Article

Police Quarterly
2017, Vol. 20(4) 420–448

DOI: 10.1177/1098611117721665
journals.sagepub.com/home/pqx

Extending the Veil of Darkness Approach: An Examination of Racial Disproportionality in Traffic Stops in Durham, NC

Travis A. Taniguchi[1], Joshua A. Hendrix[1], Alison Levin-Rector[1], Brian P. Aagaard[1], Kevin J. Strom[1], and Stephanie A. Zimmer[1]

Abstract

Developed in 2006, the veil of darkness approach is one of the most widely accepted methods for assessing the impact of driver race on traffic stops. Building on the original methodology, we innovate in three important ways to enhance the veil of darkness approach: (a) invoke generalized linear mixed models to account for the lack of independence among observations in traffic stop data sets, (b) decompose the relationship between daylight and driver race to consider the role of driver sex, and (c) assess variability in racial disproportionality across law enforcement units. Nearly 20,000 traffic stops are analyzed for the Durham (NC) Police Department. Results indicate that more than 10% of the variability in the rate of Black drivers stopped is accounted for by officer-level factors, racial disproportionality was only for male drivers, and evidence of disproportionality was found among some units, but no evidence was found among others.

Keywords

veil of darkness, traffic stops, racial disproportionality

[1]RTI International, Research Triangle Park, NC, USA

Corresponding Author:
Travis A. Taniguchi, RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709, USA.
Email: taniguchi@rti.org

Racial profiling by law enforcement occurs when authorities target individuals because of their race or ethnicity rather than their behavior. It is a social issue that is counter to the constitutional guarantee of equal treatment under the law, and when it is perceived to exist, police–community relations may suffer. Traffic stops are not only the most common reason for contact with the police (Bureau of Justice Statistics, 2016), for many people, traffic stops are likely to be the only type of interaction they ever have with law enforcement. Thus, the nature of police contact, its circumstances, and outcomes can have dramatic and possibly unrivaled influence on how people view the police (see Smith, Graham, & Adams, 1991; Worrall, 1999). It is well established that perceptions of unfair or discriminatory practices by the police have serious implications for community cooperation and civil disobedience (Cox & Fitzgerald, 1996; Tyler & Fagan, 2008), support for the police (Tyler & Wakslak, 2004), race relations, crime reporting, and public safety (see Brown & Benedict, 2002 for a review). Therefore, racial profiling by police is a critical social issue and an important research area in the social sciences.

Assessing how much racial profiling by police occurs in the United States is a complex issue, primarily because it is one of the most difficult social phenomena to study scientifically. It is an extremely challenging task to prove definitively that the nature or outcome of a traffic stop would have been different if the driver was of a different race, or that the officer has an explicit or implicit bias toward people of color that inspires discriminatory police practices. Although a Gallup survey in 2003 revealed that 59% of Americans consider racial profiling to be widespread among police stops (Ludwig, 2003; see also Weitzer & Tuch, 2002), this study cannot speak to whether perceptions of racially biased police practices reflect reality.

While an extensive body of literature exists examining the factors that influence police conduct once a community encounter occurs (Riksheim & Chermak, 1993), there is less knowledge of the factors that influence a police decision to make a traffic stop in the first place. Traditionally, the impact of race on traffic stops has been assessed by assuming that driving patterns can be approximated using census population estimates (Baumgartner & Epp, 2012a, 2012b) or traffic collision data (McDevitt & Iwama, 2016) or by conducting traffic surveys to quantify the race distribution of motorists (Smith et al., 2004). All of these approaches have serious limitations. Census populations are a poor proxy for populations at risk for traffic stops, and traffic surveys are not only expensive, they tend to have limited generalizability because of small sample sizes and geographic coverage. Use of traffic collision data is a better denominator in some respects because it more accurately reflects the driving population (Alpert, Smith, & Dunham, 2004). But this benchmark construction approach assumes that accidents, and reporting of accidents to the police, do not vary systematically. There have in fact been numerous attempts to establish an appropriate, unbiased benchmark with which to compare police traffic stop data, but no approach has proven to completely account for all characteristics of driver populations (Lange, Johnson, & Voas, 2005).

Ross, Fazzalaro, Barone, and Kalinowski (2015) highlight three specific limitations with these traditional approaches to determining a traffic denominator. First, the spatial patterns of work and recreation, pass-through commuters, and recreational traffic mean that more people are in a given geography than expected based on census residential population. Most problematically, it is challenging, if not impossible to reliably model changes in population. Second, the exposure or risk of a traffic stop may be conditional upon demographic characteristics. An effective population benchmark would need to also incorporate the differential risk of being stopped. Finally, population benchmarks lack sensitivity to the differences in risk that are associated with time of day and time of week. This is likely to amplify the effect of the other limitations of population-based models and may exacerbate errors in model specification.

In 2006, Grogger and Ridgeway attempted to address these issues through a method they termed the *veil of darkness* (VOD). This approach hypothesizes that when it is dark outside, police officers are less capable of detecting the race of a motorist before a stop occurs than when it is daytime. Taking advantage of this assumption, the VOD method aims to determine whether the race of stopped motorists varies as a function of daylight, while controlling for the time of day. To date, the VOD method has been applied in five jurisdictions: Oakland, CA (Grogger & Ridgeway, 2006; Oakland Police Department, 2004), Cincinnati, OH (Ridgeway et al., 2009), Minneapolis, MN (Ritter & Bael, 2009), Syracuse, NY (Worden, McLean, & Wheeler, 2012), and Connecticut (Ross et al., 2015). No significant difference in traffic stops by race were found in Oakland, Cincinnati, or Syracuse; however, in Minneapolis and Connecticut, researchers found that the percentage of drivers stopped after nightfall who were Black was higher than the percentage stopped during daylight, suggesting the presence of racial profiling (Table 1).

Building directly on the original methodology, we enhance the VOD approach in three important ways: (a) adjust models to account for officer-level variation, (b) determine the role of the driver's sex, independent of race, on police stops, and (c) assess variability in racial disproportionality across different law enforcement unit assignments (e.g., patrol and traffic enforcement). Collectively, our approach improves the precision of VOD estimates and facilitates more nuanced understanding regarding driver characteristics that may be targeted by officers and the departmental units that are most likely to engage in disproportionate minority contact during traffic stops. Although we specifically apply our enhanced VOD method to data on traffic stops conducted by the Durham Police Department (DPD) from January 2010 through October 2015, the issues addressed are comparable across jurisdictions, and our hope is that our improved approach and conclusions are generalizable.

Table 1. Comparison of VOD Methodology.

Source	Jurisdiction	Date range	*N*	*N* intertwilight	Subgroup analysis	Disproportionality	Novelty of analysis
Grogger and Ridgeway, 2006	Oakland, CA	June 15–December 30, 2003	7,607	1,130	Neighborhoods	No	N/A
Ridgeway, 2009	Cincinnati, OH	2008	34,099[a]	4,817	Year	No[c]	Monthly controls
Worden et al., 2012	Syracuse, NY	2006–2009	50,456	17,172	Year Officer unit	No	Time as fixed effect Day of week control
Ritter and Bael, 2009	Minneapolis, MN	2002	53,559	N/A	Varied race groupings	Yes	Poststop analysis
Ross et al., 2015	Connecticut[b]	October 2013–September 2014	595,194	136,762	Cities Varied race groupings	Yes	Explore city-level effects in aggregate analysis
Current study	Durham, NC	January 2010–October 2015	151,701	19,801	Driver sex Officer unit	Yes	Random intercept to explore officer-level effects Temporal interaction term

Note. N/A= not applicable.
[a]Number of stops with race recorded. Only includes 2008 data.
[b]Multiple jurisdictions within the state of Connecticut. VOD method was modified to account for multiple agencies.
[c]Black drivers were significantly underrepresented in daylight ITP.

