INTERNAL BENCHMARKING APPROACHES FOR IMPROVING CITIZEN SATISFACTION WITH THE POLICE AND CONTROLLING POLICE DISCRETION[1]

John M. MacDonald, Ph.D.
Jerry Lee Assistant Professor of Criminology
Department of Criminology, University of Pennsylvania,
483 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6286
(E-mail: johnmm@sas.upenn.edu)

Greg Ridgeway, Ph.D.
Director of Safety and Justice Program
Senior Statistician, RAND Statistics Group
RAND, 1776 Main Street, Santa Monica, CA 90407
(E-mail: gregr@rand.org).

June 23, 2009

[1] This Working Paper was prepared for the panel, "Cost-Effective Police Methods – Research-based Governing of Police Work" at the annual 2009 Stockholm Criminology Symposium. Points of view are those of the authors only. Please do not quote or cite.

INTERNAL BENCHMARKING APPROACHES FOR IMPROVING CITIZEN SATISFACTION WITH THE POLICE AND CONTROLLING POLICE DISCRETION

ABSTRACT

Efforts to control police discretion and improve citizen satisfaction with the police have traditionally focused on department-level variation or individual officer behavior. Studies have overlooked the importance of intra-department variation in police officer assignment and the environmental context from which citizen satisfaction with the police is determined. We address this problem by relying on an internal benchmark counterfactual method of identification that applies the probability of assignment to neighborhoods or individual officer environments for a sample of residents and police officers in Cincinnati, Ohio. The data consists of 3,000 residents surveyed in 53 neighborhoods and 133 police officers that engaged in at least 100 traffic stops. Applying the internal benchmark methodology allows us to compare the variation neighborhoods in the level of satisfaction with the police and the race distribution of individual police officer traffic stops to their appropriate reference distributions. We find that five out of nineteen identifiable neighborhoods in the city of Cincinnati appear to have significantly more unfavorable perceptions of the police, but the degree of dissatisfaction is effectively reduced to zero when these residents are compared to their appropriate statistical benchmark. We find four officers who appear to stop a larger percentage of black drivers than their statistical benchmark. Internal benchmarking methodologies provide a rigorous approach for identifying problematic neighborhoods and officers and improving police performance.

1. INTRODUCTION

The exercise of police discretion has profound consequences for citizens. Research on police behavior and the exercise of social control note a number of social processes that explain the context under which officers make decisions (see Wilson, 1968; Smith, 1984; Klinger, 1997; Walker, 1993; Worden and Shepard, 1996; Mastrofski et al., 2000; Lundman and Kaufman, 2003). At the core of the issues surrounding improving citizen satisfaction with the police and placing limits on police use of discretion is ascertaining under what contexts discretion exercised by officers is unwarranted or appears unwarranted by the public (Gould and Mastrofski, 2004; Weitzer and Tuch, 2006). Examples of unwarranted police use of discretion involve the gamut of behaviors, from the use of deadly force to the use of race as a proxy for suspicion of criminal activities in deciding whom to stop and search (Sherman, 1980; Knowles, Persico, and Todd, 2001; Persico, 2002; Alpert, MacDonald, and Dunham, 2005; Gelman, Fagan, and Kiss, 2007). Accounting for variation in citizen perceptions of the police and unfair use of discretion in routine police actions is a difficult task given that police encounters vary depending on the environmental context (Smith, 1986).

Discussions of the use of discretion by the police and citizen satisfaction in the United States often focus on racial disparities (Kennedy, 1997; Russell-Brown, 1998; Harris, 1999; Weitzer and Tuch, 2006). A National Academy of Science report noted that research is generally inadequate for answering whether race is an important factor in police discretion to use legal authority once one takes into account other aspects of police interactions, including suspect demeanor and the locale in which interactions occur (National Research Council, 2003). Similarly, the majority of methods used to assess citizens satisfaction with the police focus on departmental changes over time or variation within different segments of a city. Comparisons over time are useful but can be unduly influenced by a few high-profile media covered examples of police abuse of authority (see Weitzer and Tuch, 2006 for a review). Neighborhood environments in which the police operate can vary dramatically in large cities, such that simple mean comparisons can easily distort the degree to which citizens are dissatisfied with the police compared to how they would perceive the police in a different locale. Officer bias in the use of discretion typically focuses on comparisons of a given police department's rate of stops, searches, or arrests against some form of external benchmark, such as contraband yields, crime rates, and residential populations (Fagan and Davies, 2000; Knowles et al., 2001; Fridell, 2004). These benchmarks are all notably problematic (see Ridgeway and MacDonald, 2009).

Several decades of research in the United States and other countries note there is wide variability in how police behave within a given department (Brown, 1981; Klinger, 1997; Kane, 2002). There are, however, no widely defined empirical methods for assessing intra-neighborhood variation in satisfaction with the police or between-officer bias in routine police officer discretion. Yet, studies suggest that neighborhood context is an important determinant of police actions (Smith, 1986) and a small fraction of police officers in a given department are likely to exercise abuse of authority within specific neighborhoods (Sherman, 1978; Fyfe, 1980; Mastrofski; Ritti; and Snipes, 1994). A rigorous method for reliably assessing citizen satisfaction with the police and bias in police officer use of discretion could contribute to improving overall police performance.

Police departments face civil litigation as a result of high profile police abuse of authority incidents or class action lawsuits. The tenure of police administrators is also affected by the level of satisfaction among residents. Police managers want to minimize litigation risks and improve community satisfaction, while at the same time maintaining strict policies that ensure full enforcement of the criminal laws. There are, however, few coherent methods for assessing intra-departmental variation in citizen satisfaction with the police or detecting unfair use of police discretion.

In the present study we address this gap in the literature by relying on an internal benchmark method for detecting racial bias in police decisions to stop citizens (Ridgeway and MacDonald, 2009) and apply it to an analysis of neighborhood variation in satisfaction with the police with an individual officer analysis of routine traffic stop decisions. Specifically, we apply a potential outcomes causal model that is specified on organizational theories of police behavior and neighborhood variation in satisfaction with the police. This is the first paper to link these two strands of literature together in a single analytic method and applying to a case study of a select police department.

In Section 2 of this paper we provide background on prior research and theory on the environmental context of citizen satisfaction with the police and the exercise of police officer

discretion. Section 3 outlines our methodology for assessing intradepartmental variation in satisfaction with the police and race bias in individual police officer traffic stops. The specifications and results are presented in Section 4. Conclusions and policy implications are presented in Section 5.

2. SOCIAL CONTEXT OF POLICE DISCRETION AND PUBLIC SATISFACTION

Police discretion is one of the first aspects of the administration of justice that was systematically studied by legal scholars and social scientists (LaFave, 1965; Reiss, 1992; Walker, 1993). Research on police discretion has focused primarily on the influence that individual-level case characteristics have on the police exercise of legal control. Empirical evidence indicates that suspect demeanor, seriousness of offense, victim requests, and race of suspect and victim impact the decision that police officers make in deciding to arrest a suspect (Klinger, 1994, 1996; Lundman, 1994; Mastrofski et al., 1996; Mastrofski et al., 2000; Smith and Davidson, 1984; Smith and Visher, 1981; Worden, 1989). Scholarly debate remains about the relative importance of demeanor and race compared to other situational aspects of police-citizen interactions like the seriousness of offense (Klinger, 1994; 1996; Worden and Shepard, 1996; National Research Council, 2003). Research also suggests that an officer's experience is related to conclusions concerning a suspect's suspiciousness, crime proneness, and moral character (Werthman and Piliavin, 1967). Officer age has also been shown to be associated with police behavior. For example, empirical findings indicate that younger officers are more likely than older officers to act aggressively or use force (Alpert and Dunham, 2004). It is, however, unclear the extent to which officer age is correlated with other dynamics that may explain this association including the area of assignment, the relative antagonism of suspects, and other aspects of arrest situations.

