Measuring Disproportionate Treatment in Policing: One Department's Experience

Journal of Criminal Justice and Law: Official
Journal of the Law and Public Policy Section
of the Academy of Criminal Justice Sciences
Volume 2, Issue 2, pp. 84-101 (2018)

Heidi S. Bonner and Michele Stacey¹

Abstract

Detecting disparate impact in policing practice is essential, but those tasked with developing such inquiry face significant challenges in designing analyses. Racial profiling research has also been plagued by an inability to gain consensus regarding valid comparison groups – commonly termed the “denominator problem” in benchmarking analyses. The current inquiry details the process and outcome of using the disproportionality index (Dolan Consulting Group, 2016) to investigate several enforcement actions at a midsize department in the southeastern United States. The findings highlight the importance of combining the appropriate benchmark with the appropriate level of analysis, and the need for more scholarly inquiry into disproportionate treatment in a variety of law enforcement outcomes. The correlation between selected geographical locations and the disproportionality index in the assessment of disparity during traffic stops, field contacts, and arrest is discussed.

Keywords

Disproportionate impact, law enforcement, disparity, racial and ethnic minority, traffic stops, searches, field contacts, arrest

It is no secret that the nature of police contact with the community can have a significant influence on how the public views law enforcement (Worrall, 1999; Tyler & Wakslak, 2004; Taniguchi et al., 2017). Thus, detecting disparate impact in policing practice is essential, but researchers and practitioners tasked with developing such inquiry face significant challenges in designing analyses. Police-community encounters are intricate, and data collected about these interactions can never

¹ Thomas Harriot College of Arts and Sciences, Department of Criminal Justice, East Carolina University, Greenville, North Carolina, USA

Corresponding Author:

Heidi S. Bonner, Thomas Harriot College of Arts and Sciences, Department of Criminal Justice, East Carolina University, Greenville, North Carolina, 27858, USA
E-mail: bonnerhe@ecu.edu

The authors wish to express their appreciation to the blind reviewers of the original manuscript for their comments, which were helpful during preparation of the final version of the manuscript.

fully capture the complexity involved. Simplifying the data for the purpose of analysis can sometimes distort results which may in turn result in misguided policy intervention (Worden, McLean, & Wheeler, 2012). There have been substantial strides made in such research made over the last 15 years, but these advances have been hampered by the “denominator problem,” or an inability to gain consensus on valid comparison groups. Further, such research remains plagued by methodological limitations, including an inability to draw cause and effect inferences (Smith, Rojek, Petrocelli, & Withrow, 2017).

This research uses a different approach, a disproportionality index (DI; Dolan Consulting Group, 2016), to present a case study of the process and outcome of investigating several enforcement actions.¹ The DI is calculated by dividing the percentage observed of a particular racial category for a particular enforcement activity by the percentage expected. Previous research has compared benchmarks with actual enforcement activity and reported the results as a ratio or rate. The DI is an intuitive means of describing the degree (if any) of disparity that can be utilized across numerous enforcement activities. Although Smith et al. (2017) called for improved methodologies that could address whether observed disparities were the result of bias or discrimination, we are unable to do so with the current inquiry. However, although we cannot know its scope, training on the DI is being received by law enforcement agencies and, for many, the guidelines represent a tangible way to measure disproportionality in several enforcement decisions using data that are (or should be) available. While the measure of interest, the disproportionality index (DI), has been used in other contexts there is not (as far as we are aware) any empirical test in the disproportionate policing literature. Thus, this research uses the DI to measure disproportionality in one department to (1) review the use of alternate benchmarks for a variety of outcomes, (2) evaluate the utility of the DI as a form of analysis, and (3) consider the feasibility of measuring disparate impact in a typical midsize department in terms of geographical distribution of crime, and the availability of and access to data needed. Our findings have implications for how departments should test, and ultimately communicate, findings regarding disparate impact in policing.

Bias, Disproportionality, and Benchmarking

Prior research that has measured disparities in policing has hypothesized that disproportionality “may result from several mechanisms: racial animus or prejudice, the use of race as a shorthand for perceived criminal propensity (racial profiling), cognitive bias and stereotyping, and/or differential police deployment” (Tomaskovic-Devey, Mason, & Zingraff, 2004; Smith et al., 2017, p. 167). The term “bias” encompasses both implicit (subconscious biases that may create stereotypes affecting behavior) and explicit (conscious animosity toward a group) bias, and both types have been posited as explanations for disparities in enforcement action (Lum & Wu, 2017). Offending patterns that differ by race (Smith et al., 2017; Lum & Wu, 2017) and agency choices about where and when to police (if focused on places with high concentrations of people of color) (Lum & Wu, 2017) may result in unequal (but nonbiased) enforcement (Smith et al., 2017).

¹ Fridell (2005b) first proposed a similar index of disparity.

Although much previous research has referenced the terms “bias” and “bias policing” (or bias-based policing), it is important to state at the outset that disproportionality does not necessarily indicate bias. Early analyses of disproportionate impact focused on responses to claims of racial bias and furthered evaluation of bias-based policing. However, use of the word “bias” presupposes the motivation behind enforcement behavior. Even if a disproportionate impact is found, it cannot be assumed that it is due to racial prejudice on the part of the officer (Engel, Calnon, & Bernard, 2002). Thus, in the review of the literature, we use the term “disparate impact” or “disproportionality” throughout to reflect this understanding – prior literature that refers to bias has been changed.

In a systematic review, Smith et al. (2017) succinctly summarized the history of research on racial disparity in police enforcement. Initial interest in the topic started in the 1960s (Black, 1971; Black & Reiss, 1970; LaFave, 1965; Reiss, 1971), and racial profiling (which focused on disproportionality in traffic and pedestrian stops) was introduced in the late 1990s (Engel & Calnon, 2004; Smith & Petrocelli, 2001; Smith et al., 2004; Withrow, 2006). More recent high-profile uses of lethal force by police have led to an increased scrutiny of disproportionality in enforcement decisions, including the creation of the President’s Task Force on 21st Century Policing (President’s Task Force on 21st Century Policing, 2015; Smith et al., 2017).