Methods

Overview of the VOD Method

The VOD approach is based on the logic that police officers are less capable of determining the race of a motorist after dark than during daylight hours. The theory holds that if officers cannot observe the race of the driver during hours of darkness, then the effect of race on the decision to stop should be minimized. Traffic studies have established the difficulty of determining driver race during hours of darkness. Lamberth (2003), for example, found that driver race was easily determined during daylight but required additional lighting during darkness. Greenwald (2001) found that driver race could only be determined for 6% of drivers during darkness.

The VOD methodology takes advantage of a natural experiment that is made possible by seasonal variation in the amount of daylight in a time period known as the "intertwilight period"; limiting the analysis to this time period minimizes the variation in travel patterns that are conditional on time of day. Figure 1 shows a graphical depiction of the intertwilight period. The VOD approach compares the racial distribution of motorists stopped during the intertwilight period when it

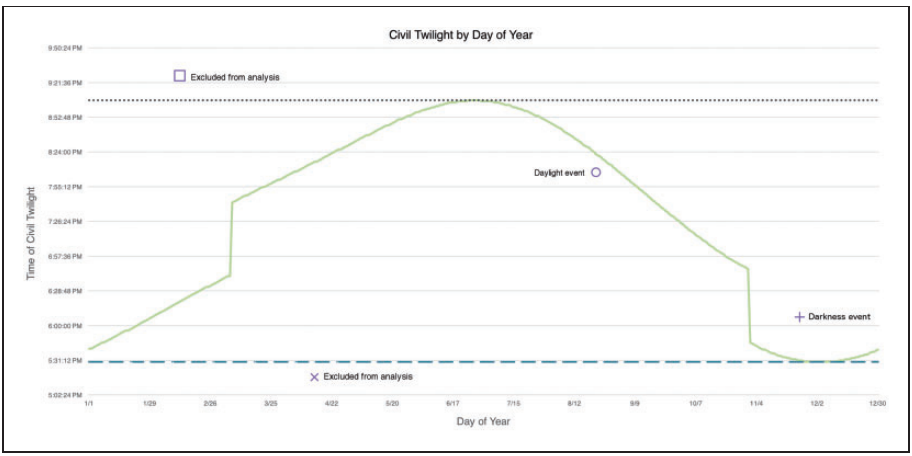

Figure 1. Civil twilight by day of year.

□ Events above the dotted line occurred *later* than the latest civil twilight—always in the dark. These events are excluded.

X Events below the dashed line occurred *before* the earliest civil twilight—always during daylight. These events are excluded.

O Events between the dotted and dashed lines occurred after the earliest civil twilight and before the latest civil twilight. Events under the curve occurred before civil twilight of that day and are considered daylight events.

+ Events outside of the curve occurred after civil twilight for that day and are considered darkness events.

is daylight with the racial distribution of motorists stopped after dark during the intertwilight period. If more minority motorists are stopped when it is daylight than when it is dark, controlling for other factors, then racial bias exists in the way that traffic stops are being carried out.

The VOD method estimates a logistic regression using data on traffic stops, which are recorded by law enforcement agencies in many jurisdictions. For example, the Bureau of Justice Statistics' state police traffic stop data collection procedures study, which was conducted from 1999 to 2004 showed that nearly 60% of state law enforcement agencies required traffic patrol officers to record information (e.g., race and ethnicity) about the motorist during traffic stops (Hickman, 2005). Although we are unaware of any systematic exploration of traffic stop data at a local level, there are at least a few states that mandate the collection of these data (e.g., North Carolina, Connecticut, and Missouri).

At a minimum, the data elements required to implement the method are the time and date of the traffic stop and the race or ethnicity of the driver. The original model proposed by Grogger and Ridgeway (2006) predicts the race of the driver using an indicator for daylight and a function of time of day, which accounts for the possibility that the mix of drivers on the road changes throughout the intertwilight period. If the daylight indicator is statistically nonsignificant, the results of the model suggest that daylight, and theoretically visibility, was not associated with the race of the driver. This simplicity of interpretation is a major strength of the VOD method, but in our view, the original method and its applications to date do not consistently or adequately control for additional, relevant information, which is necessary to gain an accurate understanding of the presence of racial profiling in traffic stops.

Limitations of Existing Studies

Three key limitations reduce the accuracy and comparability of the original VOD method and its applications to date. Prior studies do not regularly (a) account for officer-level variation in traffic stop characteristics or outcomes, (b) disentangle the role of sex and race in racial profiling in police stops, and (c) they typically treat police departments as monolithic entities rather than exploring racial disproportionality in traffic stops across different police units. The following sections describe each of these limitations and how we address them.

Refinements to Existing Method

Refinement 1: accounting for officer-level variation. Prior applications of the VOD method do not account for the inherent dependence in traffic stop data that are caused by officers making multiple traffic stops. In North Carolina and other jurisdictions, traffic stop records include an officer identification number that is linked to the officer's employment number for internal investigative purposes. These identification numbers typically preserve anonymity while being consistently linked to the same officer over time, allowing for officer-level analyses. Because officers cover different geographic areas that may vary in their racial distribution and because decisions to stop motorists are often discretionary, traffic stops conducted by the same officer are more likely to be similar to one another than are traffic stops conducted by different officers. If this relationship between observations is not accounted for, then the independence assumption of the analytic approach used in the VOD method is violated, and the analysis may produce biased standard errors and ultimately lead to inaccurate conclusions.

We resolve this limitation by including a random intercept in the model because we expect that the race of motorists stopped by the same officer may be correlated. Including a random effect does not make any assumptions about why this correlation exists, and we recognize that the racial distribution of stopped motorists within officers is a result of many factors, including geographic area, unit assignment, or individual decision-making. Regardless of the reason for its existence, officer-level variation should be accounted for to maximize model accuracy.

Refinement 2: disentangle sex and race. Existing studies that use the VOD approach do not consistently elaborate on whether racial bias in traffic stops exists for both males and females. Rather, these studies assess the impact of daylight on the driver's race for both men and women simultaneously and therefore assume that Black men and women are at an equal risk of racial profiling. This approach appears defensible on its surface. Although not specific to the views of police officers, decades of research have suggested that Americans perceive African Americans in general to be disproportionately responsible for crime in the United States (e.g., Barlow, 1998; Drummond, 1990; Hawkins, 1995; Kennedy, 1997; Mauer, 1999; Russell, 2002), and they are more likely than any other group to be described as violent, drug abusers, and criminals (Chiricos, Welch, & Gertz, 2004; Hurwitz & Peffley, 1998; Sigelman & Tuch, 1996; Sniderman & Piazza, 1993; Welch, Chiricos, & Gertz, 2002). Feeding these perceptions, a disproportionate number of African Americans are under some type of correctional supervision (see Welch, 2007), and media coverage of crime is asymmetrically tilted toward violent crime, for which African Americans are more likely to be arrested compared with other racial groups (see Chiricos & Eschholz, 2002). Hence, the link between crime and race in the United States seems generalized enough to suggest that police officers may discriminate against African Americans generally, regardless of sex.

However, there is reason to believe that the racialization of crime is also heavily gendered. Although empirical research has not definitively suggested that police officers view Black males more negatively then Black females, numerous studies have documented perceptions of differential and discriminatory treatment from police among young Black men compared with treatment of Whites and women (Brunson, 2007; Brunson & Miller, 2006; Weitzer & Tuch, 2002). Mauer (1999) argues that the perceived associations between *Blackness* and crime became more narrowly focused on males in the 1970s and 1980s, during which the term *criminal predator* became a euphemism for young Black male. Others have identified the emergence of hip-hop music and culture as a key reinforcer of attitudes toward young Black males in contemporary U.S. society (Kitwana, 2002). Moreover, social upheaval in the United States regarding the treatment of African Americans by police has been largely inspired and perpetuated by police killings of Black *males* (e.g., Walter Scott in North Charleston, Eric Garner in New York, Tamir Rice in Cleveland). Most prominently, the Ferguson riots of 2014 were a direct consequence of community outrage regarding the killing of 18-year-old Michael Brown.