A growing body of literature has also found that neighborhood and organizational context shape the use of police legal authority. The few rigorous empirical studies that examined the influence of organizational context on aspects of legal control compare a limited number of agencies and settings (Wilson, 1968; Smith, 1984; Mastrofski et al., 1987). Findings suggest that variations in departmental management styles and culture explain a significant amount of variation in the proclivity of police officers to exercise their arrest powers. For example, Smith's (1984) analysis of observational data on over 1,000 police-citizen encounters found that police officers in "legalistic departments" were two to three times more likely than officers in other types of departments to arrest juveniles. Mastrofski et al. (1987) also found that police officers working in small departments were more likely than officers working in large departments to follow written departmental policies regarding arresting individuals for drunk driving. Other work, however, indicates that departmental policies and organizational characteristics have only a minimal effect on officer behavior (Worden, 1995).

Geographic location plays an important role in how police exercise their use of legal authority. Bittner's (1967) early observational study of police behavior noted that the deployment of police personnel could be attributed to perceptions about the relationship between race, crime, and neighborhoods. Smith's (1986) study of 60 neighborhoods in three cities found that neighborhood characteristics were significantly related to police decisions to exercise their legal authority. Citizens who committed offenses in lower socioeconomic status neighborhoods were three times more likely to be arrested compared those who offended in high socioeconomic status neighborhoods, independent of the effects of

individual case attributes like the seriousness of offense, suspect antagonism, the victim requests arrests, and race of suspect. Police were also more likely to use or threaten force toward suspects in neighborhoods with a greater percentage of minorities and where racial (black-white) income inequality was higher. Responses to victim requests were also related to neighborhood characteristics. Police were less likely to file an official report for a crime in higher crime neighborhood when all other individual-level variables were held constant. According to this study, neighborhood characteristics may interact with police administrative styles to create different geographical patterns of police use of discretionary authority.

Prior research also indicates that satisfaction with the police and perceptions of police more generally are also conditional on neighborhood environments. Dunham and Alpert's (1988) analysis of five ethnically and racially distinct neighborhoods in Miami found a higher degree of consensus within each neighborhood in perceptions of the police than existed between neighborhoods. Predominately black lower and middle-income neighborhoods held less favorable views on the issues of police use of discretion and the average officer's level of respectful demeanor compared to neighborhoods dominated by other race and ethnic groups. Weitzer's (1999) interviews with residents in a lower-class black, middle-class black, and middle-class white neighborhoods in Washington, DC found similar perceptions regarding racially biased police practices, but each neighborhood had different explanations for this bias. Black respondents living in a lower-class neighborhood thought that law-abiding blacks were unfairly targeted by the police because of the disproportionate involvement of blacks in street crime. In contrast, black respondents living in a middle-class community did not perceive racial bias by the police in their own neighborhood (Weitzer 1999). More recent work by Weitzer and Tuch (2006) found substantial variation within African Americans across a nationally representative sample of residents in different cities and neighborhood settings. Sampson and Bartusch's (1998) analysis of satisfaction with the police in Chicago neighborhoods also indicates that neighborhood context can explain differences in attitudes toward the police. They found the level of violent crime and relative concentration of poverty between neighborhoods was associated with significantly more negative views of the police (Sampson and Bartusch 1998). Work by Reisig and Parks (2000) also found that satisfaction with the police was associated with the concentration of poverty in neighborhoods. Findings of the importance of neighborhood setting are also indicated in work by MacDonald et al. (2007) in Cincinnati, Ohio. Work by Carr et al. (2007) also indicates that distrust of the police various within different predominately black neighborhoods in Philadelphia.

The importance of neighborhood context in shaping attitudes and perceptions of the police appears to be indisputable, it is noteworthy much of prior work in this areas does not distinctly model the selection into different neighborhood environments. Instead, the majority of prior research relies on statistical modeling adjustments for neighborhood environments. Given the dissimilar neighborhood contexts to which residents select, it is important in assessing the causal relationship between neighborhood environments and perceptions of the police that one adequately accounts for selection. In the United States discussions of neighborhood context and satisfaction with the police are often discussed within the rubric of racial disparities, but it is equally important to acknowledge variation within race groups (Weitzer, 199; Weitzer and Tuch, 2006). The relationship between satisfaction and perceptions of the police must be balanced within their proper neighborhood

environment. Previous efforts have attempted to do this through the use of multilevel modeling (Sampson and Bartusch 1998; Reisig and Parks 2000). Given the selection into different neighborhood environments, it is possible that these multilevel models suffered from an insufficient number of individuals living in comparable contexts to uniquely assess neighborhood effects. Regression models, including multilevel models, have the potential to extrapolate beyond the limitations of the data if there is an insufficient balance of attributes among respondents sampled in different neighborhoods.

In addition to properly identifying neighborhood context for variations in attitudes and opinions of the police, it is also well recognized that police behave differently based on area context. Klinger (1997) explains a behavioral model of police discretion that provides one of the few theoretical explanations for assessing how officers will respond differently depending on the environmental context in which they operate. Klinger's model suggests that police assignments in most urban cities are divided into distinct administrative geographic boundaries (e.g., divisions, districts, or precincts). These administrative boundaries result in a territorial division of labor that creates "distinct community-based work groups" (p. 286). Thus, each geographic boundary or district creates its own informal set of rules that govern policing. These informal rules of conduct are shaped by the internal workgroup dynamics of the district as well as the environment that the patrol officer is working. For example, patrol officers working in higher crime areas generally have heavier workloads and also develop a sense that they are working in a "troubled" area. According to Klinger's theory of police behavior the workgroup dynamic develops such that officers working in high crime districts will be less vigilant about minor criminal infractions than officers working in low crime districts. This behavioral model provides a concise explanation for why police officers in similar circumstances will act differently depending on the geographic location in which they are situated. Officers may develop different standards of conduct on the basis of both organizational principles and the social and economic makeup of the neighborhoods they work in (Smith, 1984; 1986; Klinger, 1997; Kane, 2002). These same principles may also explain how racial bias operates in police decision-making.

Minority status is at the core discussion of police behavior and citizen satisfaction with the police because there are entrenched racial disparities in social class, geography, and crime in American society (Russell-Brown, 1998; Kennedy, 1997; Sampson and Wilson, 1995). Although research findings have produced mixed results concerning the individual-level influence of race on police behavior (National Research Council, 2003; p. 3), racial bias is at the forefront of issues regarding police use of their legal authority and the subject of tense police-community relations and civil litigation in many U.S. cities. All of the major urban riots that have occurred in the United States since the 1970s have caused by minority grievances high-profile police abuse cases.

Research on the police use of legal authority suggests that decisions are based on prior experience, time and place of assignment, and the seriousness of the incident to which the police officer is responding. Therefore, efforts to compare race bias in officer behavior should be sensitive to the variation in geographic environments to which officers are situated. Assessing race bias in police officer behavior requires an analytic approach that appropriately develops a counterfactual where officers have equivalent exposures to neighborhood

conditions. The geographic dissimilarity between officer assignments has the potential to over-identify race bias in individual officer decision-making.

Identifying Racial Bias in Police Discretion

There have been many efforts to assess whether entire police forces have racially biased practices with regard to stopping motorists or other aspects of routine police discretion. These studies have compared the percentage of drivers that officers have stopped that are black with the percentage of black residents or the percentage of blacks among not-at-fault traffic accidents (Alpert, Smith, and Dunham, 2007), assuming they represent a random sample of the population of drivers at risk for motor vehicle stops. Other approaches have involved the comparing the race distribution of drivers recorded in traffic stops to those recorded through systematic social observations (Lamberth, 1994) or examinations of the search rates of blacks and whites to the hit rates between race groups (Knowles, Persico, and Todd, 2001). More recently, Grogger and Ridgeway (2006) used changes to and from Daylight Savings Time to detect whether the ability to identify race in advance of the stop influenced the percentage of black drivers that were stopped. All of these methods focus on the police force as a whole and report whether there is or is not evidence of racially biased policing.

Some police departments and policy makers, however, argue that if there a problem of racial bias in police decisions to stop motorists exists then it stems from "a few bad apples" since a small fraction of police officers in a given department contribute to a disproportionate share of cases of abuse of authority (Sherman 1978; Christopher, 1991). If racial bias is a result of a few "bad officers" then comparisons at the department level are not useful because they do not help to identify these problem officers. Additionally, it is well recognized that a few problematic officers or a few high profile police cases of abuse of authority can transform entire neighborhoods' and communities' sense of their local police (Weitzer and Tuch, 2006).