Fridell (2005a) notes two criteria for taking viable measurements of data related to disparate treatment: first, that “they be linked to the race/ethnicity of the suspect or perpetrator” and second, that the “measures reflect as closely as possible *actual crime* as opposed to crime responded to by police” (p. 9). In order to effectively conduct research of disproportionality, it is essential that a baseline comparison group (often called a benchmark) be established. Benchmarking is the “process of measuring data against an established standard for the purpose of evaluation or judgment” (International Association of Chiefs of Police, 2006, p. 178), and benchmarks are used to determine whether disparity exists in enforcement decisions. An effective benchmark should be scientifically credible, have practical value, be politically sound (Walker, 2003), should account for any measurement challenges that exist within the jurisdiction under study (Withrow & Williams, 2015; Smith et al., 2017), and should indicate the race or ethnic proportion if no disparate treatment exists. However, the development of effective benchmarks that reflect the population at risk for a particular enforcement activity has been fraught with difficulty, and this problem can be considered the most controversial aspect of research on disproportionate impact (Withrow, 2006). The following sections summarize the existing research utilizing benchmarking analyses for several law enforcement outcomes.

Measuring Disproportionality in Traffic Stops

Most of the existing work on disproportionate treatment in policing has looked at traffic stops. Traffic stops are especially important to investigate because they hold “the greatest potential for police racial bias, or perceptions of it” (Fridell, Lunney, Diamond, & Kubu, 2001, p. 122). Early analyses of stops (which focused exclusively on potential disparity related to race) compared those who were stopped to the residential population. However, using Census data as a benchmark is inherently misleading because it does not appropriately consider those who do not drive, those in the driving population who are not residents, or variations in driving behavior (Farrell, Rumminger,

& McDevitt, 2004; Fridell, 2004; Engel & Calnon, 2004; Tillyer, Engel, & Cherkauskas, 2010; Worden, McLean, & Wheeler, 2012; Baumgartner, Epp, & Shoub, 2018). Some researchers weighted Census data to reflect the driving population more accurately (Novak, 2004; Rojek, Rosenfeld, & Decker, 2004; Smith & Petrocelli, 2001; Walker, 2001), but this practice is also problematic (Walker, 2001; Withrow, 2006; Smith et al., 2017).

Other studies have utilized field research to more accurately determine the racial composition of the driving population to generate more valid benchmarks (Lange, Blackman, & Johnson, 2001; Zingraff, Smith, & Tomaskovic-Devey, 2000; Smith et al., 2004). Such field observations, however, are expensive with limited geographical coverage, which limits their generalizability (Engel & Calnon, 2004; Taniguchi et al., 2017). Further, such approaches, although an improvement, still do not appropriately represent who is at risk of being the subject of a particular enforcement action (Fridell, 2005b; Ross, Fazzalaro, Barone, & Kalinowski, 2015).

Other options for benchmarks include DMV data and blind enforcement mechanisms (e.g., red light cameras, radar, air patrol) (Fridell, 2005a). Each of these, however, is not without limitations. Using DMV data is preferable if license data can be linked to racial and ethnic data. However, DMV data cannot assess who is driving in a specific jurisdiction, and some blind enforcement methods are limited in geographical scope (e.g., specific intersections). Comparing high-discretion to low-discretion stops, or daylight versus darkness stops, can expand the geographical area under consideration. Comparing stops to accident data, for example, has been proposed as a strong benchmark because accidents more accurately reflect the driving population and are less likely to be statistically biased in racial/ethnic composition (Alpert, Smith, & Dunham, 2004; International Association of Chiefs of Police, 2006; McDevitt & Iwama, 2016). Researchers have utilized data on not-at-fault drivers in two-vehicle crashes (Alpert et al., 2004) and at-fault and not-at-fault drivers in accident records (Withrow & Williams, 2015). Although crashes are a stronger benchmark than the others used for stops, the use of crashes assumes a lack of variation in accidents and in how (or if) they are reported (Taniguchi et al., 2017). The majority of studies have shown that Black motorists are disproportionately stopped regardless of the benchmark used (see Smith et al., 2017, for a review), although two have determined that they are not overrepresented (Lamberth, 2003; Withrow, 2003).

Veil of Darkness

The veil of darkness (VOD) approach, proposed by Grogger and Ridgeway (2006), offers a potential solution to the problem of determining appropriate benchmarks in the assessment of racial profiling in traffic stops. The basic premise is that police are less likely to know the race of a motorist before deciding to make a stop after dark than they are during daylight, and a test for racial profiling can be done by comparing the racial breakdown of stops made during the day to stops made after dark. Thus, the VOD is an analytical approach that uses changes in natural lighting to assess disproportionality in traffic stops. Specifically, the “inter-twilight time” (the period during the year between the earliest end of civil twilight and the latest end of civil twilight) serves as the analysis period. Theoretically, restriction of the analysis to this period adds in natural controls for the quality, quantity, and location of drivers. If there were no disparity, the proportion of people of color stopped during daylight would be the same as the proportion stopped during darkness.

Stopping a significantly greater proportion during the day than at night suggests disproportionate impact because people of color are being stopped more often when an officer can determine the race of the driver.

Since the VOD method was proposed, several studies have utilized it to assess racial profiling in traffic stops (e.g., Grogger & Ridgeway, 2006; Ridgeway, 2009; Ritter & Bael, 2009; Ross et al., 2015; Worden et al., 2012). In 2017, Taniguchi et al. noted that the traditional VOD approach does not account for officer-level variation in stops, does not consider the intersection of sex and race in disparate impact, and does not explore disproportionality across different units. When they extended the VOD approach to account for these limitations, they found that the traditional approach showed no disproportionate impact, but the expanded analysis indicated that Black drivers were disproportionately affected by stops and that the overrepresentation was confined specifically to Black males. Further, racial disproportionality varied considerably between units and across officers. The authors additionally found that disproportionality was present even when they restricted their sample to control for seasonality in traffic patterns.