Using both self-report data and official agency records, research on other stages of traffic stops has found relationships between driver sex and traffic stop outcomes. Studies using official police data have found that female drivers were less likely to be ticketed or searched (Briggs, 2013), or arrested (Smith, Makarios, & Alpert, 2006) compared with male drivers. Studies using self-report data have found that being male *and* Black had an additive effect. Younger African American males, in particular, were much more likely to have additional enforcement actions taken after the initial stop (Engel & Calnon, 2004; Lundman & Kaufman, 2003; Tillyer & Engel, 2010). Not all results have provided such clear results. Blalock, DeVaro, Leventhal, and Simon (2011) found that female drivers were *more* likely to receive traffic tickets in three out of five jurisdictions they analyzed. These existing studies have focused on poststop outcomes. The current study advances this line of research by exploring the effect of driver sex on the decision to make the traffic stop.

In sum, it is possible that male and female African Americans are not at equal risk for profiling by the police, because officers may have different views regarding the extent to which African American males and females are involved in crime. Prior applications of the VOD approach ignore this possibility, while our refined application explores it by modeling stops for Black males and Black females separately. This explicit modeling allows us to better understand the relationship between sex and race of the driver.

Refinement 3: include officers' unit assignment. With the exception of Worden et al. (2012), which assessed the relationship between daylight and driver race for two specific units, prior applications of the VOD approach do not consider how unit assignment may impact racial patterns in traffic stops during the intertwilight period. Assessing variation in disproportionate minority contact across units is imperative because it helps to address whether identified racial bias toward minority drivers is organization wide or rather contained in specific domains within the department. One might hypothesize, for example, that driver characteristics such as sex and race might be less influential on the decision to conduct a stop for an officer tasked with traffic safety enforcement because these stops are theoretically predicated on relatively objective driving violations. Officers assigned with proactive crime reduction tasks, on the other hand, may be more likely to disproportionately target Black, and specifically Black male, drivers when making the decision to conduct a stop.

Furthermore, unique subcultures across units may contribute to disproportionate minority contact. Klinger's (1997) Ecological Theory of Policing, for example, argues that subgroup norms develop within organizational units and that these subgroup norms drive variations in activity and officer outputs. These subcultures may be perpetuated by normalizing officer behaviors that experience similar situations or the subculture may be the result of certain types of like-minded officers gravitating toward specific units (i.e., a self-selection process). Making these distinctions could have practical implications regarding staffing and task assignment decisions at the municipal leadership or executive level.

In the present study, we incorporate an enhanced VOD method to study the racial composition of traffic stops in Durham, North Carolina. Our analysis not only answers questions related to the extent to which racial bias is existent in Durham traffic stops, it also demonstrates the utility of our enhancements for improving the precision of VOD estimates.

Setting

To study the racial distribution of traffic stops in Durham, we implemented an enhanced version of the VOD approach using data describing 151,701 traffic stops conducted by the DPD from January 2010 through October 2015. Durham has a racially heterogeneous population of just under 250,000, served by 513 sworn officers and 104 civilian staff from the DPD. In addition to officers serving in the typical uniform patrol capacity, officers can also assigned to specialized teams, including high enforcement abatement team (HEAT), traffic services, and interdiction. Both traffic services and interdiction focus on conducting traffic stops as a primary public safety tool although their intentions are different. Traffic services, responsible mainly for traffic safety enforcement, operates out of one of the city's five districts while interdiction, responsible for investigating the transportation of illegal substances, is organized under the special operations division. Uniform patrol and HEAT operate out of each of the city's five geographic districts at the direction of district commanders. Uniform patrol units are primarily responsible for responding to calls for service while proactive activity is carried out as time allows. HEAT focuses on proactive problem-solving, quality of life complaints, and street-level drug transactions.

Importantly, the HEAT unit does not respond to routine calls for service, instead they are expected to largely self-direct their time and efforts.

Approximately 42% of Durham's population is White and 41% is African American, with Latinos and individuals of other races making up the remaining 17%. Although Durham is one of the birthplaces of the civil rights movement, which catalyzed significant strides in social equality, a series of incidents involving Durham police officers and people of color in recent years have acutely aroused racial tensions.

North Carolina state law requires law enforcement agencies to document the demographic characteristics, stop characteristics, and stop outcomes of all traffic stops. The stop data analyzed were sourced directly from the DPD and included additional information (e.g., the unit assignment of the officer who conducted the stop) not typically reported in the traffic stop data submitted to the state. After incorporating local civil twilight data from a public database maintained by the U.S. naval observatory, our analysis data set includes 19,801 traffic stops that occurred during the intertwilight period.

Analytic Strategy

We use a model building approach to explore and enhance the VOD methodology. Consistent with existing research, we use a logistic regression to estimate the relationship between daylight and driver race. Next, we demonstrate that there is statistically significant variability at the officer level and that failing to account for this variability violates the assumption of independent observations.

To explore the impact of officer-level correlation, we use multilevel modeling. Our outcome y is a binary variable that indicates whether the driver is Black. The grouping variable in our study is the officer; 509 officers indexed by the j subscript each conduct n_j traffic stops. The number of traffic stops conducted by officers ranges from 1 to 1,165 stops, with an average of about 39 stops.[1]

Fully specified models include 10 fixed effects predictors: Daylight (0 = did not occur during daylight of the intertwilight period, 1 = occurred during daylight of the intertwilight period); Day of the Week (dummy variables for Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday); Year (year in which traffic stop occurred); Time Bin; and Time Bin Quadratic.[2] We incorporate random intercepts for officers into the model because it is possible that traffic stops conducted by the same officer are more similar to another (i.e., more or less likely to involve a Black driver) than they are between officers. We pay special attention to covariance parameter estimates and the accompanying intraclass correlation (ICC) coefficients because these suggest the amount of variation in the probability of the driver being Black that is accounted for by the officer conducting the stop.

We present a series of regression models that collectively address the following research questions:

1. What is the relationship between light visibility and race of the driver stopped?
2. Is the relationship between light visibility and race of the driver contingent on the sex of the driver?
3. For male drivers, what is the relationship between light visibility and the race of the driver stopped by officers in different unit assignments?
4. Does the relationship between daylight and race change over time?

Results

Descriptive Statistics

Table 2 describes the race and sex distribution of drivers involved in DPD traffic stops from January 2010 through October 2015. The first set of columns represents all 151,701 traffic stops that occurred in the time period studied. The right two columns columns display frequency distributions and percentages for the 19,801 traffic stops that occurred during the intertwilight period. Traffic stops more commonly involved men than women in both the overall sample and in stops limited to the intertwilight period. The racial composition between all traffic stops and traffic stops occurring in the intertwilight period was relatively

Table 2. Race and Sex of People Stopped.

	Overall ($N = 151{,}701$)		ITP stops ($n = 19{,}801$)	
Characteristic	Frequency	%	Frequency	%
Sex				
Male	93,486	61.63	12,516	63.21
Female	58,215	38.37	7,285	36.79
Total	151,701	100.00	19,801	100.00
Race				
Asian	2,636	1.74	376	1.90
Black	89,236	58.82	11,859	59.89
Native American	884	0.58	114	.58
Unknown	647	0.43	71	.36
White	58,298	38.43	7,381	37.28
Total	151,701	100.00	19,844	100.00

ITP = intertwilight period.

Table 3. Race of People Stopped During Light and Dark Portions of the Intertwilight Period.

	N (%)	*N* (%) ITP	*N* (%) daylight ITP	*N* (%) darkness ITP
Black	89,236 (58.8)	11,859 (59.9)	5,847 (60.0)	6,012 (59.8)
White	58,298 (38.4)	7,381 (37.3)	3,611 (37.1)	3,770 (37.5)
All other	4,167 (2.8)	561 (2.8)	286 (2.9)	275 (2.7)
Total	151,701 (100)	19,801 (100)	9,744	10,057

ITP = intertwilight period.

consistent. For instance, the majority of motorists in both samples were Black. Asian, Native American, and motorists of an unknown race were involved in only a small percentage of all traffic stops, both overall and those occurring during the intertwilight period.