Present Study

In the current study we build recently published work (Ridgeway and MacDonald, 2009) that constructs statistical benchmarks based on Walker's (2003) conceptualization of an internal benchmark method for comparing police officer decisions. In the current study, we expand on this methodology to compare both neighborhood variation in satisfaction with the police and racial disparities in individual officer traffic stop decisions. The methodology here provides an analytic model that tests whether problematic neighborhoods and officers can be identified, once one takes into account the geographic context of citizen exposure to the police and the daily work actions of individual police officers. We rely on a propensity score weighting method that compares the average level of satisfaction with the police among residents in a target neighborhood to an appropriate statistical benchmark of similarly situated citizens. Additionally, we compare the race distribution of drivers individual police officers have stopped in a given district, neighborhood, and time with the race distribution of drivers that other officers have stopped under similar conditions.

3. DATA AND METHOD

The data for our analysis come from a study of community-police relations in Cincinnati. Under a settlement agreement, the Cincinnati Police Department (CPD) agreed to sponsor a

citywide survey of citizens and measure their level of satisfaction with the CPD and collect data on all vehicle stops including where and when the stop took place, the reason for the stop, and the race of the driver and passengers. The following analysis is based on citizen surveys and traffic stop data that were collected in 2005. Traffic stop data were collected as part of routine police activities and included indicators of the time, place, reason, and racial characteristics of the driver stopped by the police (see Riley et al., 2005).

The citizen satisfaction survey consisted of 3,000 residents in Cincinnati that were selected via random-digit dialing (RDD) and list-assisted sampling methods with a designed quota sample to develop acceptably precise estimates of residents living in 53 defined Cincinnati neighborhoods (see Riley et al., 2005 for technical details). From this survey data we selected the observations with complete measurement on demographic and other respondent attributes of individuals to estimate satisfaction with the police across different neighborhoods in Cincinnati. The final sample was reduced to 2,167 observations with complete data across 46 neighborhoods.

Satisfaction with the police was assessed by combining the following five questions:

- How would you rate the performance of the Cincinnati Police on working with residents to address local crime problems – would you say it is excellent, good, far, or poor?
- In general, how would you rate the quality of police protection in Cincinnati – would you say it's excellent, good, fair, or poor?
- When it comes to getting its share of police services, would you say that your neighborhood gets, more than it needs, about the right amount, or not enough?
- In your opinion, would you say the Cincinnati police officers are generally very polite toward people like yourself, somewhat polite, somewhat rude, or very rude?
- How much do you trust police officers working for the Cincinnati Police Department - a lot, somewhat, a little bit, or not at all?

Each participant's responses to these five items were combined to create a single summed scale. The alpha reliability for this scale was 0.83. Higher scores on this scale reflected lower levels of satisfaction with the police in Cincinnati.

To control for the effect of race and ethnicity on perceptions of the police, respondents were categorized into black vs. white and other (other making up less than 5% of respondents). Education was measured on a five-point scale from less than high school diploma to graduate or professional degree; household income, measured on a six-point scale from $20,000 or less to $100,000 or more; whether the respondent owned or rented their home; and whether they were employed, either full or part-time.

To measure perceived neighborhood environments respondents were asked to indicate their awareness of local police, perceptions of neighborhood disorder and crime, and social cohesion. To measure exposure to police in one's neighborhood, respondents were asked whether they knew any of the police officers in their neighborhood by name or by sight. To

measure disorder, respondents indicated how often they witnessed garbage in the streets and empty beer bottles; kids hanging out on street corners without adult supervision; graffiti on walls, bus stops, and mailboxes; drug transactions, or activities that appeared to be drug dealing; and people acting disrespectfully toward police (e.g., yelling obscenities). Response options ranged from almost never to almost always. All five items were combined into a single summed scale. This scale's alpha reliability was 0.81. Higher scores on this scale indicated that respondents witnessed more disorder in their neighborhood.

Two separate items measured respondents' assessment of neighborhood crime conditions. Respondents indicated their fear of crime, by asking them "How safe would you feel being out alone in your neighborhood at night: very safe, reasonably safe, somewhat unsafe, very unsafe?" To assess the degree of actual crime exposure, respondents were asked if, during the last 12 months, they were aware of any armed robberies, murders, sexual assaults, or burglaries that occurred in their neighborhoods (yes-no). All four crime exposure items were summed together for an aggregate measure of exposure to neighborhood crime.

Finally, three separate items were used to assess the extent of neighborhood social cohesion. Respondents were asked if they participated in any neighborhood associations or activities, how often they got together with neighbors (daily to never), and how much they trusted people in their neighborhood (a lot to not at all). The average characteristics of the sample and the distribution across the 46 neighborhoods are displayed in Table 1.

Table 1. Characteristics of Sampled Citizens

Variable	Obs	Mean	Std. Dev.	Min	Max
Satisfaction with police	1971	11.333	3.371	5	19
Black	2167	.438	.496	0	1
Employed	2167	.685	.464	0	1
Education (level)	2167	3.033	1.224	1	5
Income (level)	2167	2.830	1.551	1	6
Disorder in neighborhood	2167	10.155	4.302	5	20
Neighborhood crime	2167	1.610	1.392	0	4
Fear of crime in neighborhood	2150	2.443	.957	1	4
Participation in neighborhood groups	2167	.2441	.429	0	1
Frequency of neighborhood socialization	2167	2.406	1.097	1	4
Trust of neighbors	2167	2.644	1.070	1	4
Know police by name or sight	2167	.336	.472	0	1
Married	2167	.352	.477	0	1
Home ownership	2167	.526	.499	0	1

Number of children	2162	.735	1.190	0	7
Male	2167	.383	.486	0	1
Age (years)	2167	46.471	16.721	18	105
Neighborhood	--		Percent		
Avondale	123		5.68		
Bondhill	58		2.68		
C.B.D./Riverfront	20		0.92		
Camp Washington	9		0.42		
Carthage	13		0.6		
Clifton	60		2.77		
Clifton/University Heights	64		2.95		
College Hill	83		3.83		
Columbia/Tusculum	17		0.78		
Corryville	22		1.02		
East Price Hill	116		5.35		
East Walnut Hills	23		1.06		
Evanston	40		1.85		
Fairview	33		1.52		
Fay Apartments	14		0.65		
Hartwell	28		1.29		
Hyde Park	87		4.01		
Kennedy Heights	35		1.62		
Lindwood	8		0.37		
Lower Price Hill	8		0.37		
Madisonville	71		3.28		
Mount Adams	9		0.42		
Mount Airy	65		3		
Mount Auburn	44		2.03		
Mt. Lookout	22		1.02		
Mt. Washjngton/East End	93		4.29		
North Fairmount/ English Woods	23		1.06		
Northside	69		3.18		
O'bryonville	17		0.78		
Oakley	78		3.6		
Other (Specify)	38		1.75		
Over the Rhine	56		2.58		
Paddock Hills	17		0.78		
Pleasant Ridge	56		2.58		
Riverside/Sedamsville	29		1.34		
Roselawn	42		1.94		
S.Cumminsville/Millvale	22		1.02		
Sayler Park	22		1.02		
South Fairmount	21		0.97		
Walnut Hills	50		2.31		
West End/Queensgate	49		2.26		

West Price Hill	113	5.21
Westwood	242	11.17
Winton Hills	38	1.75
Winton Place	17	0.78
Total	2,167	100

Analytic Framework

The fundamental goal of internal benchmarking is to compare the appropriate statistical benchmarks for both the level of neighborhood satisfaction with the police and racial bias in individual officer decision-making on whom to stop in a motor vehicle. It is necessary to compare a neighborhood's level of satisfaction with the police with residents who are exposed to similar environmental circumstances, and because no two neighborhoods are exactly alike we design a comparison group of residents with similar demographic backgrounds, perceptions of neighborhoods, and knowledge of local police. For constructing an internal benchmark of officers in stop decisions it is necessary to compare the percentage of minority stops for a particular officer with the percentage of minority stops of other officers patrolling the same areas at the same times. Matching in this way assures us that both the neighborhoods of interest and officers under investigation are being compared to the same set of environmental circumstances including disorder, criminal offenses, and offenders. But all matching procedures are not created equal. Direct matching to construct internal benchmarks can run low on suitable matches for some neighborhoods as well as officers who don't have simple comparison cases.