Horrace and Rohlin (2016) also attempted to improve the original VOD test by redefining the definition of darkness. Specifically, they sought to simulate ambient lighting by using geographical information system (GIS) software to determine streetlight and traffic stop locations, then restricted their testing to locations that had limited artificial lighting. They found that the restricted test (which did not include stops that occurred in well-lit areas) concluded that Black drivers were stopped disproportionately, whereas the full test did not. However, Horrace and Rohlin (2016) note that they cannot explain why this difference exists. It may be due to “differential police behaviors in poorly lit areas, but it could be due to differential driving behaviors or some other unobservable features of poorly lit areas” (p. 231). Kalinowski, Ross, and Ross (2017) also proposed a variation on VOD, controlling for the severity of the infraction that resulted in a stop by limiting the sample to speeding offenses.

Measuring Disproportionate Impact in Other Enforcement Activities

Searches

Although traffic stops have received most of the research attention when it comes to assessing disproportionate impact among people of color it is not the only enforcement activity of interest. Black community members are also at a greater risk of being searched (Rojek, Rosenfeld, & Decker, 2004; Tillyer, Klahm, & Engel, 2012). Research that has examined disproportionality in searches notes that the percentage searched of specific racial and ethnic groups cannot be used alone to identify the causes of disparities because other factors may account for disproportionate search behavior (Higgins, Vito, & Walsh, 2008; Higgins, Vito, Grossi, & Vito, 2012). Examining the “hit rate” of searches (percentage of searches in which officers find something) is a better way to determine if a disparity exists in the productivity of searches (Fridell, 2005a, p. 10; Knowles, Persico, & Todd, 2001). However, research on disparity in searches has been limited (see Baumgartner et al., 2018, for an exception), particularly on the use of benchmarks to assess disproportionate treatment. The same is true for other law enforcement outcomes.

Stop and Frisk

Disproportionality in stop and frisk has been examined at length in New York City as part of its practice of zero tolerance policing (Smith et al., 2017), and researchers have found that the practice disproportionately affects people of color (Fagan, 2010, 2012; Fagan, Geller, Davies, & West, 2010; Gelman, Fagan, & Kiss, 2007; Smith et al., 2017). Smith et al. (2017), in their review of research examining disparate impact in police enforcement actions, note that different benchmarks for stop and frisk lead to different results (Ridgeway, 2007). When precinct-level census comparisons are used (Fagan, 2010, 2012), Black community members are overrepresented (Ridgeway, 2007). However, census-based benchmarks do not sufficiently account for rates of criminal participation (Fridel, 2004; Smith et al., 2017). When arrestees were used as a benchmark, Black community members were underrepresented in total stops but overrepresented in certain types (e.g., drug-related) of crimes (Ridgeway, 2007). However, arrestees might not best represent the population at risk. In New York City, most individuals stopped by the New York Police Department were never arrested or cited, so arrestees were not a good benchmark for stop and frisk (Gelman et al., 2007). Finally, when reports of criminal suspects (as reported by victims) were used, Blacks were significantly underrepresented (Ridgeway, 2007).

Arrest and Use of Force

A substantial amount of research has included race as a control variable in predicting arrest and use of force. A meta-analysis by Kochel et al. (2011) concluded that people of color were more likely to be arrested than Whites. The results of research on use of force, on the other hand, have been mixed, with generally null findings in regard to race and use of force (Smith et al., 2017). A recent review did not reveal any benchmarking analyses of arrest or use of force (Smith et al., 2017).

Disproportionality Index

Expanding on previous inquiry, the current research utilizes the DI to assess multiple outcomes. As mentioned, the DI has not been used in prior research on disproportionate contact with police. Similar indices have been used to examine racial disparities in child maltreatment victimization (Fluke, Yuan, Hedderson, & Curtis, 2003), disproportionality in foster care representation (Summers, 2015), and disproportionate representation in special education programs (Artiles, Rueda, Salazar, & Higareda, 2005). Different benchmarks and geographical areas were utilized to determine whether there was disparate treatment in vehicle stops, field contacts, arrests, and use of force.

Methods**Research Site**

Research on disproportionate policing practices was conducted at a midsize department in the southeastern United States (hereafter described with the pseudonym Darden Police Department, or DPD). As of 2017, according to the U.S. Census Bureau, the population of Darden was just under

100,000 people. The city is predominantly White, and 39% of the population is Black. The median household income was just over \$35,000.

The DPD provides a definition of biased policing in its policy and procedures manual, noting the difference between criminal profiling and biased policing. Specifically, the former is recognized as one of the many tools officers may use, supported by actionable intelligence, when carrying out their duties. For example, when a suspect is described as a “Black male in his mid-twenties wearing a red shirt,” an officer may stop individuals in the vicinity of the crime who match that description. Officers at DPD, however, are prohibited from using the latter in enforcement activities. Biased policing is described as selecting an individual for enforcement action on the basis of a trait that is common to a specific group (e.g., race, ethnic background, gender, sexual orientation) without a specific cause (e.g., actionable intelligence) to support consideration of that trait.

Officers at DPD receive annual training on issues related to biased policing, and complaints of such are investigated by the department’s Office of Internal Affairs. The department has a generally positive relationship with the community, and assessments of community satisfaction reflect that relationship (Blinded, 2017). There was no specific crisis or event that prompted the current research. The chief, like those in many other jurisdictions, undertook an assessment of disproportionate treatment to proactively address any issues that might arise. Although disparate enforcement practice can be considered for many traits, the focus for the current inquiry was an assessment of disparate treatment among the Black² community in the jurisdiction.