Table 3 describes the race of drivers during daylight and darkness portions of the intertwilight period ($n = 19{,}801$). The racial composition of drivers during daylight and darkness is almost identical. As shown, about 60% of drivers stopped during daylight and darkness were Black and about 37% of drivers were White. Roughly 3% of stops during daylight and darkness involved drivers of all other races.

Standard Logistic Regression Approach

Table 4 displays the results of a standard logistic regression approach to estimating the impact of daylight on driver race. The single-level logistic models find no association between daylight and driver race (Table 4, Model A). Based on these results, we would conclude that Black drivers are *not* overrepresented in Durham traffic stops. Models were respecified to estimate the effect of daylight on the male-only subsample (Table 4, Model B). Results demonstrate a statistically significant relationship between driver race and daylight, with drivers more likely to be Black when daylight, and presumably visibility, is higher ($OR = 1.14$).

Multilevel Models

An additional set of models were specified to explore the impact of nested observations (i.e., multiple traffic stops conducted by officers) within the data set. Table 5 presents a series of multilevel logistic regression models predicting the odds of the driver being Black during traffic stops for the overall sample, a subsample of stops in which the driver was male, a subsample of stops in which the driver was female, and subsamples of traffic stops conducted by

Table 4. Logistic Regression Predicting Race of Driver.

Model	Model A	Model B
Sex	All	Male
N	19,801	12,516
Variable	Coefficient (odds ratio)	Coefficient (odds ratio)
Occurred daylight (darkness = 0, daylight = 1)	0.05 (1.05)	0.1348*** (1.14)
Day of the week[a]		
Monday	−0.06 (0.94)	−0.01 (0.99)
Tuesday	−0.10 (0.90)	−0.09 (0.92)
Wednesday	−0.06 (0.95)	−0.01 (1.00)
Thursday	−0.05 (0.96)	0.02 (1.02)
Friday	0.10 (1.11)	0.14 (1.15)
Saturday	0.06 (1.06)	0.03 (1.03)
Year	0.03 (1.03)	0.03** (1.04)
Time bin[b]	−0.03 (0.97)	−0.07 (0.94)
Time bin quadratic[c]	0.01 (1.05)	0.01* (1.01)
Intercept	0.29**	0.17

Note.
[a]Sunday set as reference category.
[b]Time bin created by classifying each event into eight roughly 45-minute groups; earliest events occurring during the intertwilight period equal to 1.
[c]Time bin quadratic calculated by squaring the time bin.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

different units within the department. The final model in Table 5 includes an interactive term for daylight and year to assess whether the relationship between daylight and race of the driver changes over time.

We begin by demonstrating the statistical necessity for controlling officer-level effects. The first step is to estimate an unconditional analysis of variance (ANOVA) model (Model I), which allows for estimating the Black driver stop rate for a "typical officer" and provides information about the variability in the racial composition of stopped drivers between officers. The estimated Level-2 intercept (0.4334) represents the log odds of the driver being Black during a traffic stop conducted by a "typical officer." To make this estimate more meaningful, we calculate the predicted probabilities for Black drivers:

$$\frac{Exp.L2\ Intercept}{1 + Exp.L2\ Intercept}$$

Table 5. Logistic Regression With Random Intercepts Predicting Race of Driver.

				Male-only subsample					
	Model I: ANOVA	Model II: Overall	Model III: Females	Model IV: Males	Model V: HEAT	Model VI: Traffic	Model VII: Patrol	Model VIII: Interdiction	Model IX: Temporal
Sex	All	All	Female	Male	Male	Male	Male	Male	Male
Unit	All	All	All	All	HEAT	Traffic	Patrol	Interdiction	All
N	19,801	19,801	7,285	12,516	1,777	2,285	6,824	425	12,516
Measures (Odds Ratios [Coefficients])									
Occurred daylight	–	1.12***	.99	1.20***	1.44**	1.11	1.17**	2.66***	1.40***
(darkness = 0, daylight = 1)		(.03)	(.06)	(.04)	(.14)	(.10)	(.06)	(.26)	(.09)
Day of the week[a]									
Monday	–	1.01	.87	1.06	1.40	.62	1.14	.05	1.06
		(.07)	(.12)	(.09)	(.51)	(.49)	(.10)	(1.14)	(.09)
Tuesday	–	.99	.93	.99	.96	.56	1.07	.19	.99
		(.07)	(.11)	(.08)	(.43)	(.48)	(.10)	(.88)	(.08)
Wednesday	–	1.04	.95	1.07	1.05	.54	1.06	.36	1.07
		(.06)	(.11)	(.08)	(.40)	(.47)	(.10)	(.86)	(.08)
Thursday	–	1.06	.93	1.12	1.22	.59	1.15	.27	1.12
		(.06)	(.11)	(.08)	(.40)	(.47)	(.10)	(.86)	(.08)
Friday	–	1.16*	1.06	1.20*	1.24	.59	1.17	.46	1.20*
		(.06)	(.11)	(.08)	(.40)	(.48)	(.10)	(.89)	(.08)
Saturday	–	1.12	1.18	1.08	1.14	.67	1.04	.46	1.08
		(.06)	(.11)	(.08)	(.41)	(.48)	(.09)	(.89)	(.08)
Year	–	1.01	.97	1.03*	1.14**	1.00	1.01	1.11	1.05**
		(.01)	(.02)	(.01)	(.05)	(.03)	(.02)	(.09)	(.02)

(continued)

Table 5. (continued)

				Male-only subsample					
	Model I: ANOVA	Model II: Overall	Model III: Females	Model IV: Males	Model V: HEAT	Model VI: Traffic	Model VII: Patrol	Model VIII: Interdiction	Model IX: Temporal
Time bin[b]	–	.98	1.01	.96	.93	1.00	.96	.93	.96
		(.03)	(.06)	(.04)	(.12)	(.09)	(.07)	(.24)	(.04)
Time bin quadratic[c]	–	1.01	1.00	1.01	1.02	1.00	1.01	1.02	1.01
		(.00)	(.01)	(.00)	(.01)	(.09)	(.01)	(.03)	(.01)
Intercept	1.54***	1.27*	2.00***	1.01	1.09	1.11	1.08	2.12***	.93
	(0.37)	(.11)	(.17)	(.13)	(.51)	(.52)	(.19)	(1.04)	(.14)
Year × Daylight	–	–	–	–	–	–	–	–	.96†
									(.02)
Number of officers (Level 2)	509	509	448	486	74	19	376	14	486
Covariance parameter estimates									
Intercept	0.4334	.3674	.3666	.3796	.6543	.0021	.3288	.5720	.3800
ICC variance explained		10.27%	10.02%	10.34%	16.59%	0.10%	9.10%	14.81%	10.35%

Note. ICC = intraclass correlation.
[a]Sunday set as reference category.
[b]Time bin created by classifying each event into eight roughly 45-minute groups; earliest events occurring during the intertwilight period equal to 1.
[c]Time bin quadratic calculated by squaring the time bin.
†$p < .10$. *$p < .05$. **$p < .01$. ***$p < .001$.

After accounting for officer-level clustering, the ANOVA model indicates that there is a 60.7% chance that the driver involved in a traffic stop is Black. Alternatively, we could say that the predicted probability of the driver being non-Black is 39.3%

Turning to the Level-1 intercept, we note that there is statistically significant variability in the log odds of the driver being Black between officers (intercept = .3770, $z = 9.61$, $p < .0001$). This suggests that the probability of the driver being Black for a typical officer is 60.7%, but the probability of the driver being Black varied significantly across officers.