In order to increase the set of appropriate statistical benchmarks for residents and officers, we rely on comparisons of citizens and officers under similar constraints. Previous efforts to match officers to each other often relied on wide parameters, such as entire police precincts. Prior research suggests that officer behavior can vary by both the district and neighborhood in which police-citizen interactions occur. More precise measures of the location of the stop are, therefore, necessary if such propositions are true. Additionally, matching at the officer level is inappropriate because officers' assignments can vary by geography, time of the year, etc. Therefore, we develop an algorithm that matches each officer's *collection* of stops to a collection of stops made by other officers at the same time and place. Stops were matched on month, day of week, time of day, and all 53 Cincinnati neighborhoods plus eight highway segments, and the reason the officer initiated the stop.

We match the joint distribution of the neighborhood perceptions of the police as well stop features of individual officers to the joint distribution features for the set of respective comparison groups. We construct our internal benchmark by reweighting the neighborhoods and stops by the police to a potential comparison group so that the joint distributions align. Specifically,

$$f(\mathbf{x} \mid t = 1) = w(\mathbf{x}) f(\mathbf{x} \mid t = 0) \quad (1)$$

where $\mathbf{x}$ is the vector of neighborhood resident characteristics or stop features, t is an indicator for a stop involving the target neighborhood or officer, and $w(\mathbf{x})$ is the weight

function for which we will solve to equalize the feature distributions. Solving for $w(\mathbf{x})$ and applying Bayes theorem to the two conditional distributions of $\mathbf{x}$ yields

$$w(\mathbf{x}) = \frac{f(t=1 \mid \mathbf{x})}{f(t=0 \mid \mathbf{x})} K \tag{2}$$

K is a constant that does not depend on $\mathbf{x}$ and will cancel in the outcomes analyses. The probability that a citizen respondent or stop having features $\mathbf{x}$ involves the target neighborhood or officer $f(t = 1|\mathbf{x})$ we refer to as the propensity score (Rosenbaum and Rubin 1983). We will denote the propensity score for stop i as p_i. According to (2), weighting average neighborhood's level of satisfaction with the police to that made by a comparison set of residents and stops made by individual officers to a comparison set officers by $p_i/(1 - p_i)$. This weighting scheme re-aligns the distribution of comparison cases characteristics with the distribution of the target neighborhood or officer's stop characteristics. Comparison cases having features (time, location, stop reason) that are quite different from the characteristics of the target group will have propensity scores near 0 and therefore will receive weights near 0. Comparison cases with large propensity scores, on the other hand, have features that are very similar to the target neighborhood or officer's stops and will have larger weights. For more detailed analysis of propensity score weights see Hirano and Imbens (2001), Wooldridge (2002, pp. 614-621), and McCaffrey et al (2004). Through the application of propensity score weights the average of each neighborhoods level of satisfaction with the police and the average race distribution of traffic stops by individual officers becomes the only observable factor that differs between the treatment groups and their statistical benchmark.

It is also worth noting that weighted means have greater sampling variance than un-weighted means from a sample of equal size. The effective sample size (ESS) captures this increase in variance.

$$\text{ESS} = \frac{\left(\sum_{i \in \text{comparison}} w_i \right)^2}{\sum_{i \in \text{comparison}} w_i^2} \tag{3}$$

The ESS is the number of observations from a simple random sample needed to obtain an estimate of the comparison minority stop rate with precision equal to the precision obtained with the weighted comparison observations. The increase in variance captured through the ESS means that we have a more conservative statistical test of racial bias. Conceptually, the ESS offers an estimate of the number of comparison participants that are comparable to the treatment group.

We first compare the average perceptions of citizens in each of 19 out of the 46 neighborhoods for which there were at least 50 respondents to their statistical benchmark defined by their weighted average comparison group (c). Because the comparisons involve a small set of neighborhoods that in some circumstances has only 50 respondents we rely on a student's t-test for small samples.

$$t = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2}{n_1} + \frac{s_c^2}{ESS}}} \qquad (4)$$

Next we compare officers to their statistical benchmark by relying on the Bonferroni approach to adjust for multiple comparisons. However, it has long been known that the Bonderroni correction is excessively conservative. Benjamini and Hochberg (1995) pioneered the use of the false discovery rate (fdr) as an alternative multiple comparison adjustment technique. The fdr is the probability of no group difference given the value of an observed test statistic, z. Consider computing z-statistics for each officer comparison as

$$z = \frac{p_t - p_c}{\sqrt{\frac{p_t(1-p_t)}{N} + \frac{p_c(1-p_c)}{ESS}}} \qquad (5)$$

where p_t and p_c are, respectively, the proportion of stops involving minority drivers for the target and the weighted comparison officers. These z-scores have been promoted as a measure for flagging problem officers, particularly those officers exceeding 2.0 (Fridell 2004). In standard circumstances z will have a $N(0,1)$ distribution and such a cutoff could be used for testing differences in the minority stop rates. However, in a collection of 133 z-scores, each of which is correlated with the other, the empirical distribution of the z may be much wider than would be predicted by statistical theory (Efron 2006).

4. ANALYSIS AND RESULTS

Neighborhood Benchmark:

Table 2 provides an example comparison of how the distributions of a target neighborhood (Over the Rhine) and comparison group align on covariates. One can see that prior to constructing a statistical benchmark that respondents from this target neighborhood are largely different from average respondent in other Cincinnati neighborhoods. Respondents in the target neighborhood are more likely to be black, unemployed, lower income, less educated, perceive more neighborhood disorder, have higher fear of crime, express less trust of their neighbors, less likely to be married, etc. However, table 2 also shows that after matching this target neighborhood to a weighted sample of similarly situated respondents the differences in average respondent attributes is effectively reduced to zero. This sets the example for why an appropriate comparison group for neighborhood residents is necessary.

Table 2: Internal Benchmark for an Example Neighborhood (Over the Rhine)

	N=56		N=2,111		N= 246.9 [(1)]	
Variable	Target	SD	Unadjusted	Effect Size	Comparison	W Effect Size
Black, %	0.8	0.41	0.43	0.90	0.79	0.02
Employed, %	0.65	0.48	0.69	-0.08	0.63	0.04
Education (level)	2.47	1.07	3.05	-0.54	2.5	-0.02
Income (level)	1.63	1.2	2.86	-1.02	1.66	-0.02
Disorder in neigh.	15.56	4.24	10.02	1.30	15.53	0.00
Neighborhood crime	2.21	1.41	1.59	0.43	2.21	0.00
Fear of crime	2.76	1.07	2.43	0.30	2.71	0.04
Part. In neighborhood	0.25	0.44	0.24	0.02	0.25	0.00
Neigh. socialization	2.61	1.18	2.40	0.17	2.61	0.00
Trust of neighbors	1.85	0.92	2.66	-0.88	1.85	0.00
Know police, %	0.47	0.5	0.33	0.28	0.47	0.00
Married, %	0.09	0.28	0.35	-0.92	0.09	0.00
Home ownership, %	0.13	0.33	0.53	-1.21	0.13	0.00
Number of children	0.98	1.3	0.73	0.19	0.95	0.02
Male, %	0.44	0.5	0.38	0.12	0.45	-0.02
Age (years)	40.56	13.96	46.52	-0.42	40.53	0.00

[(1)]For the comparison respondents *N* represents the effective sample size.

Table 3 displays the results from the comparison between each of the 19 neighborhoods' average level of satisfaction with the police, residents in other areas of the city (unadjusted), and their adjusted comparison group. The comparisons displayed in table 3 indicates that average citizen's perceptions of the police are significantly different (t-value $p<.05$) in 10 out of the 19 neighborhoods with at least 50 respondents. Five of the 19 neighborhoods rate significantly higher in their level of dissatisfaction with the police. These neighborhoods on average have a higher percentage of black residents, more poverty, and located in higher crime sections of the city of Cincinnati. However, when we adjust these average neighborhood levels of dissatisfaction with the police to their statistical benchmark each of these five neighborhoods is no longer significantly different from zero (all t statistics drop to less than 2.0 standard deviations). These findings suggest that neighborhood variation in satisfaction with the police is driven by selection mechanisms. The results from table 3 are also displayed graphically in Figure 1.