The chief of DPD sought to examine several enforcement actions, including traffic stops, field contacts, arrests, searches of persons and cars, and use of force. The benchmarks for each of these analyses were developed in collaboration with the executive leadership at the DPD, and according to best practice. As with other work of this nature, benchmarks serve as a low-discretion/no-discretion comparison to the police activity that is being evaluated for disproportionality. Since best practices regarding analyses of disparities indicate that smaller geographical areas of analysis provide more robust findings, the analyses for DPD were based on areas of strategic high enforcement – namely, hot spots and Data-Driven Approaches to Crime and Traffic Safety (DDACTS) areas.³

Benchmarks

The benchmark for traffic stops was crashes. Law enforcement is not involved in the events leading to crashes, so it is an occurrence that involves no discretion by officers. Crash data have been used in several prior studies assessing disproportionality in traffic stops (Alpert, Smith, & Dunham, 2004; Mosher & Pickerill, 2011; Smith et al., 2004). Neither checkpoint nor investigative stops were included in the analysis because these stops provide little opportunity for discretion. The

² The Hispanic population in the city is quite small, so the analysis does not include an assessment of disproportionate treatment based on ethnicity.

³ Neighborhood groups, which were identified in collaboration with DPD, were also used to conduct analyses. Neighborhood analyses, however, were based on small sample sizes, which allowed calculation of the DI in only some of the neighborhoods. The results indicated variability throughout the city, with Black community members overrepresented in some neighborhoods but underrepresented in others. Detailed results are available upon request.

benchmark for field contacts was any known demographic of a suspect of a crime (identified from crime reports) committed in the area (Ridgeway, 2007). Because suspect characteristics (when they are known) are reported by victims, they are also a no-discretion indicator.

Most of the research examining racial disparities in arrests has relied on multivariate analyses (e.g., Brown, 2005; Klinger, 1996; Smith, 1984; Worden & Pollitz, 1984), and these methodologies often require additional information gathered through observation and often not readily available to police department crime analysts. As such, in an effort to choose a no-discretion or low-discretion benchmark that would be readily accessible to police departments, the benchmark for discretionary arrests was arrests based on warrants (issued by a court, or privately) and arrests pursuant to investigation. These arrests are directed and represent low-discretion decisions by officers.

Finally, two benchmarks were used to analyze use of force. The first was assaults on police, and the second was known characteristics of a suspect for Part 1 violent crime (aggravated assaults and robberies) or weapons charges. Both benchmarks represent violent behavior on the part of a suspect but using assaults on police as a benchmark can be problematic. For example, Legewie (2016) found that the shooting of police officers by Black suspects was directly associated with an increase in the use of force by police officers during pedestrian stops of Black individuals in New York City. Thus, an additional no-discretion benchmark was utilized to counter any potential concerns about using assaults on police as a benchmark for use of force.

Plan of the Analysis

The primary analytical strategy used to assess the behavior in question against the benchmark was the DI, which is calculated by dividing the percentage observed of a specific racial category for a particular enforcement activity by the percentage expected. Consider a hypothetical example in which 35% of the drivers stopped for speeding are Hispanic (the percentage observed) and 29% of speeding drivers in the area are Hispanic (the percentage expected). To calculate the DI, 35% is divided by 29% to get 1.21 (Dolan Consulting Group, 2016). A DI of 1.0 indicates no disproportionate outcome (i.e., the racial category under examination is no more or less likely to receive enforcement). A DI above 1.0 indicates that enforcement is more likely than expected, whereas a DI below 1.0 indicates that enforcement is less likely than expected. Thus, in the preceding hypothetical example, Hispanic drivers are more likely than would be expected to be stopped for speeding. Very little empirical literature about the DI is available, so determining a specific threshold for concern is difficult. However, a report that focused on providing technical assistance regarding the collection and analysis of data for the assessment of disparate impact noted that some studies used a criterion of 5% (McMahon, Garner, Davis, & Kraus, 2002). Thus, for these purposes, a DI of 1.05 or less indicates that a particular treatment can reasonably be attributed to chance, whereas a DI of 1.10 indicates a moderate amount of disproportionate treatment.

TABLE 1: Comparing Traffic Stops to Crashes in DDACTS Areas

DDACTS	Traffic Stops		Crashes		DI
	N	% Black	N	% Black	
Jan-June 2015 Primary	915	54.32	634	32.18	1.69
Jan-Dec 2015 Secondary	3,016	52.75	1,298	37.44	1.41
Jan-Dec 2016 Secondary	2,015	65.26	1,247	45.79	1.43
Jan-June 2017 Secondary	2,959	48.23	3,485	31.88	1.51
July 2015=June 2017 Primary	2,716	53.68	1,456	35.51	1.51
OVERALL	11,621	54.11	8,120	35.58	1.52

Results

Traffic Stops

To assess disproportionate treatment in traffic stops, crash data were chosen as the benchmark. The race of any driver involved in a traffic crash within the neighborhood group or DDACTS area was compared to the race of the driver in traffic stops. Table 1 presents the results of the analysis comparing traffic stops with crash data in the DDACTS areas, which are high-enforcement zones for traffic. The smallest DI is for the January-December 2015 secondary DDACTS area, at 1.41 (indicating that the likelihood of a Black driver being stopped is 41% greater than would be expected), whereas the largest DI is for the January-June 2015 primary DDACTS area, at 1.69. The average DI across all the DDACTS areas for the full study period is 1.52. Thus, within the DDACTS areas for this time period, a Black driver was 52% more likely to be stopped than would be expected.

Field Contacts

Known characteristics of Part 1 crime suspects were used as the benchmark to determine disproportionality in hot spots. Table 2 indicates that many of the DIs for hot spots fall outside the margin of error (0.95-1.05). The largest DIs, both 1.11, are for the January-June 2015 secondary hot spot and the July-December 2015 secondary hot spot. The primary hot spot for July-December 2015, the secondary hot spot for July-December 2016, the secondary hot spot for January-June 2017, and the 2016 and 2017 combined hot spots all have DIs between 0.95 and 1.05. Taken together, all the hot spot zones in the full time period indicate that a Black community member was 6% more likely to be stopped than would be expected.