Next, we calculate the ICC to explore how much of the total variation in the probability of driver race is accounted for by the officer (Level 2) effect. We use the Level-1 intercept (0.38) to calculate the ICC[3]:

$$\frac{L1\ Intercept}{(L1\ Intercept + 3.29)}$$

For the ANOVA model, an ICC of 10.3% was calculated, suggesting that just over 10% of the variability in the proportion of Black drivers stopped is accounted for by officer-level effects. Stated another way, 90% of the variability in the proportion of drivers stopped is accounted for by event-level characteristics. Taken together, these results support the use of multilevel modeling for studying traffic stop outcomes.

With an established statistical reason for conducting multilevel models, we now turn to fully specified models with traffic stops characteristics modeled at Level 1 and officer effects modeled at Level 2. A likelihood ratio test confirmed that the fully specified model fit the data better than the unconditional model (LRT = 90.22, $df = 10$, $p = <.001$). Contrary to the overall single-level logistic model (Table 4, Model B), Model II (Table 5), including the overall sample ($n = 19{,}801$) without restrictions on unit or sex, shows a statistically significant relationship between daylight and driver race. The odds of the driver being Black were 12% higher when the stop occurred during daylight than when the stop occurred during darkness.[4] An assessment of the predicted probabilities for this model showed that stops conducted during daylight had a 62% chance of involving a Black motorist, whereas stops occurring during darkness had a 59% chance of involving a Black motorist.[5] Put in another context, we expect officers to stop 146 Black drivers for every 100 White drivers during hours of darkness when officers are less capable of determining driver race. During daylight hours, when visibility of the driver is higher, we expect officers to stop 164 Black drivers for every 100 White drivers. This increase in Black drivers stopped during daylight hours is indicative of racial disparity in traffic stops. While the differences here are small in magnitude, their statistical significance warrants further analysis to try and better quantify and describe this finding within more specific groups of drivers and officers.

Multilevel Models: Race–Sex Interaction

Models were respecified to explore the effect of driver sex on the relationship between daylight and driver race. The VOD method was applied to the subsample for females only (Table 5, Model III; $n = 7{,}285$) and for males only (Table 5, Model IV; $n = 12{,}516$). We did not find a statistically significant relationship between daylight and driver race among the female subsample. In the male-only subsample, the odds of the driver being Black were 20% higher when the traffic stop occurred during the daylight hours of the intertwilight period compared with the dark hours of the intertwilight period. It is important to point out that although the effect of daylight on the race of the driver was statistically significant among males for both the standard logistic regression model and the multilevel model, the latter model that accounts for officer-level variation indicates an odds ratio of 1.20, a difference of + .06 from the standard logit model. The predicted probabilities of the male-only multilevel model suggest that during darkness, there was a 55% chance that the motorist was a Black male. During daylight hours, the chance that the motorist was a Black male increased to 69%. In other words, we expect officers to stop 124 Black male drivers for every 100 White male drivers during dark hours. This figure increases to 149 Black male drivers during daylight hours. Together, these findings suggest that Black male and Black female drivers are not an equal risk for being stopped by the police.

Multilevel Models: Analysis by Unit Assignment

Models V to VIII in Table 5 present results from four logistic regression models predicting the odds of the driver being Black during traffic stops stratified by four DPD units: HEAT ($n = 1{,}777$), traffic ($n = 2{,}285$), patrol ($n = 6{,}824$), and interdiction ($n = 425$). Models were restricted to traffic stops involving male drivers.[6] These models show evidence of racial disproportionately in stops conducted by the HEAT, patrol, and interdiction units. For the HEAT (Model V),[7] the odds of the being Black during daylight hours were 44% higher than the odds during darkness. An assessment of the predicted probabilities for this model shows that stops conducted by the HEAT during darkness had a 69% chance of involving a Black male motorist, compared with a 76% chance in daylight. The relationship between lighting daylight and race of the driver was positive, but statistically nonsignificant among the traffic services unit (Model VI). Results from the subsample analysis of the patrol unit (Model VII) indicated that the odds of daylight stops' involving a Black male driver were 17% higher than the odds during darkness. More specifically, stops during darkness had a 55% chance of involving a Black male driver, compared with a 58% chance in daylight.[8] Finally, for the interdiction unit (Model VII), the odds of daylight stops involving a Black male driver were 166% higher than the odds during darkness. Stops during darkness had a 53% chance of involving a Black male driver, compared with a 73% in daylight.

Together, these results indicate that racial disproportionately in traffic stops is not equally present in traffic stops conducted by all units in the DPD. In addition, it is notable that the risk of Black male drivers being stopped is not equal across different types of units. Whereas the odds of a driver being a Black male during daylight are high among the HEAT and interdiction units, they are considerably lower for the patrols unit, and nonexistent in the traffic unit.

Multilevel Models: Temporal Effects

Temporal effects in the relationship between driver race and daylight were explored through additional models that included an interaction term between daylight and year. Results from the male-only subset analysis are presented. Model IX shows that the interactive term for year and daylight was statistically significant only at a relaxed *p*-value cutoff point of $p < .10$. Figure 2 displays predicted probabilities of the driver's being Black during daylight and darkness from 2010 to 2015.

In 2010, there was a 59% chance that the traffic stop involved a Black male driver if the stop occurred during daylight, compared with a 53% chance during darkness. In 2011, there was a 58% chance the driver was a Black male during daylight, compared with a 54% chance during darkness. The disparity was

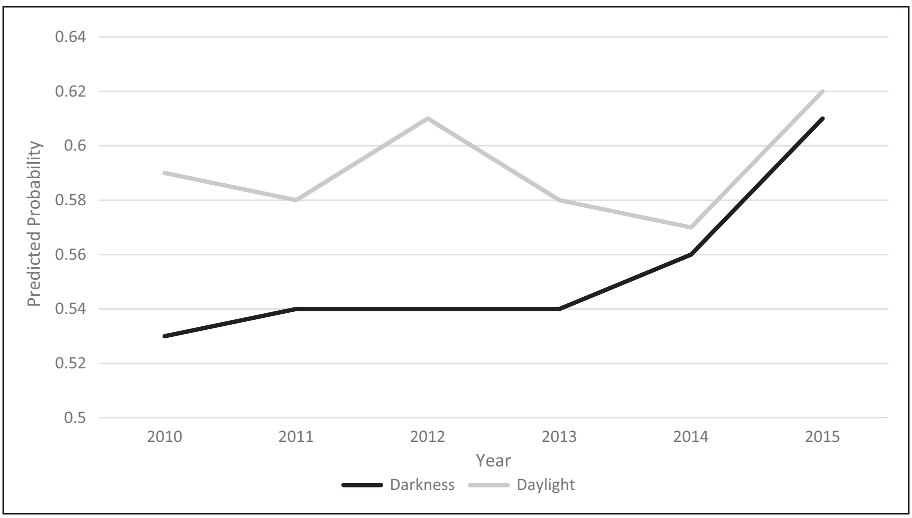

Figure 2. Effect of daylight on the predicted probability of driver race, over time for traffic stops involving males ($n = 12{,}516$).

largest in 2012; there was a 61% chance the driver was a Black male during daylight, compared with a 54% chance during darkness. The disparity between predicted probabilities during daylight and darkness in 2013 is very similar to that in 2011. As shown, in 2014, the chance of the driver's being Black during daylight was only about 1% higher than during darkness. Although the predicted probability for the traffic stop involving a Black driver in daylight increased in 2015 (relative to 2014), so did the predicted probability for the traffic stop's involving a Black driver during darkness. Separate models were run for each year of data (results omitted), which indicated that the relationship between daylight and driver race was not statistically significant in 2014 or 2015. This suggests that racial disparity in traffic stops was most prevalent during the earlier years of the analysis and that this disparity was declining over time.

Sensitivity Analyses

We conducted sensitivity analysis to assess the robustness of our findings because small variations in the VOD methodology have the potential to alter results. Four supplementary models were assessed (Table 6). First, the time between sunset and intertwilight (roughly 30 min) can be difficult to classify as light or dark. Therefore, Model X was specified so as to drops events that occurred between sunset and intertwilight. Compared with Table 5, results are substantively the same. Model XI considers only events that occurred within 30

Table 6. Additional Model Specifications.

