops. The Lamberth Team reports that odds ratios between 1.0 and 1.5 are "benign." Odds ratios between 1.5 and 2.0 require "a review of stops in these locations." A value greater than 2.0, according to the Lamberth Team, indicates "potential targeting" of minorities requiring "further action…. from the agency."

In a number of locations studied in Washington, D.C., the Lamberth Team produced "no disparity" results. In the assessment of *racial disparity* in *vehicle* stops in Washington, D.C., five locations of 20 produced odds ratios in the 1.5 to 2.0 range (Lamberth 2004, p. 45). In the assessment of *ethnic disparity* in *vehicle* stops, three of seven locations

[30] It is unfortunate that the MPD did not, at least during the study period, maximize the trust-building potential of this group. As one example, although MPD had previously agreed to meet with the Biased Policing Task Force to discuss the final report and to share it with the group in advance of its release, the MPD did not follow through on this commitment.

produced odds ratios in this range (p. 49).[31] In one of the two locations where viable results were produced for assessing *racial disparity* in *pedestrian* stops, the odds ratio was in the 1.5 to 2.0 range. Finally, in the single intersection at which *ethnic bias* was assessed for *pedestrian* stops, the odds ratio was 2.0.

The highest odds ratio was found in the assessment of racial bias in pedestrian stops. At Wisconsin Avenue and M Street NW, which is the main intersection in Georgetown, the odds ratio was 5.8. According to the Lamberth Team's odds ratio formulation, this means that African Americans walking in the vicinity of Wisconsin and M streets are almost six times as likely to be stopped as non-African Americans. The Lamberth Team refers to these results as "potential targeting." In reflecting upon these disparity results, the task force might ask whether African Americans are at risk for "race out of place" stops in this major tourist/shopping area of the city.[32]

With regard to alternative sources of information, the task force should review the *Biased Policing Project* report of the Police Foundation (2004).[33] The Police Foundation (PF) conducted focus groups of community residents and of police personnel. The PF team collected and analyzed agency documents and reports and conducted a survey of community residents. On the positive side, the PF team found that majorities of the community members surveyed "felt that the police do a good job preventing crime," that police were polite and professional, and otherwise had positive attitudes toward police. On the other hand, the PF team reports that "most survey respondents felt that MPD officers are more likely to stop non-whites for traffic violations" (2004: 4).

Ultimately, the role of the task force is to make recommendations to the Chief that will facilitate the department's efforts to achieve state-of-the-art practices–tailored to the city–in fair and impartial policing. Within the various realms of a comprehensive response, the group would assess current policies/practices and identify areas that need strengthening. Their assessment and recommendations would address:

- Policy prohibiting racially biased policing,[34]
- Education and training,
- Supervision and accountability,
- Recruitment and hiring,
- Outreach to diverse communities, and
- Data collection and analysis.

[31] Assessments were made at 20 intersections, but only at seven locations were viable results produced. See Table 8, page 49 of the Lamberth Team's report.

[32] The task force should review the Lamberth Team's recommendations regarding "operational practices relative to pedestrian contacts" (p. 61).

[33] By then, the group should also have access to the report of the Police Executive Research Forum based on focus groups held around the city.

[34] The MPD policy on "Unbiased Policing" (GO-OPS-304.15) has some strong elements. If it was not developed with the input of resident stakeholders, a joint review of the policy should be conducted by the MPD and the task force as the city moves forward to implement a comprehensive response to racially biased policing.

Comments on Continued Data Collection

The task force should decide, in concert with the Chief, whether data collection efforts should continue in light of perceived benefits and resources. I would like to offer two priorities for data collection efforts, if the MPD continues efforts in this realm. I would have the department prioritize the implementation of an internal benchmarking system and collect/analyze data on police searches.

The priority to implement an internal benchmarking system reflects the recommendation of the Police Foundation that the department "incorporate data on stops into the Personnel Performance Management system in order to assess racial profiling of individuals or units" (PF 2004: 117). While internal benchmarking, as reported in Section II of this report has some weaknesses, its strength is in pinpointing the particular source(s) of disparities. Recall, that with this method, the department compares similarly situated officers (or units) to each other. Chapter 8 of *By the Numbers* provides specific guidance regarding the use of this method.