Arrests

Table 3 compares discretionary arrests to those that are mandatory (i.e., warrant-based). Only two of the DIs fall outside the specified margin of error. The January-June primary hot spot in 2015 shows 27% lower odds of a Black community member being arrested than would be expected,

whereas the combined hot spots in 2015 show 10% lower odds of a Black community member being arrested than would be expected.

TABLE 2: Comparing Field Contacts to Known Part I Crime Suspects in Hot Spots

	Field Contacts		Part I Crime Suspects		DI
	N	% Black	N	% Black	
2015					
Jan-June Primary	124	88.70	106	83.02	1.07
Jan-June Secondary	128	95.00	112	85.70	1.11
July-Dec Primary	188	90.10	165	86.70	1.04
July-Dec Secondary	124	87.00	107	78.50	1.11
Combined	564	90.10	490	83.10	1.08
2016					
Jan-June Primary	193	90.00	128	84.40	1.07
Jan-June Secondary	122	90.60	155	83.20	1.09
July-Dec Primary	130	90.70	107	83.20	1.09
July-Dec Secondary	190	87.20	127	89.80	.97
Combined	635	89.40	517	84.80	1.05
2017					
Jan-June Primary	171	95.10	138	89.10	1.07
Jan-June Secondary	165	85.40	167	84.40	1.01
Combined	336	90.20	305	86.60	1.04
OVERALL	1,535	89.90	1,312	84.60	1.06

Use of Force

To examine disproportionality in use of force, two benchmarks were used. Few incidents of use of force occurred within the DPD during the 2.5-year time frame, so that it was not possible to examine use of force within smaller geographical units. Table 4 compares the percentage of Black community members involved in use of force incidents versus the percentage of those charged with assaulting a police officer. The DI is 1.01, which is well within the margin of error. Table 5 compares the percentage of Blacks involved in a use of force incident with the percentage of Black suspects in a violent crime or a weapons charge. The DI is 0.88, which indicates 12% lower odds of a Black community member being involved in a use of force incident than would be expected.

Supplemental Analyses

In addition to the DI, specifically for traffic stops, a second analysis was performed in which the VOD method was used. The “inter-twilight” period for the city during the study was the period between 5:24 p.m. (the earliest end of civil twilight) and 8:59 p.m. (the latest end of civil twilight). The sample was further restricted by the removal of stops conducted as part of an investigation,

as these do not involve discretion. Consistent with prior research (Grogger & Ridgeway, 2006; Worden et al., 2012), the VOD method utilizes a logistic regression model to detect any difference between stops in daylight and stops in darkness. Time of day in 30-minute increments is included as a control.

TABLE 3: Comparing Discretionary to Mandatory Criminal Arrests in Hot Spots

	Discretionary		Mandatory		DI
	N	% Black	N	% Black	
2015					
Jan-June Primary	353	62.30	153	85.50	.73
Jan-June Secondary	170	92.30	206	89.20	1.03
July-Dec Primary	210	90.00	288	89.20	1.01
July-Dec Secondary	205	87.70	242	87.40	1.00
Combined	938	79.60	889	88.10	.90
2016					
Jan-June Primary	202	88.10	291	87.80	1.00
Jan-June Secondary	223	87.40	253	90.90	.96
July-Dec Primary	176	85.20	201	84.50	1.01
July-Dec Secondary	217	85.50	228	87.60	.98
Combined	818	86.60	973	87.90	.99
2017					
Jan-June Primary	58	91.40	102	90.10	1.01
Jan-June Secondary	196	81.60	230	84.10	.97
Combined	254	83.90	332	86.00	.98
OVERALL	2,010	83.00	2,194	87.70	.95

Tables 6 and 7 present the results of the VOD analysis. Table 6 displays the percentages of Black drivers stopped during daylight and during darkness for the “inter-twilight period” from January 2015 to June 2017. When these percentages are examined, it appears that the percentage of stops involving Black drivers was lower during darkness than during daylight in 2016, a possible sign of disproportionality, whereas it was higher during darkness in 2015 and only slightly higher in 2017. Table 7 tests these relationships using a logistic regression model predicting the odds of a driver being Black versus White in daylight over darkness controlling for time of day. The results indicate that the odds of a Black driver being stopped during daylight rather than darkness were significantly higher in 2016, but that the difference in the odds did not reach statistical significance in 2015 or 2017. Specifically, the odds of a Black driver being stopped during daylight rather than during darkness were 45.5% higher in 2016.

TABLE 4: Comparing Use of Force to Assault on a Police Officer (2015-2017)

	Use of Force		Assault on Police Officer		DI
	N	% Black	N	% Black	
Overall	83	73.02	86	72.09	1.10

TABLE 5: Comparing Use of Force to Known Violent Part I Crime and Weapons Charges Suspects (2015-2017)

	Use of Force		Violent Crime/Weapons Charges		DI
	N	% Black	N	% Black	
Overall	83	75.90	1,396	85.78	.88

TABLE 6: Racial Makeup (% Black) of Traffic Stops During Daylight and Darkness for the Inter-twilight Period

Year	% Daylight	% Darkness
2015	50.99	54.64
2016	64.39	59.11
2015	58.76	59.13

TABLE 7: Odds of Blacks Versus Others Being Stopped During Daylight Versus Darkness

Year	Odds Ratio
2015	.93
2016	1.45*
2015	1.40

Note: Controlled for clock time using 30-minute intervals.

*p<.05

Discussion

The results of the analyses do not indicate widespread disproportionate treatment of Blacks by the DPD. However, evidence of disproportionality is found in some parts of the city for some time periods and for some outcomes. Whereas the DI analysis indicates some disproportionality in traffic stops in all the high-enforcement DDACTS areas, the VOD analysis paints a slightly more nuanced picture. This analysis suggests that although disproportionate treatment of Blacks may have occurred in 2016, the disparity was no longer present in 2017. It is important to note that the department conducted additional diversity training in 2017, which may account for this finding; however, an examination of the effect of the training is beyond the scope of the current analysis.