	Overall		Male		Female	
Model specification	*n*	*OR*	*n*	*OR*	*n*	*OR*
Model X—Exclude the period of sunset to intertwilight	17,414	1.125**	11,003	1.218***	6,411	0.971
Model XI—Events within 30 days of switch to daylight saving time, exclude sunset to intertwilight	6,015	1.245*	3,761	1.258*	2,254	1.209
Model XII—Exclude drivers that were Hispanic, other than White or Black	16,578	1.166***	10,085	1.256***	6,493	1.015
Model XIII—Include controls for month	19,801	1.128*	12,516	1.221**	7,285	0.935

Note. Results presented as odds ratios. Models controlled for day of week, year, and time of stop as linear and quadratic terms. These coefficients are omitted for brevity. Models were specified as generalized linear mixed models where officer ID was treated as a random effect.
*$p < .05$. **$p < .01$. ***$p < .001$.

days before or after the switch to daylight saving time (DST). The DST switch causes a well-defined difference in daylight at the same time of day and limits the amount of variance that could be caused by seasonal differences in traffic patterns. Furthermore, we excluded cases that occurred between sunset and inter-twilight for reasons discussed earlier. Although a considerable number of cases were dropped, the results were generally the same. There were minor changes in the odds ratios, but the overall findings were consistent with less-restrictive models presented earlier.

Model XII dropped events that involved drivers who were Hispanic, Asian, Native American, or of unknown race so that models predicted the odds of the driver being Black versus the odds of the driver being White. Again, results were substantively the same.

Model XIII recognizes that there could be seasonal or monthly variation in driver characteristics (e.g., the influx of tourists during the summer) that could have an impact on the relationship between daylight and driver race. This model includes dummy variables that control for month of year in addition to standard temporal controls for time of day, day of week, and year. No noteworthy changes were observed from previously discussed models.

An alternative scenario that may invalidate our primary findings is that there is systematic underreporting of traffic stops that involve Black drivers. Given the lower confidence interval of the primary model (Table 5, Model II), DPD officers would need to have systematically underreport stops of Black drivers by 4.54% during dark hours (see Grogger & Ridgeway, 2006) for this to have occurred. For example, if officers reported 100% of traffic stops involving Black drivers during daylight, they would need to have to reported 95.46% of stops involving Black drivers during dark hours. Given state laws mandating this reporting, and the DPD's internal audits and checks on stop data submission, this level of systematic underreporting by officers is unlikely.

Note that across all models used for sensitivity testing, the relationship between driver race and daylight was nonsignificant in the female subsample. We found no evidence that race and daylight were associated for female drivers. The relationship between daylight and driver race was only present for male drivers.

Discussion

The VOD approach to analyzing traffic stops is powerful because it does not require an external benchmark to model the driving population at risk. Instead, it takes advantage of natural variation in lighting that occurs over time throughout the year is used to identify periods when it is dark at some times of the year and light during other times. This natural variation creates changes in the ability of officers to determine driver race, while avoiding the issue of comparing different times that may have different driving populations. For these reasons, the

VOD approach has received widespread and increasing acceptance in the criminal justice field in recent years.

The results of these analyses suggest that Black motorists are overrepresented in the Durham traffic stops that occur during the intertwilight period from January 2010 through October 2015. The overrepresentation of Black motorists in daylight traffic stops is confined to Black males suggesting that there is a sex–race interaction. This finding is consistent with other research that has found Black males in particular receive differential and discriminatory treatment by the police.

We also found evidence that the overrepresentation of Black drivers is present in most, but not all, unit assignments, and that the strength of the racial disproportionality varied considerably between units. From a practical perspective, these findings suggest that there may be training or other interventions that can be developed to help reduce racial disproportionality. It may be easier and more productive to intervene in the group dynamics that lead to racial disproportionality rather than to confront individual officers. Doing so may avoid the perception that officers are being singled out for their traffic stop activity and may help to avoid the "de-policing" that can occur when officers receive increased scrutiny (Rushin & Edwards, 2017). From a research perspective, these findings are interesting in that aggregate agency results may mask important variation in racial disproportionality. Future implementations of the VOD methodology should carefully consider organizational structure when investigating disproportionality.

Inclusion of an interaction term between daylight and year indicated that the racial disproportionality was declining year-over-year. Racial disproportionality was larger in the earlier years of analysis and by 2014 to 2015, there was a convergence of the predicted probabilities for daylight and darkness. This suggests that there is little evidence of Black overrepresentation in traffic stops conducted during these 2 years. The causes for this convergence are unknown and warrant further study. During this time period, DPD reported changes to training and policies designed to reduce racial disproportionality. Although the VOD method is not optimal for evaluating temporal changes, it is at least illustrative that further research in this area is warranted.

The use of multilevel modeling demonstrated that accounting for clustering of traffic stops within officers was important. Other studies that have not accounted for officer-level variance may be missing important information about the scale and strength of racial disproportionality in traffic stops. Had we only conducted the traditional single-level logistic model, we would have concluded that there was no evidence of disproportionality. We found that controlling for officer-level observational dependencies resulted in larger odds ratios and more models with statistically significant relationships between daylight and driver sex. It would be informative to go back to prior applications and conduct a similar set of analyses to determine if this pattern exists in other settings. These results suggest that existing literature may be under reporting racial disproportionality. Future research could also capitalize on the ability of multilevel models to better understand the officer characteristics that are associated with traffic stop behavior. Unfortunately, it was not possible to link these data sets in the current study. Other jurisdictions may not have this limitation, and the proposed methodology would be useful for clarifying these relationships.

The two largest effects of daylight on driver race were found in the HEAT and interdiction units; the models for these units also showed the highest proportion of officer-level effects (demonstrated by the model ICC). Although only speculation, there are at least two groups of factors that may be at play here. The first set of factors relate to officer qualities and characteristics, such as training and experience. Officer work location, especially, is likely to effect the racial composition of motorists stopped. Second, external factors, such as work group norms that may develop at the district level (Klinger, 1997), may influence the racial composition of drivers stopped. More research with greater attention to these potential factors would be needed to understand how and why the racial composition of motorists stopped by an officer varies so much between officers.

The results of these analyses also demonstrate the importance of capturing additional contextual details in the traffic stop data set. Our analyses demonstrate that the organizational unit is a key factor in exploring racial bias. This information is not available in the data that law enforcement agencies are required to submit to the state of North Carolina, and without it, it would not have been possible to determine that the traffic unit showed no evidence of racial disproportionality in traffic stops. Other factors that have been demonstrated as important in a VOD analysis (e.g., street lighting) can only be incorporated if stop data are spatially referenced (Horrace & Rohlin, 2014).

Limitations

First, the VOD approach assumes that there are no seasonal differences in the risk of drivers being stopped that are conditional upon daylight or darkness. Certain conditions (e.g., large population changes based on a university schedule or large seasonal changes in population) may invalidate this assumption. To test this assumption, we modeled events that occurred within 30 days before and after the switch to DST. These models, which are reported in Table 6, demonstrated no substantive differences from the results presented earlier.

Second, the VOD method is sensitive to factors that reduce the correlation between presumed visibility and darkness. Two situations are worth discussing. First, street lighting may alter the ability of officers to detect the race of drivers in darkness (Horrace & Rohlin, 2014). The extent to which artificial lighting enhances officers' abilities to detect driver race will negatively affect the assumptions underlying the model. Second, adverse weather conditions may make it darker earlier than indicated by the day's civil twilight. We note, however, that under both of these conditions, the exclusion of this ancillary information biases results toward null findings and that the inclusion of these ancillary data should *strengthen* the ability to detect a relationship between visibility and driver race. The results presented here are thus a *conservative* estimate of the impact of daylight on the race of the driver stopped.