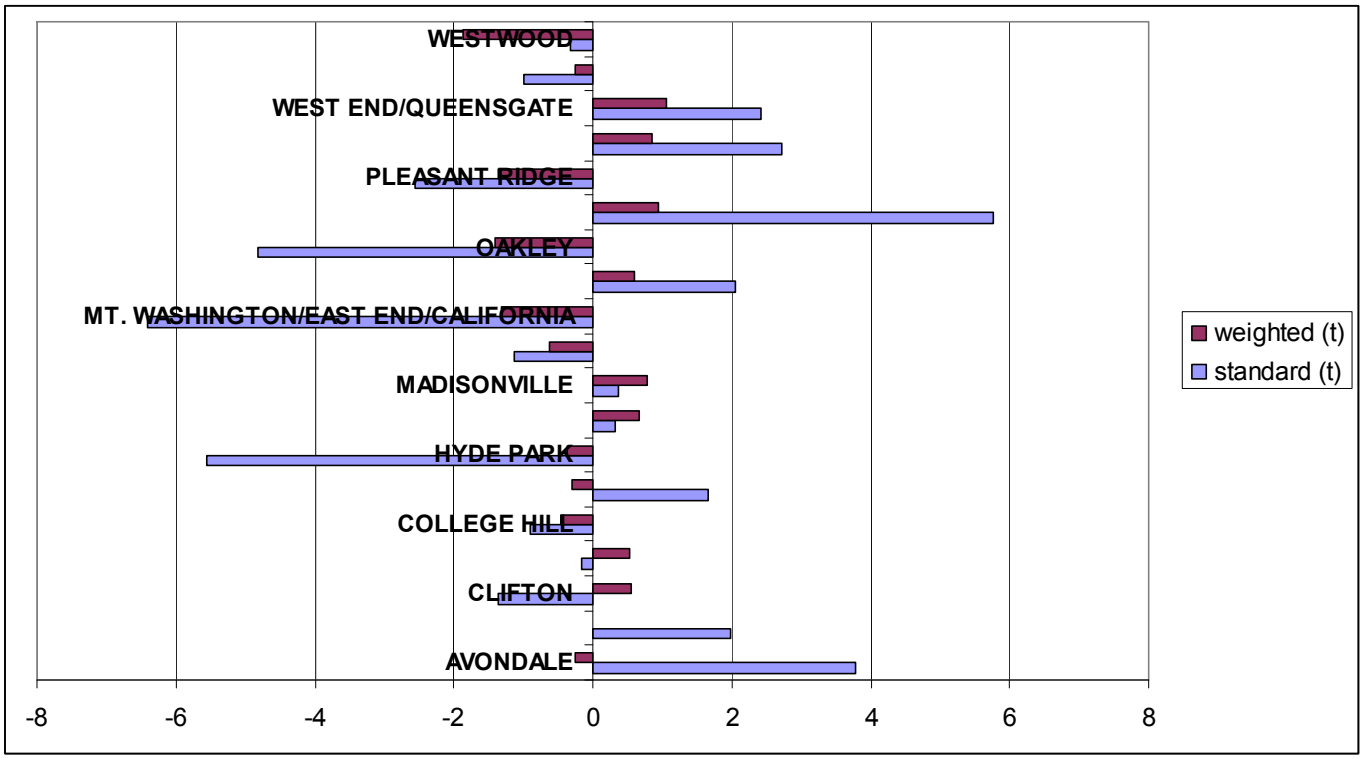
Table 3: Internal Benchmark for Neighborhoods in Cincinnati

Neighborhood	standard (t)	weighted (t)
MT. WASHINGTON/EAST END/CALIFORNIA	**-6.41**	**-1.31**
HYDE PARK	**-5.55**	**-0.38**
OAKLEY	**-4.82**	**-1.4**

PLEASANT RIDGE	**-2.56**	**-1.36**
CLIFTON	-1.35	0.55
MOUNT AIRY	-1.13	-0.63
WEST PRICE HILL	-0.98	-0.26
COLLEGE HILL	-0.91	-0.47
WESTWOOD	-0.33	-1.87
CLIFTON/UNIVERSITY H	-0.16	0.52
KENNEDY HEIGHTS	0.33	0.66
MADISONVILLE	0.36	0.79
EAST PRICE HILL	1.67	-0.29
BONDHILL	**1.98**	**0.01**
NORTHSIDE	**2.06**	**0.6**
WEST END/QUEENSGATE	**2.42**	**1.07**
WALNUT HILLS	**2.73**	**0.86**
AVONDALE	**3.79**	**-0.26**
OVER THE RHINE	**5.77**	**0.94**

Note: Neighborhoods in bold significantly different at p<.05 level prior to comparison to statistical benchmark.

Figure 1: Standard and Adjusted Average Difference between Perceptions of CPD by Neighborhood



Note: 1,568 respondent perceptions of CPD in 19 neighborhoods against the standard (t) average difference and their weighted (t) statistical benchmark.

Officer Benchmark:

Table 4 presents an example of the internal benchmark for a particular target CPD officer. This officer made 111 stops in 2005, most of them were during the night shift, particularly on Mondays and Thursdays, and perhaps this officer was off during much of June and July. Most of the officer's stops occurred in neighborhood J (49%) and neighborhood K (33%). This officer seems particular attune to equipment violations. This officer's stops involved black drivers 71% of the time. To assess whether 71% is a reasonable fraction we construct a weighted set of stops for comparison.

Table 4: Example of Internal Benchmarking for an Example Officer

Variable		Target	Comparison	Weighted effect size
	N =	111	571.3[(2)]	
Time	(12-4pm]	0.09	0.09	0.01
	(4-8pm]	0.57	0.56	0.01
	(8pm-12am]	0.34	0.35	-0.02
Day	Mon	0.20	0.20	0.00
	Tue	0.12	0.11	0.02
	Wed	0.12	0.12	-0.00
	Thu	0.20	0.21	-0.03
	Fri	0.14	0.14	-0.01
	Sat	0.11	0.11	-0.01
	Sun	0.13	0.12	0.03
Month	Jan	0.12	0.12	0.01
	Feb	0.14	0.15	-0.02
	Mar	0.07	0.07	-0.01
	Apr	0.06	0.06	0.00
	May	0.08	0.07	0.05
	Jun	0.03	0.03	-0.03
	Jul	0.04	0.04	-0.02
	Aug	0.10	0.10	0.00
	Sep	0.06	0.06	0.03
	Oct	0.04	0.05	-0.03
	Nov	0.14	0.14	0.01
	Dec	0.11	0.11	-0.01
Neighborhood[(1)]	H	0.01	0.01	-0.01
	I	0.01	0.01	-0.01
	J	0.49	0.48	0.02
	K	0.33	0.34	-0.02
	L	0.05	0.05	0.01
	M	0.11	0.11	-0.01
Stop reason	Equipment	0.64	0.63	0.01
	Moving	0.26	0.27	-0.01
	Other	0.10	0.10	-0.00