The results for field contacts show little disproportionality within the hot spots, with only a few hot spot areas showing an increase of more than 5% in the likelihood of being stopped. Similarly, the results of the arrest analysis show little disproportionality in the hot spots. The use of force analysis, restricted to the city, likewise shows little disproportionality for either

benchmark. Because the relative infrequency of use of force did not allow for additional levels of analysis, we were unable to consider disparate impact in high-enforcement areas. Future analyses should consider the benchmarks, geographical foci, and data time frames that are appropriate for the jurisdiction.

Given that this is a case study, generalizability to other police departments is limited. However, because the DPD is a midsize police department, it is like many other police departments in the United States. Several additional limitations influenced our analyses. First, it is important to note that race/ethnicity is not recorded on driver licenses in the study jurisdiction. As such, officers must identify race/ethnicity according to their perception. It is impossible to know to what extent (if any) the data were affected by mistaken identification of the race/ethnicity of a community member by an officer. Second, attempts to conduct analyses within additional geographical areas were hampered by a lack of truly defined neighborhoods. Analyses of disproportionate impact should ultimately examine geographical areas that are not focused just on enforcement (i.e., DDACTS and hot spots), as was the case in our analyses. Relatedly, such analyses should also not be conducted on a city-wide basis. This was necessary given the low overall number of use of force incidents for the DPD (combined with an inability to capture a full 10 years of data, as is suggested by the Dolan Consulting Group [2016]), but it is not ideal.

Third, we were hampered by geocoding issues for many of the analyses. It is not always possible to geocode an incomplete or incorrect address, which meant that most of these cases had to be dropped from the analysis. Fourth, the DI is unstable with small numbers, so we could not complete some of our planned analyses. For example, we had planned to utilize DWI-related crashes as a benchmark for traffic arrests. However, after we had merged the necessary data from two systems and coded for the defined geographical areas, the number of DWI-related crashes during the study period was inadequate to allow us to estimate that DI reliably. Finally, the selection of the benchmark is important when the DI is used. However, the selection is dictated by the availability of data from law enforcement systems to allow measurement of the benchmark, along with the reality of the number of incidents within defined geographical areas. For some analyses, we planned and discarded numerous benchmarks because of inefficiencies in the data. Undoubtedly, some of our planned analyses would have been stronger had we been able to use our planned primary benchmark.

The following limitation did not pertain to the findings presented, but we note the issue here to provide additional context for other researchers or practitioners who might be planning similar analyses. Our inability to conduct other planned analyses stemmed from newer data collection practices at the DPD that had not been in place long enough to generate sufficient data for analysis. For example, a traffic form policy that mandated the collection of data to inform analysis of searches of cars was established in the middle of the study period, which made it impossible to conduct the analysis for the current inquiry. The same was true of discretionary search information. This information is provided in the department's Field Contact form, which was modified in the middle of the study period to include whether an officer safety search (weapons pat down) or consensual search was conducted. As with traffic searches, the limited amount of available data for the study time frame made it impossible to complete an analysis of disparate impact in person searches. However, even if this limitation did not exist, there are currently no

mandatory fields for the data entry into the Field Contact form – the amount of information provided is at the officer's discretion. This is problematic because it means that information on searches may not be collected in a systematic manner. Future inquiry should carefully consider data limitations such as these when research seeking to examine disparate impact is undertaken.

Conclusions

The findings highlight the importance of combining the appropriate benchmark with the appropriate level of analysis, and the need for more scholarly inquiry into disproportionate treatment in law enforcement, especially for outcomes that have not previously received much research attention. As stated previously, although both the DI and the VOD measure disproportionate impact, they do not measure bias. Neither can determine if disproportionality is due to racial profiling or criminal profiling on the part of officers, and it is important to determine if a cause and effect relationship exists between racial disparity and racial bias (Smith et al., 2017). Additional research that more deeply investigates individual officer decision making is necessary to determine *why* a disproportionate outcome exists. That said, it is quite difficult to determine empirically if an officer has an implicit or explicit bias against people of color that results in discriminatory behavior. Further, departments may encounter several challenges, such as those experienced at the study site, that make it difficult to present nuanced results to the public with confidence. However, this is true of any analysis that seeks to investigate racial and ethnic disparity in the criminal justice system.

Although it is not easy to conclude that bias exists where a disproportionate outcome exists, that does not mean that departments cannot (and should not) use disproportionate results to take strides in improving police-community relations. What such analyses do provide is an ability to focus on potential areas of concern, and departments that are transparent about disproportionate enforcement activities and action items that seek to address such disproportionality, whatever the cause, signal to their respective communities that responding positively to community concerns about disparate impact is important. Thus, it is important for departments and, if appropriate, their research partners to continue to investigate disproportionality in policing, and for the field to continue to refine and test measures of both bias and disproportionate impact. Future research should, for example, endeavor to determine the link between bias (implicit or explicit) and disproportionate treatment by examining the influence of anti-bias training on disparity in law enforcement outcomes (Smith et al., 2017). Further, body-worn cameras are expected to make low-visibility decisions more observable. The introduction of body cameras in an agency could be used to assess any changes in patterns of enforcement under the assumption that cameras might reduce bias if it is occurring (Lum, Koper, Merola, Scherer, & Reioux, 2015).

Ultimately, we have demonstrated that it is possible to use the DI to assess disproportionality in several outcomes, including those in which researchers have not previously used benchmarking analyses. Several alternate benchmarks were used to conduct our analyses – benchmarks that were developed in collaboration with the DPD with a focus on utilizing data that would be readily accessible to police departments. We noted several limitations in our analyses, however, which highlights the need for agency-specific foci when disproportionate impact is measured, along with a realistic assessment of the availability of necessary data.

Declaration of Conflicting Interests

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support with respect to the research, authorship, and/or publication of this article.