The Lamberth Team, too, recommends attention to searches. Chapter 11 of *By the Numbers* provides guidance for poststop analysis; unfortunately, there are as many social science constraints associated with analyzing poststop data as there are in analyzing data on decisions to stop. That said, I believe that there are some methods for analyzing police searches that can provide valuable guidance to agencies. If data collection efforts continue, I suggest that the MPD collect information on searches that will allow it to conduct proper "hit rate" analyses of them. 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. Hit rates can be calculated for each racial/ethnic group and, when calculated in accordance with the directions contained in Chapter 11 of *By the Numbers*, can determine if there is *problematic* disparity in searches conducted by police.[35]

Importantly, if the MPD continues with data collection efforts, it should make sure it has a viable and transparent auditing system to (1) ensure that the data that are collected are reliable and valid, and (2) to reassure stakeholders that any conclusions drawn are based on data that are reliable and valid.

Summary of Recommendations

My specific recommendations to the MPD are to:

- ✓ Release to interested stakeholders the data used by the Lamberth Team for its research.

- ✓ Release the results of the auditing analyses that the department conducted on these data.

- ✓ Maintain and expand the membership of the Biased Policing Task Force.

[35] In *By the Numbers*, I acknowledge that there are still ongoing debates among social scientists regarding the value of hit rates.

- ✓ Charge this task force with advising the chief on the department's continued efforts to ensure fair and impartial policing.

My recommendations to the MPD in concert with the Biased Policing Task Force are to:

- ✓ Review and discuss the findings presented in the Lamberth Team's report.

- ✓ Draw upon other information that pertains to racially biased policing and the perceptions of its practice.

- ✓ Identify what the department needs to do to achieve state-of-the-art practices with the realms of:
    - Policy prohibiting racially biased policing,
    - Education and training,
    - Supervision and accountability,
    - Recruitment and hiring,
    - Outreach to diverse communities, and
    - Data collection and analysis.

- ✓ With regard to data collection, the Chief and task force should decide jointly whether and/or what continued data collection efforts are worthwhile. I recommend as priorities internal benchmarking and hit rate analyses of searches.

- ✓ Implement state-of-the-art practices within the MPD to address racially biased policing and the perceptions of its practice.

References

Farrell, Amy and Jack McDevitt (2006). Racial Profiling Data Collection: A Decade of Research. A paper presented at the Annual Meeting of the American Society of Criminology, Los Angeles, November.

Fridell, Lorie and Mike Scott (2005). "Law Enforcement Agency Responses to Racially Biased Policing and the Perceptions of its Practice." In Roger G. Dunham and Geoffrey P. Alpert (Eds). Critical Issues in Policing, 5th Edition, pp. 304-321. Prospect Heights, IL: Waveland Press.

Fridell, Lorie (2004). By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops. Washington, DC: The Police Executive Research Forum.

Fridell, Lorie (2005). Understanding Race Data from Vehicle Stops: A Stakeholder's Guide. Washington, DC: The Police Executive Research Forum. (This guide summarizes the key content of *By the Numbers*.)

Fridell, Lorie, Bob Lunney, Drew Diamond and Bruce Kubu with Michael Scott and Colleen Laing (2001). Racially Biased Policing: A Principled Response. Washington, DC: The Police Executive Research Forum.

Lamberth, John (2003). Racial Profiling Data Analysis Study. Report submitted to the San Antonio Police Department by Lamberth Consulting.

Lamberth, John (2004a). Ann Arbor Police Department Traffic Stop Data Collection Methods and Analysis Study. Report submitted to the Ann Arbor Police Department by Lamberth Consulting.

Lamberth, John (2004b). Grand Rapids Police Department Traffic Stop Data Collection Program. Report submitted to the City of Grand Rapids by Lamberth Consulting.

Police Foundation (2004). Biased Policing Project. A final report submitted to the Metropolitan Police Department by the Police Foundation on September 1, 2004.

Appendix A

About the Author

Dr. Lorie Fridell is an Associate Professor in the Department of Criminology at the University of South Florida (USF). Prior to joining USF in August of 2005, she served for six years as the Director of Research at the Police Executive Research Forum (PERF). Dr. Fridell has 20 years of experience conducting research on law enforcement. Her subject areas include police use of deadly force, use of less-than-lethal weapons, police-minority relations, police pursuits, and violence against police. Additionally, Dr. Fridell is also a national expert on racial profiling. She is the first author of the PERF book entitled *Racially Biased Policing: A Principled Response*, which guides law enforcement agencies in their response to both racially biased policing and the perceptions of its practice and the author of *By the Numbers: A Guide for Analyzing Race Data From Vehicle Stops*. (These and other publications are available for free downloading from the PERF website at www.policeforum.org.) Dr. Fridell speaks nationally and consults with agencies on the topic of racially biased policing; she testified before a subcommittee of the Senate Judiciary Committee on the topic. She completed her bachelor's degree in Psychology at Linfield College in McMinnville, Oregon, and both her master's and Ph.D. in Social Ecology at the University of California at Irvine.