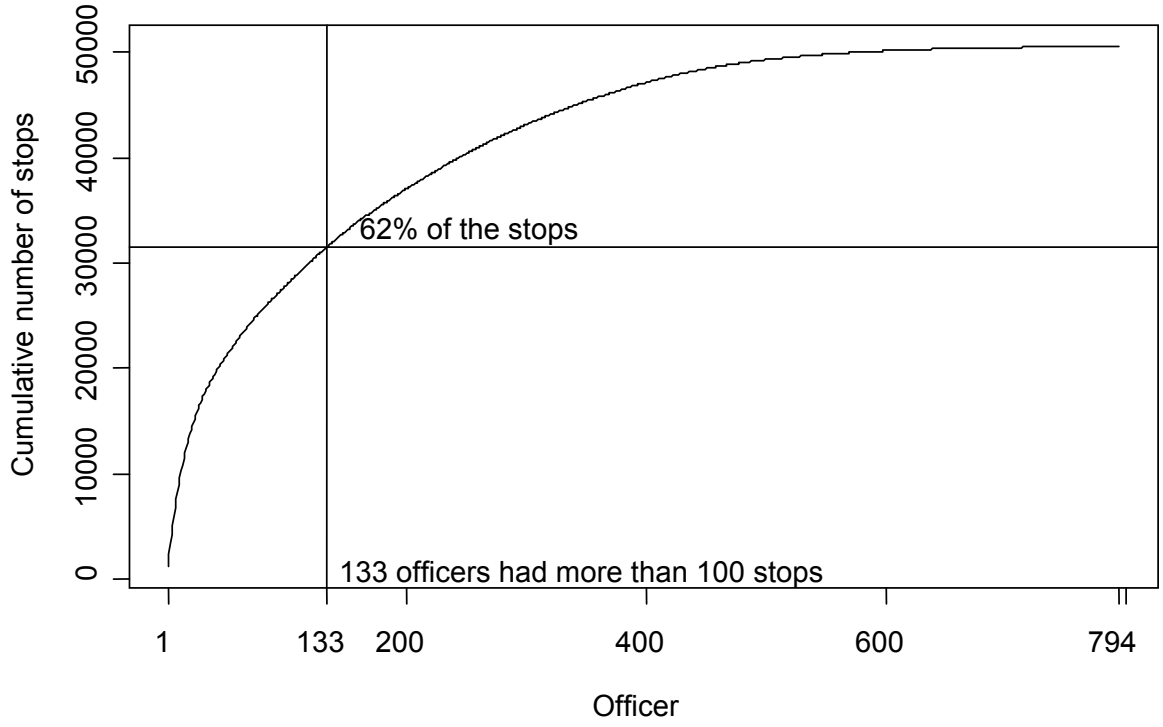
Outcome	Stops involving black drivers	71%	46%

[(1)]The neighborhoods have been given random letter codes to mask the officer's identity.
[(2)]For the comparison stops *N* represents the effective sample size.

The "Comparison" column in shows the distribution of features for stops made by other officers weighted as in (2). There were 74 officers that made of 90% of the weighted comparison stops. The comparison stops have almost exactly the same distribution of features as the target officers. They were made in the same places, times of day, days of the week, months of the year, and for the same reason. Importantly, we created the weighting without looking at the race of the drivers involved in the stops. This mitigates the risk of manipulation of the form of the covariates in the model in order to obtain a significant race effect. Those weighted comparison officers had stops involving black drivers 46% of the time. Even though the stops occurred at the same places and times as the target officer, the rate of black drivers among those stopped was much lower than the target officer.

For the actual analysis we selected all CPD officers with more than 100 reported stops in 2005 for the analysis. A total of 133 officers exceeded 100-stop cutoff, which focuses the analysis on those officers most frequently interacting with drivers in Cincinnati. It also assures us that we have at least a minimum level of statistical power for detecting differences if they exist. Figure 2 shows the distribution of the number of stops by officer. These 133 officers amount to only 16 percent of the CPD officers that reported a stop 2005 but account for 62 percent of all the 2005 traffic stops.

Figure 2: Cumulative Number of Stops by Officer



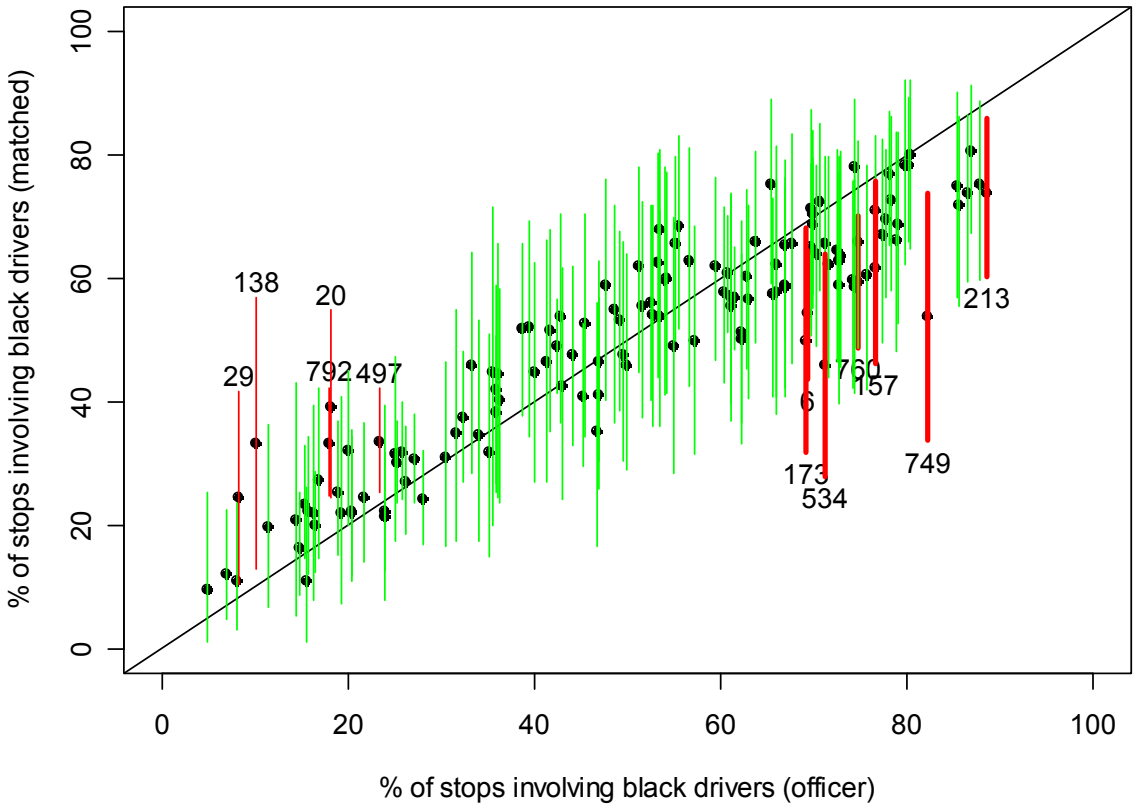
Note: 794 CPD officers made a total of 50,724 analyzable stops. 16% of the police force (133 officers) was responsible for 62% of the stops. Each of these 133 CPD officers conducted more than 100 stops.

Figure 3 shows a graphical representation of the results. Each solid dot represents one of the 133 officers. The horizontal axis indicates the percentage of stops that the officer made that involved a black driver. The vertical axis is the same percentage of black drivers among the matched stops. In the absence of differences between officers, all of the dots would line up on the diagonal line. We expect some variability and the vertical lines in Figure 3 indicate a range of percentages that are plausible if the particular officer were not profiling. The extremes of the ranges are computed as the 2.501 and 97.599 percentiles of the Beta-Binomial with parameters N = the number of target officer stops, α = the effective sample size of control group stops of black drivers, and β = the effective sample size of control group stops of non-black drivers.

The selected quantiles are essentially a Bonferroni correction for the 133 comparisons. Seven of the officers marked with thick vertical lines on the right side of the Figure 3 appear to have stopped a larger percentage of black drivers than other officers making stops at the same times and places. There may be a legitimate reason for these differences, but our analysis has

confirmed that it cannot be due to a unique exposure to a racial distribution of offenders by place, time, reason for the stop, or random chance.