References

- Alpert, G. P., Smith, M. R., & Dunham, R. G. (2004). Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy*, 6, 43–69.
- Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children*, 71, 283–300.
- Baumgartner, F. R., Epp, D. A., & Shoub, K. (2018). *Suspect citizens: What 20 million traffic stops tell us about policing and race*. New York, NY: Cambridge University Press.
- Black, D. (1971). The social organization of arrest. *Stanford Law Review*, 23, 1087–1111.
- Black, D., & Reiss, A. J. (1970). Police control of juveniles. *American Sociological Review*, 35, 63–77.
- Blinded. (2017). *Citizen fear of crime and satisfaction with the police: Final report*. Report to the Darden Police Department.
- Brown, R. A. (2005). Black, white, and unequal: Examining situational determinants of arrest decisions from police-suspect encounter. *Criminal Justice Studies*, 18, 51–68.
- Dolan Consulting Group. (2016). *Biased based policing reports: Best practices*.
- Engel, R. S., & Calnon, J. M. (2004). Comparing benchmark methodologies for police–citizen contacts: Traffic stop data collection for the Pennsylvania State Police. *Police Quarterly*, 7, 97–125.
- Engel, R. S., Calnon, J. M., & Bernard, T. J. (2002). Theory and racial profiling: Shortcomings and future directions in research. *Justice Quarterly*, 19, 249–273.
- Fagan, J. (2010). *Floyd v. New York*, No. 08 Civ. 1034, US District Court S.D.N.Y. Report of Jeffrey Fagan, PhD Expert Rep.
- Fagan, J. (2012). *Floyd v. New York*, 959 F. Supp. 2d 540 (SDNY 2013). Second supplemental report of Jeffrey Fagan, PhD Expert Rep.
- Fagan, J., Geller, A., Davies, G., & West, V. (2010). Street stops and broken windows revisited: The demography and logic of proactive policing in a safe and changing city. In S. K. Rice & M. D. White (Eds.), *Race, ethnicity, and policing: New and essential readings* (pp. 309–348). New York, NY: New York University Press.
- Farrell, A., Rumminger, J., & McDevitt, J. (2004). *New challenges in confronting racial profiling in the 21st century: Lessons from research and practice*. Boston, MA: Northeastern University.
- Fluke, J. D., Yuan, Y. T., Hedderson, J., & Curtis, P. A. (2003). Disproportionate representation of race and ethnicity in child maltreatment: Investigation and victimization. *Children and Youth Services Review*, 25, 359–373.

- Fridell, L. (2004). *By the numbers: A guide for analyzing race data from vehicle stops*. Washington, DC: Police Executive Research Forum.
- Fridell, L. A. (2005a). *Racially biased policing: Guidance for analyzing race data from vehicle stops*. Washington, DC: U.S. Department of Justice, Office of Community Oriented Policing Services.
- Fridell, L. A. (2005b). *Understanding race data from vehicle stops: A stakeholder's guide*. Washington, DC: U.S. Department of Justice Office of Community Oriented Policing Services.
- Fridell, L., Lunney, R., Diamond, D., & Kubu, B. (2001). *Racially biased policing: A principled response*. Washington, DC: U.S. Department of Justice Office of Community Oriented Policing Services.
- Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the NYPD's stop and frisk policy in the context of claims of racial bias. *Journal of the American Statistical Association*, 102, 813–822.
- Grogger, J., & 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.
- Higgins, G. E., Vito, G. F., Grossi, E. L., & Vito, A. G. (2012). Searches and traffic stops: Racial profiling and capriciousness. *Journal of Ethnicity in Criminal Justice*, 10(3), 163–179.
- Higgins, G. E., Vito, G. F., & Walsh, W. F. (2008). Searches: An understudied area of racial profiling. *Journal of Ethnicity in Criminal Justice*, 6(1), 23–39.
- Horrace, W. C., & Rohlin, S. M. (2016). How dark is dark? Bright lights, big city, racial profiling. *Review of Economics and Statistics*, 98(2), 226–232.
- International Association of Chiefs of Police. (2006). *Addressing racial profiling: Creating a comprehensive commitment to bias-free policing*. Washington, DC: U.S. Department of Justice Office of Community Oriented Policing Services.
- Kalinowski, J., Ross, S., & Ross, M. (2017). *Endogenous driving behavior in veil of darkness tests for racial profiling* (No. 2017-03). Storrs, CT: University of Connecticut, Department of Economics.
- Klinger, D. A. (1996). More on demeanor and arrest in Dade County. *Criminology*, 34, 61–82.
- Knowles, J., Persico, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, 1(109), 203–229.
- Kochel, T. R., Wilson, D. B., & Mastrofski, S. D. (2011). Effect of suspect race on officers' arrest decisions. *Criminology*, 49(2), 473–512.
- LaFave, W. (1965). *Arrest: The decision to take a suspect into custody*. Boston, MA: Little, Brown.
- Lamberth, J. (2003). *Racial profiling data analysis: Final report for the San Antonio Police Department* (unpublished report). Chadds Ford, PA: Lamberth Consulting.
- Lange, J. D., Blackman, K. O., & Johnson, M. B. (2001). *Speed violations of the New Jersey Turnpike: Final report*. Calverton, MD: Public Services Research Institute.
- Legewie, J. (2016). Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination. *American Journal of Sociology*, 122(2), 379–424.
- Lum, C., Koper, C., Merola, L., Scherer, A., & Reieux, A. (2015). *Existing and ongoing body worn camera research: Knowledge gaps and opportunities. A Research Agenda for the Laura and John Arnold Foundation*. Fairfax, VA: Center for Evidence-Based Crime Policy, George Mason University.