Third, the VOD addresses only the question of racial disproportionality that may be occurring during the intertwilight period. Because the method requires variations in daylight during the same period, we are unable to assess racial disproportionality in time periods when it is always daylight (e.g., 3:00 p.m.) or always dark (e.g., 11:30 p.m.). Although these results are suggestive, it would be inappropriate to extrapolate findings to times outside of the analysis window.

Fourth, the VOD approach explores only the patterns of traffic stops and the racial composition of the drivers. This methodology makes no claims about the reasons, causes, or other factors that may influence the officer's decision to make a traffic stop. Relatedly, the VOD approach tells us nothing about other important stop characteristics, including the length of the stop, the outcome of the stop, or decisions to search. This approach also does not consider the impact of passengers on the decision to stop. Additional methods have been developed to assess these downstream decision points. For example, the Knowles, Persico and Todd (KPT) method of assessing bias in decisions to conduct vehicle searches as part of a traffic stops (Knowles, Persico, & Todd, 2001). These methods may provide a more definitive assessment of racial bias. We note, however, that the clear majority of traffic stops do not result in searches. Therefore, a holistic view of the event, which includes the initial stop and later actions taken by the officer, should be adopted.

Finally, location of the traffic stop was not collected by the DPD from 2010 to 2014. This omission prevents us from exploring the district-level variance in the relationship between daylight and driver race. Instead, we controlled for officer-level variance as a proxy for natural differences in the likelihood of stopping Black motorists that can be influenced by factors such as location. The DPD began recording location data in 2015, and this additional information may allow for further refinement of the VOD methodology and a better understanding of traffic stop patterns.

Conclusion

The VOD methodology provides a framework for assessing racial disproportionality in traffic stops that does not require an external benchmark. This article refines the VOD methodology by incorporating driver sex, disaggregating by unit assignment, exploring temporal trends through interactive terms, and using a random intercepts model to control for officer-level differences in stop activity. Findings suggest that Black males, not Black females, are overrepresented in traffic stops that occur during the daylight period of the intertwilight period relative to the darkness period. Future research in this area should consider the implications of structuring analyses in a way that explores the relationship between sex and race and racial disproportionality. Finding differences between units suggest that racial disproportionality is not uniform throughout the department but may vary based on the presumed mandates and goals of each group. Training and methods to remediate this racial disproportionality may need to be customized to the activities with which officers are tasked. In addition, there was evidence of longer term temporal trends that suggest the level of disproportionality was going down over time. Finally, we found that controlling for officer-level variance was critical to fully understanding the patterns of traffic stops. Prior research that ignored officer-level effects may be underestimating the scale of racial disproportionality.

Acknowledgments

The authors would like to thank the Durham Police Department (DPD) for providing access to the data used in this study. Jason Schiess, DPD's Analytic Services Manager, was especially helpful in understanding the strengths and nuances of these data.

Declaration of Conflicting Interests

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

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Notes

1. These numbers were produced using the restricted data set that included only traffic stops that occurred during the intertwilight period ($n = 19{,}801$).
2. Time bin was created by dividing the intertwilight period into eight equal temporal groups. The roughly 3.5 hours of civil twilight range was decomposed into eight equal blocks. The earliest block was assigned one, the second block two, and so on. The time bin quadratic variable was created by taking the square of the time bin. We explored alternative methods of controlling for time. We specified models with seven binary variables to represent time bins, 6-point linear time splines (consistent with Grogger and Ridgeway's original methodology), and 6-point quadratic time splines. These alternative specifications produced models that were largely the same and made no difference to any of the key findings.
3. Hierarchical generalized linear models assume no error at Level 1. Because of this, the ICC calculations must be amended slightly. As described by Ene, Leighton, Blue, and Bell (2014), ICCs are calculated assuming that the "outcome comes from an unknown latent continuous variable with a level-1 residual that follows a logistic distribution with a mean of 0 and a variance of 3.29" (p. 6).

4. A likelihood ratio test was performed to test the time control-only model (day of week, year, and time of stop indicators) against the fully fitted model that added the dark or light indicators. Models including the day or light indicator resulted in a statistically significant improvement in model fit.
5. This was calculated by taking 1/(1−mean) for that condition. Odds were calculated using the least squares MEANS command in SAS. All other model variables were set to their means.
6. Analyses were also conducted for stops involving female drivers separated by unit assignment. These models did not show statistically significant relationships between daylight and race of the driver.
7. Interdiction and traffic have relatively few Level-2 units, 14 and 19, respectively. There are no definitive rules on a minimum number of acceptable Level-2 units Snijders (2005). However, some guidance does exist. Snijders and Bosker (1999) recommend at least 10 Level-2 units. Simulation modeling has been done to explore the effect of varying Level-2 units. These studies have found that small (<50) Level-2 counts had negligible impact on regression coefficients, variance components, and Level-1 standard errors (Maas & Hox, 2005). Based on these findings, the authors conclude "if one is only interested in the fixed effects of the model, ten groups can lead to good estimates. If one is also interested in contextual effects, 30 groups are needed . . ." (p. 135). Given that the current analysis is only focused on the fixed effects model, the small Level-2 count found in the interdiction and traffic models is not problematic.
8. All predicted probabilities for the night or day indicator were calculated while holding the other variables in the model at their means.

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.

Barlow, M. H. (1998). Race and the problem of crime in *Time* and *Newsweek* cover stories, 1946 to 1995. *Social Justice*, *25*, 149–183.

Baumgartner, F. R., & Epp, D. (2012a). *North Carolina traffic stop analysis*. Retrieved from http://www.unc.edu/~fbaum/papers/Baumgartner-Traffic-Stops-Statistics-1-Feb-2012.pdf

Baumgartner, F. R., & Epp, D. (2012b). *North Carolina traffic stop statistics analysis*. Final report to the North Carolina Advocates for Justice Task Force on Racial and Ethnic Bias. Retrieved from http://www.unc.edu/~fbaum/papers/Baumgartner-Traffic-Stops-Statistics-1-Feb-2012.pdf

Blalock, G., DeVaro, J., Leventhal, S., & Simon, D. H. (2011). Gender bias in power relationships: Evidence from police traffic stops. *Applied Economics*, *43*(29), 4469–4485.

Briggs, J. S. (2013). *Routine justice: The intersection of race, gender, and police discretion in traffic stops* (doctoral dissertation). Retrieved from http://krex.k-state.edu/dspace/handle/2097/16228

Brown, B., & Benedict, W. R. (2002). Perceptions of the police: Past findings, methodological issues, conceptual issues and policy implications. *Policing: An International Journal of Police Strategies & Management*, *25*, 543–580.

Brunson, R. K. (2007). Police don't like Black people: African-American young men's accumulated police experiences. *Criminology and Public Policy*, *6*, 71–101.

Brunson, R. K., & Miller, J. (2006). Young Black men and urban policing in the United States. *The British Journal of Criminology*, *46*, 613–640.

Bureau of Justice Statistics. (2016). *Traffic stops*. Retrieved from http://www.bjs.gov/index.cfm?ty=tp&tid=702

Chiricos, T., & Eschholz, S. (2002). The racial and ethnic typification of crime and the criminal typification of race and ethnicity in local television news. *Journal of Research in Crime and Delinquency*, *39*, 400–420.

Chiricos, T., Welch, K., & Gertz, M. (2004). The racial typification of crime and support for punitive measures. *Criminology*, *42*, 358–390.

Cox, S. M., & Fitzgerald, J. D. (1996). *Police in community relations: Critical issues* (3rd ed.). Dubuque, IA: Brown and Benchmark Publishers.

Drummond, W. J. (1990). About face: Blacks and the news media. *American Enterprise*, *1*, 23–29.

Ene, M., Leighton, E., Blue, G., & Bell, B. (2014). *Multilevel models for categorical data using SAS PROC GLIMMIX: The basics* (Paper 134–2014). Retrieved from https://support.sas.com/resources/papers/proceedings15/3430-2015.pdf

Engel, R. S., & Calnon, J. M. (2004). Examining the influence of drivers' characteristics during traffic stops with police: Results from a national survey. *Justice Quarterly*, *21*, 49–90.