Figure 3: Internal benchmark comparisons for the 133 CPD officers with more than 100 vehicle stops during 2005



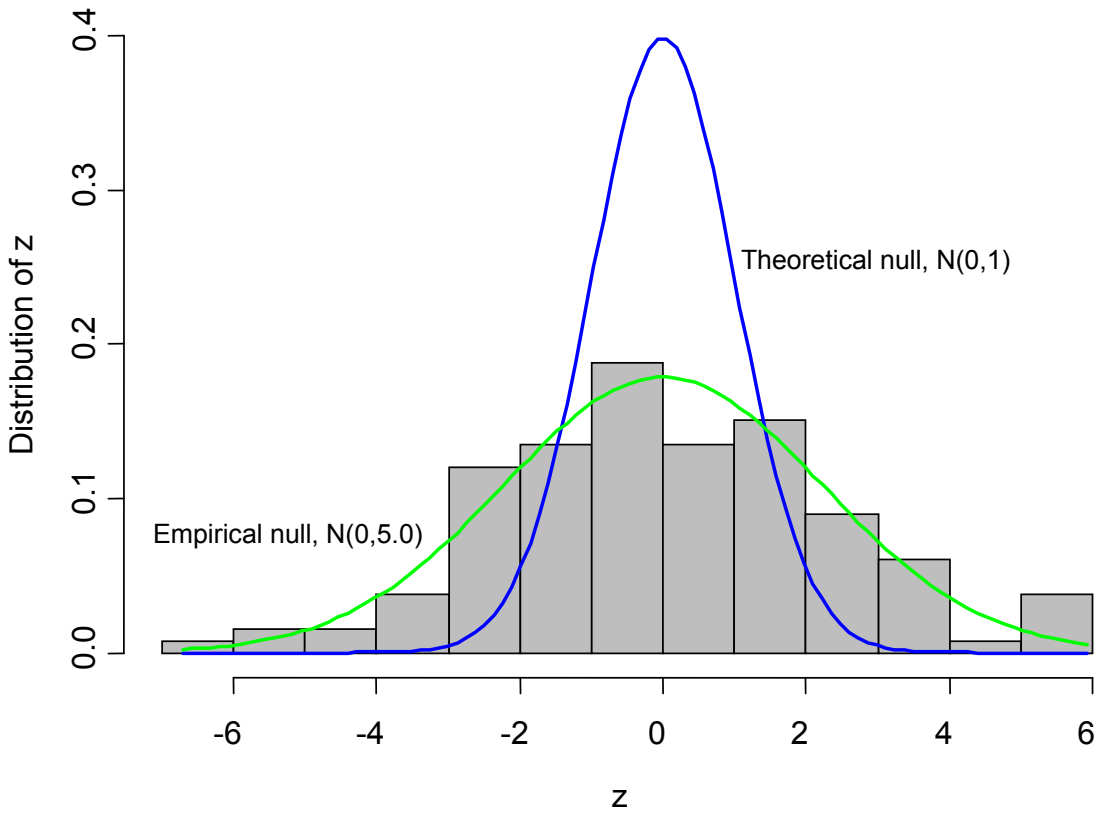
Note: The ID numbers are random identifiers and do not represent a CPD badge number. The vertical lines represent 99.9 percent confidence intervals, equivalent to a 0.05 Type I familywise error rate, to account for the 133 comparisons.

Figure 4 displays a histogram of the 133 z-scores computed for the CPD data. Overlaid on the histogram is the theoretical null for independent tests, $N(0,1)$, and an estimate of the empirical null with mean 0 and standard deviation estimated as the inter-quartile range divided by 1.349 (Tukey 1977). The mass for $z > 5$ is of particular concern. This mass corresponds to five of the 133 CPD officers, all of whom were also identified in Figure 3. We can derive the probability of an officer being problematic as

$$
\begin{aligned}
P(\text{problem} \mid z) &= 1 - P(\text{no problem} \mid z) \\
&= 1 - \frac{f(z \mid \text{no problem}) f(\text{no problem})}{f(z)} \\
&\geq 1 - \frac{f_0(z)}{f(z)}
\end{aligned}
\tag{4}
$$

where $f_0(z)$ is the distribution of z for non-problem officers and $f(z)$ is the distribution of z for all officers (Efron 2004). If the fraction of problem officers is small (less than 10%) then the bound in the last line of Figure 4 is near equality. We estimate $f_0(z)$ with the empirical null assuming a mean 0 normal but with a variance estimated using only the central data of the distribution. We estimate $f(z)$ with the histogram shown in Figure 4. For the five officers in the extreme right tail their probabilities of being a problem officer range from 70% to 86%.

Figure 4: Distribution of 133 *z*-statistics and their reference distributions



At this stage, we do not know whether there is a problem with these five officers, as we can only detect a disparity up to the resolution of the data. For example, Officer 213's assignment might have been to a particular corner frequented more by black drivers than non-black drivers, but the resolution of our analysis limits us to neighborhood-level analyses.

5. CONCLUSION

This paper was predicated on the notion that methods for analyzing the public's level of satisfaction with the police and assessing police officer discretion need to be sensitive to the environmental context in which citizens live and officers carry out their daily activities. Prior research suggested that citizens often view the police differently depending on their neighborhood context and that officers behave differently depending on the location in which they are working. The internal benchmark method controlled for these potential confounds on citizens' perceptions of the CPD and individual officer behavior in traffic stop decisions. By comparing the average level of dissatisfaction with the police in each neighborhood to a statistical benchmark of similarly situated residents we were able to remove the confounding effects of neighborhood context and empirically examine the independent effect of neighborhood setting on citizen attitudes. Similarly, by comparing police officers' decisions to stop vehicles to officers patrolling the same areas at the same times we are able to construct a reasonable benchmark of what the race distribution of their stops should be. After using propensity score weighting to compare neighborhood variation in satisfaction with the police we find that these differences are largely explainable based on selection into different neighborhoods. By comparing the racial distribution of the vehicle stops of officers 133 officers, and controlling for false-discovery, we found five officers who appeared to be stopping a significantly larger fraction of black drivers when compared with stops made by other officers made at the same time and place.

The results from this study suggest that "problematic neighborhoods" and "problematic officers" are identifiable when one conditions on the environmental context of citizen exposure to the police and individual officer work assignments. Importantly, the differences identified in the current study are not the result of chance differences in the exposure to unique times or places in Cincinnati. This analysis suggests that the work performance of the five officers identified by our approach should be evaluated by police management.

While this analysis is limited to within-agency variation in citizen perceptions of the police and individual officer differences in traffic stop decisions, we think this is an important avenue for future legal and empirical inquiry on managing police discretion and improving citizen satisfaction. The internal benchmark approach used here moves the debate away from discussions about the relevant aggregate benchmark for residents (since no two neighborhood are exactly alike) and officers and focuses on providing a rigorous method for comparing neighborhoods and officers to an appropriate statistical benchmark. Importantly, this approach can be applied over time to monitor the performance of departments in improving the average resident's satisfaction with the police in different neighborhood environments and use of routine day-to-day discretion among police officers. After all, a few problematic officers can change the sentiments of an entire neighborhood. Close monitoring and scrutiny against a defensible statistical benchmark could help improve the professionalism of the police and produce more effective and human police practices.

REFERENCES

Alpert, G. and R. Dunham (2004). *Understanding Police Use of Force*. New York: Cambridge University Press.

Alpert, Geoffrey, John M. MacDonald, and Roger Dunham. (2005). "Police Suspicion and Discretionary Decision Making During Citizen Stops." *Criminology* 43: 407-434.

Alpert, G. P., Smith, M. R., and Dunham, R. G. (2007). "Investigating Racial Profiling by the Miami-Dade Police Department: A Multimethod Approach," Criminology and Public Policy 6: 25-56.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society, Series B* 57:289-300.

Bittner, E. (1967). The Functions of Police in Modern Society. Washington, DC: National Institute of Mental Health.

Brown, M.K. (1981). Working the Street: Police Discretion and the Dilemmas of Reform. New York: Russell Sage Foundation.

Carr, P., L. Napolitano, et al. 2007. "We never call the cops and here is why: A qualitative examination of legal cynicism in three Philadelphia neighborhoods." Criminology 45(2): 445-480.

Christopher, Warren. (1991). Report of the Independent Commission on the Los Angeles Police Department. Available at http://www.parc.info/client_files/Special%20Reports/1%20%20Chistopher%20Commision.pdf

Cincinnati Police Department, "Statistics," undated Web site. Online at http://www.cincinnati-oh.gov/police/pages/-4258-as of November 8, 2005.

Efron, B. (2004). "Large-scale simultaneous hypothesis testing: the choice of a null hypothesis," *Journal of the American Statistical Association* 99:96-104.

Efron, B. (2006). "Correlation and Large-Scale Simultaneous Significance Testing," Working paper available at http://www-stat.stanford.edu/~brad/papers.

Fagan, J. and G. Davies (2000). "Street Stops and Broken Windows: Terry, Race and Disorder in New York City," *Fordham Urban Law Journal* 28:457-504.

Fridell, L. (2004). *By the Numbers: A Guide for Analyzing Race Data from Vehicle Stops*, Washington, D.C.: Police Executive Research Forum.