- Lum, C., & Wu, X. (April 2017). *Basic analysis of traffic citation data for the Alexandria police department (2011-2015)*. Fairfax, VA: Center for Evidence-Based Crime Policy, George Mason University.
- McMahon, J., Garner, J., Davis, R., & Kraus, A. (2002). *How to correctly collect and analyze racial profiling data: Your reputation depends on it!*. U.S. Department of Justice, Office of Community Oriented Policing Services.
- McDevitt, J., & Iwama, J. (2016). *Vermont state police: An examination of traffic stop data*. Report prepared for Institute on Race and Justice. Boston, MA: Northeastern University.
- Mosher, C., & Pickerill, J. M. (2011). Methodological issues in biased policing research with applications to the Washington State Patrol. *Seattle University Law Review*, 35, 769.
- Novak, K. J. (2004). Disparity and racial profiling in traffic enforcement. *Police Quarterly*, 7, 65–96.
- President's Task Force on 21st Century Policing. (2015). *Final report of the president's task force on 21st century policing*. Washington, DC: Office of Community Oriented Policing Services.
- Reiss, A. (1971). *The police and the public*. New Haven, CT: Yale University Press.
- Ridgeway, G. (2007). *Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices*. Santa Monica, CA: RAND.
- Ridgeway, G. (2009). *Cincinnati Police Department traffic stops: Applying RAND's framework to analyze racial disparities*. Santa Monica, CA: RAND.
- Ritter, J. A., & Bael, D. (2009). Detecting racial profiling in Minneapolis traffic stops: A new approach. *CURA Reporter*, 11–17.
- Rojek, J., Rosenfeld, R., & Decker, S. (2004). The influence of driver's race on traffic stops in Missouri. *Police Quarterly*, 7(1), 126–147.
- Ross, M., Fazzalari, J., Barone, K., & Kalinowski, J. (2015). *State of Connecticut: Traffic stop data analysis and findings, 2013–2014*. Report prepared for Institute for Municipal and Regional Policy. New Britain, CT: Central Connecticut State University.
- Smith, M. R., & Petrocelli, M. (2001). "Racial profiling?: a multivariate analysis of police traffic stop data. *Police Quarterly*, 4, 4–27.
- Smith, M. R., Rojek, J. J., Petrocelli, M., & Withrow, B. (2017). Measuring disparities in police activities: A state of the art review. *Policing: An International Journal of Police Strategies & Management*, 40(2), 166–183.
- Smith, W. R., Tomaskovic-Devey, D., Zingraff, M. T., Mason, H. M., Warren, P. Y., & Wright, C. P. (2004). *The North Carolina highway traffic study*. Final report to the National Institute of Justice, U.S. Department of Justice.
- Smith, D. A., Visser, C. A., & Davidson, L. A. (1984). Equity and discretionary justice: The influence of race on police arrest decisions. *Journal of Criminal Law and Criminology*, 75, 234–249.
- Summers, A. (2015). *Disproportionality rates for children of color in foster care (fiscal year 2013)*. National Council of Juvenile and Family Court Judges.
- Taniguchi, T. A., Hendrix, J. A., Levin-Rector, A., Aagaard, B. P., Strom, K. J., & Zimmer, S. A. (2017). Extending the veil of darkness approach: An examination of racial disproportionality in traffic stops in Durham, NC. *Police Quarterly*, 20(4), 420–448.

- Tillyer, R., Engel, R. S., & Cherkauskas, J. C. (2010). Best practices in vehicle stop data collection and analysis. *Policing: An International Journal of Police Strategies & Management*, 33, 69–92.
- Tillyer, R., Klahm, C. F., & Engel, R. S. (2012). The discretion to search: A multilevel examination of driver demographics and officer characteristics. *Journal of Contemporary Criminal Justice*, 28, 184–205.
- Tomaskovic-Devey, D., Mason, M., & Zingraff, M. (2004). Looking for driving while black phenomena: Conceptualizing racial bias processes and their associated distributions. *Police Quarterly*, 7, 3–29.
- Tyler, T. R., & Wakslak, C. J. (2004). Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority. *Criminology*, 42, 253–281.
- Walker, S. (2001). Searching for the denominator: Problems with police traffic stop data and an early warning system solution. *Justice Research and Policy*, 3, 63–95.
- Walker, S. (2003). *Internal benchmarking for traffic stop data: An early intervention systems approach*. Omaha, NE: Police Professionalism Initiative, University of Nebraska at Omaha.
- Withrow, B. L. (2003). *Sedgwick County (Kansas) Sheriff's Department: Racial profiling study*. Wichita, KS: Wichita State University, Midwest Criminal Justice Institute (as cited in Baumgartner, Epp, & Shoub, 2018).
- Withrow, B. L. (2006). *Racial profiling: From rhetoric to reason*. Upper Saddle River, NJ: Pearson/Prentice Hall.
- Withrow, B. L., & Williams, H. (2015). Proposing a benchmark based on vehicle collision data in racial profiling research. *Criminal Justice Review*, 40, 449–469.
- Worden, R. E., McLean, S. J., & Wheeler, A. P. (2012). Testing for racial profiling with the veil of darkness method. *Police Quarterly*, 15(1), 92–111.
- Worden, R. E., & Pollitz, A. A. (1984). Police arrests in domestic disturbances: A further look. *Law and Society Review*, 18, 105–119.
- Worrall, J. L. (1999). Public perceptions of police efficacy and image: The 'fuzziness' of support for the police. *American Journal of Criminal Justice*, 24, 47–66.
- Zingraff, M., Smith, W., & Tomaskovic-Devey, D. (2000). North Carolina highway traffic and patrol study: Driving while black. *Criminologist*, 25(3), 1–4.

Authors' Biographies

Heidi S. Bonner, PhD, is an associate professor in the Department of Criminal Justice at East Carolina University and a research fellow at the John F. Finn Institute for Public Safety. Her research focuses on individual and organizational criminal justice decision-making behavior and outcomes, with an emphasis on law enforcement operations.

Michele Stacey, PhD, is an assistant professor in the Department of Criminal Justice at East Carolina University. Her research focuses on the victimization and social control of minority and special population groups.