Greenwald, H. P. (2001). *Final report: Police vehicle stops in Sacramento, California*. Los Angeles, CA: University of Southern California.

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*, 878–887. doi:10.1198/016214506000000168

Hawkins, D. F. (1995). Ethnicity, race, and crime: A review of selected studies. In D. Hawkins (Ed.), *Ethnicity, race, and crime: Perspectives across time and place* (pp. 11–45). Albany, NY: State University of New York Press.

Hickman, M. J. (2005). Traffic stop data collection policies for state police, 2004. Bureau of Justice Statistics Fact Sheet. U.S. Department of Justice, Office of Justice Programs. NCJ 209156. Retrieved from http://www.bjs.gov/content/pub/pdf/tsdcp04.pdf

Horrace, W. C., & Rohlin, S. M. (2014, January). *How dark is dark? Bright lights, big city, racial profiling* (Unpublished manuscript). Retrieved from https://www.colgate.edu/docs/default-source/default-document-library/horracerohlindarkness-1-14-14.pdf?sfvrsn=0

Hurwitz, J., & Peffley, M. (Eds.). (1998). *Perception and prejudice: Race and politics in the United States*. New Haven, CT/London, England: Yale University Press.

Kennedy, R. (1997). *Race, crime and the law*. New York, NY: Vintage.

Kitwana, B. (2002). *The hip hop generation: Young Blacks and the Crisis in African-American Culture*. New York, NY: Basic Civitas Books.

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., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, *1*(109), 203–229.

Lamberth, J. (2003). Measuring the racial/ethnic makeup of traffic: The how, what and why. Paper presented at Confronting Racial Profiling in the 21st Century: Implications for Racial Justice, Boston, MA.

Lange, J. E., Johnson, M. B., & Voas, R. B. (2005). Testing the racial profiling hypothesis for seemingly disparate traffic stops on the New Jersey Turnpike. *Justice Quarterly*, *22*(2), 193–223.

Ludwig, J. (2003, May 13). Americans see racial profiling as widespread. *Gallup*. Retrieved from http://www.gallup.com/poll/8389/Americans-See-Racial-Profiling-Widespread.aspx

Lundman, R. J., & Kaufman, R. L. (2003). Driving while black: Effects of race, ethnicity, and gender on citizen self-reports of traffic stops and police actions. *Criminology*, *41*(1), 195–220.

Maas, C. J., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica*, *52*(2), 127–137.

Maas, C. J., & Hox, J. J. (2005). Sufficient sample size for multilevel modeling. *Methodology*, *1*(3), 86–92.

Mauer, M. (1999). *Race to incarcerate*. New York, NY: New Press.

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.

Oakland Police Department. (2004). *Promoting cooperative strategies to reduce racial profiling: A technical guide*. Oakland, CA: Author.

Ridgeway, G., Schell, T. L., Gifford, B., Saunders, J., Turner, S., Riley, K. J., & Dixon, T. L. (2009). *Police-community relations in Cincinnati*. Santa Monica, CA: RAND Center on Quality Policing Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG853.pdf

Riksheim, E., & Chermack, S. M. (1993). Causes of Police behavior revisited. *Journal of Criminal Justice*, *21*, 353–382.

Ritter, J. A., & Bael, D. (2009). Detecting racial profiling in Minneapolis traffic stops: A new approach. *CURA Reporter,* 11–17. Retrieved from http://hup.umn.edu/docs/DetectingRacialProfiling.pdf

Ross, M., Fazzalaro, 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.

Rushin, S., & Edwards, G. S. (2017). De-policing. *Cornell Law Review*. Retrieved from https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2757809

Russell, K. K. (2002). The racial hoax as crime: The law as affirmation. In S. L. Gabbidon, H. T. Greene & V. D. Young (Eds.), *African American classics in criminology and criminal justice* (pp. 349–376). Thousand Oaks, CA: Sage.

Sigelman, L., & Tuch, S. A. (1996). Metastereotypes: Blacks' perceptions of Whites' stereotypes of Blacks. *Public Opinion Quarterly*, *61*, 87–101.

Smith, D. A., Graham, N., & Adams, B. (1991). Minorities and the police: Attitudinal and behavioral questions. In M. J. Lynch & E. B. Patterson (Eds.), *Race and criminal justice* (pp. 22–35). Albany, NY: Harrow and Heston Publishers.

Smith, M. R., Makarios, M., & Alpert, G. P. (2006). Differential suspicion: Theory specification and gender effects in the traffic stop context. *Justice Quarterly*, *23*(2), 271–295.

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.

Sniderman, P., & Piazza, T. (1993). *The scar of race*. Cambridge, MA: Harvard University Press.

Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 3, pp. 1570–1573). Chicester, England: Wiley.

Snijders, T. A. B., & Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Tillyer, R., & Engel, R. S. (2010). The impact of drivers' race, gender, and age during traffic stops: Assessing interaction terms and the social conditioning model. *Crime and Delinquency*, *59*, 369–395.

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities? *Ohio State Journal of Criminal Law*, *6*, 231–275.

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.

Weitzer, R., & Tuch, S. (2002). Perceptions of racial profiling: Race, class, and personal experience. *Criminology*, *40*, 435–456.

Welch, K. (2007). Black criminal stereotypes and racial profiling. *Journal of Contemporary Criminal Justice*, *23*, 276–288.

Welch, K., Chiricos, T., & Gertz, M. (2002). *Racial typification of crime and punitive attitudes*. Paper presented at annual meeting of the American Society of Criminology, Chicago, IL.

Worden, R. E., McLean, S. J., & Wheeler, A. (2012). Testing for racial profiling with the veil-of-darkness method. *Police Quarterly*, *15*, 92–111.

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.

Author Biographies

Travis A. Taniguchi, PhD, is a research criminologist at the Center for Justice, Safety, and Resilience at RTI International. His research focuses on evaluations of law enforcement policy and effectiveness, crime and place, and methods of improving police–practitioner partnerships. He previously worked for the Redlands (CA) Police Department and the Police Foundation. His publications can be found in *Criminology*, *Justice Quarterly*, and *Deviant Behavior*.

Joshua A. Hendrix is a research criminologist at the Center for Justice, Safety, and Resilience at RTI International. His research areas are in deviance, delinquency, substance abuse, and informal and formal mechanisms of social control. He is the project director or principal investigator for multiple research studies funded by the Bureau of Justice Statistics or the National Institute of Justice, and his recent work has been featured in the *Journal of Experimental Criminology, Deviant Behavior,* and the *Journal of Family Issues*.

Alison Levin-Rector is a research public health analyst in the Center for Justice, Safety, and Resilience at RTI International. She has expertise in disease and syndromic surveillance, demographic methodology, advanced statistical modeling techniques, and survey design. She earned her MPH in Global Health Metrics and Evaluation in 2011 from the University of Washington.

Brian P. Aagaard is a research public health analyst at the Center for Justice, Safety, and Resilience at RTI International. He has experience working for law enforcement at local, county, state, and federal level. He earned his MA in criminal justice in 2008 for the State University of New York at Albany.

Kevin J. Strom is the program director of the Policing Research Program at RTI International. He has led numerous projects for the U.S. Department of Justice, including studies to increase efficiencies for processing forensic evidence and to develop strategic and tactical applications for 911 calls-for-service (CFS) data. He currently leads the National Crime Statistics Exchange (NCS-X) and the National Forensic Laboratory Information System (NFLIS) study, and his recent work has been featured in *Policing, Journal of Forensic Sciences*, and *Forensic Science Policy & Management*.

Stephanie A. Zimmer, PhD, is a research statistician in the Division of Statistics and Data Science at RTI International. She has experience in sampling design and analysis. Her recent work has focused on supporting nationally representative surveys of inmates, law enforcement agencies, and crime victims. She recently presented an analysis comparing the diversity of police forces to the communities they serve at the 2016 Joint Statistics Meetings.