Fyfe, J. (1980). "Geographic Correlates of Police Shootings: A Microanalysis," *Crime and Delinquency* 17: 101-113.

Gelman, A., Fagan, J., and Kiss, A. (2007), ''An Analysis of the New York City Police Department's ''Stop-and-Frisk'' Policy in the Context of Claims of Racial Bias,'' Journal of the American Statistical Association, 102, 813–823.

Gould, J. and S. Mastrofski (2004). "Suspect Searches: Assessing Police Behavior under the U.S. Constitution," *Criminology & Public Policy* 3:315–361.

Grogger, J., and Ridgeway, G. (2006), ‘‘Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness,’’ *Journal of the American Statistical Association*, 101(475), 878–887.

Kane, R. (2002). “The Social Ecology of Police Misconduct.” *Criminology* 40:867-896.

Kennedy, R. (1997). *Race, Crime, and the Law*, New York: Pantheon Books.

Klinger, D.A. (1994). “Demeanor or Crime? Why ‘Hostile’ Citizens Are More Likely To Be Arrested,” *Criminology* 32:475-493.

Klinger, D.A. (1996). “More on Demeanor and Arrest in Dade County,” *Criminology* 34: 61-82.

Klinger, D.A. (1997). “Negotiating Order in Patrol Work: An Ecological Theory of Police Response to Deviance,” *Criminology* 35(2):277–306.

Knowles, J., Persico, N., and Todd, P. (2001), ‘‘Racial Bias in Motor Vehicle Searches: Theory and Evidence,’’ *The Journal of Political Economy*, 109,203–229.

LaFave, W.R. (1965). *Arrest: The Decision to Take a Suspect into Custody*. American Bar Foundation.

Lamberth, J. (1994). “Revised statistical analysis of the incidence of police stops and arrests of black drivers/travelers on the New Jersey turnpike between exits or interchanges 1 and 3 from the years 1988 through 1991,” Technical report, Temple University, Department of Psychology.

Massey, D.S. and N.A. Denton (1993). *American Apartheid: Segregation and the Making of the Underclass*, Cambridge, Mass.: Harvard University Press.

McCaffrey, D.F., G. Ridgeway, and A.R. Morral (2004). “Propensity Score Estimation with Boosted Regression for Evaluating Causal Effects in Observational Studies,” *Psychological Methods* 9(4):403–425.

Reiss, A.J. Jr. (1992). “Police Organization in the Twentieth Century,” in Michael H. Tonry and Norval Morris, eds., *Modern Policing: Crime and Justice: A Review of Research* 15, Chicago: University of Chicago Press, pp. 51–98.

Ridgeway, G. (2006). “Assessing the Effect of Race Bias in Post-Traffic Stop Outcomes Using Propensity Scores,” *Journal of Quantitative Criminology* 22(1):1-29.

Ridgeway, Greg, and John M. MacDonald. (2009). “Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops.” *Journal of the American Statistical Association* 104, No. 486.

Riley, K.J., S. Turner, J. MacDonald, G. Ridgeway, T. Schell, J. Wilson, T.L. Dixon, T. Fain, D. Barnes-Proby, B. Fulton (2005). *Police-Community Relations in Cincinnati*. RAND TR-333 http://www.rand.org/pubs/technical_reports/TR333.

Rosenbaum, P.R. and D.B. Rubin (1983). “The central role of the propensity score in observation studies for causal effects,” *Biometrika* 70:41-55.

Sampson, Robert J. and Dawn Jeglum Bartusch. (1998). "Legal Cynicism and (Subcultural?) Tolerance of Deviance: The Neighborhood Context of Racial Differences," *Law and Society Review* 32(4):777–804.

Sampson, R.J. and W.J. Wilson (1995). "Toward a Theory of Race, Crime, and Urban Inequality," in *Crime and Inequality* (John Hagan and Ruth D. Peterson, eds.), Stanford, Calif.: Stanford University Press, pp. 37–54.

Sherman, L.W. (1978). *Scandal and Reform: Controlling Police Corruption*. Berkeley:University of California Press.

Lundman, R.J. (1994). "Demeanor or Crime? The Midwest City Police-Citizen Encounters Study," *Criminology* 32: 31-656.

Lundman, R.J. and R.L. Kaufman. (2003). "Driving While Black: Effects of Race, Ethnicity, and Gender on Citizen Self-Reports of Traffic Stops and Police Actions," *Criminology* 41:195-220.

Mastrofski, S.D., R.R. Ritti. and Hoffmaster, D. (1987). "Organizational determinants of police discretion: the case of drinking and driving," *Journal of Criminal Justice* 15:387-402.

Mastrofski, S.D., R.R. Ritti, and J.B. Snipes (1994). "Expectancy Theory and Police Productivity in DUI Enforcement," *Law and Society Review* 28:113-148.

National Research Council. (2003). *Fairness and Effectiveness in Policing: The Evidence*. Washington, DC: The National Academies Press.

Novak, K. (2004). "Disparity and Racial Profiling in Traffic Enforcement," *Police Quarterly* xx: 65-96.

Reisig, Michael D. and Roger B. Parks. (2000). "Experience, Quality of Life, and Neighborhood Context: A Hierarchical Analysis of Satisfaction with Police." *Justice Quarterly* 17(3):607–630.

Riley, K. Jack, Susan Turner, John MacDonald, Greg Ridgeway, Terry Schell, Jeremy Wilson, Travis L. Dixon, Terry Fain, Dionne Barnes-Proby, and Brent Fulton. 2005. *Police-Community Relations in Cincinnati.* RAND Corporation, Santa Monica, CA.

Russell-Brown, K. (1998). *The Color of Crime: Racial Hoaxes, White Fear, Black Protectionism, Police Harassment and Other Macro-Aggressions*. New York: New York.

Smith, D. (1984). "The Organizational Aspects of Legal Control," *Criminology* 22:19-38.

Smith, D. (1987). "Police Response to Interpersonal Violence: Defining the Parameters of Legal Control," *Social Forces* 65:767-782.

Smith, D. (1986). "The Neighborhood Context of Police Behavior," in *An Annual Review of Research*. pp. 313-341, Vol. 6. (A.J. Reiss and M. Tonry, eds). Chicago: University of Chicago Press.

Smith, D. and Laura Davidson (1984). "Equity and Discretionary Justice: The Influence of Race on Police Arrest Decisions," *Journal of Criminal Law* 75: 234-249.

Smith, D. and C. Visher (1981). "Street Level Justice: Situational Determinants of Police

Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading: Addison-Wesley.

Walker, S. (1993). *Taming the System: The Control of Discretion in Criminal Justice, 1950-1990*. Oxford: Oxford University Press.

Walker, S. (2003). *Internal Benchmarking for Traffic Stop Data: An Early Intervention System Approach*, Discussion paper available at http://www.policeforum.com/internalb.pdf.

Weiss, A. and A. Grumet-Morris (2005). *Illinois traffic stop statistics act report for the year 2004*. Technical report, Northwestern University Center for Public Safety, Evanston, IL. http://www.dot.state.il.us/trafficstop/2004summary.pdf.

Weitzer, R. (1999). Citizen's perceptions of police misconduct: Race and neighbourhood context', *Justice Quarterly* 16: 819-846.

— 2002. 'Incidents of police misconduct and public opinion', *Journal of Criminal Justice* 30: 397-408.

Weitzer, R., and Tuch, S.A. (2006). Race and policing in America: Conflict and reform. New York: Cambridge University Press.

Wilson, W.J. (1987). *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, Chicago: University of Chicago Press.

Wilson, J.Q. (1968). *Varieties of Police Behavior*. Cambridge: Harvard University Press.

Worden, R.E. (1995). "Police Officers' Belief Systems: A Framework for Analysis," *American Journal of Police* XIV: 49-81.

Worden, R.E. (1989). "Situational and Attitudinal Explanations of Police Behavior: A Theoretical Reappraisal and Empirical Assessment," *Law and Society Review* 23: 667-711.

Worden, R.E. and R. Shepard (1996). "Demeanor, Crime, and Police Behavior: A Reexamination of the Police Services Study Data," *Criminology* 34: 83-105.