



ARTICLE

An Empirical Assessment of Pretextual Stops and Racial Profiling

Stephen Rushin & Griffin Edwards*

Abstract. This Article empirically illustrates that legal doctrines permitting police officers to engage in pretextual traffic stops may contribute to an increase in racial profiling. In 1996, the U.S. Supreme Court held in *Whren v. United States* that pretextual traffic stops do not violate the Fourth Amendment. As long as police officers identify an objective violation of a traffic law, they may lawfully stop a motorist—even if their actual intention is to use the stop to investigate a hunch that by itself does not amount to probable cause or reasonable suspicion.

Scholars and civil rights activists have sharply criticized *Whren*, arguing that it gives police officers permission to engage in racial profiling. But social scientists have struggled to empirically evaluate how *Whren* has influenced police behavior.

A series of court decisions in the State of Washington presents an opportunity to test the effects of pretextual-stop doctrines on police behavior. In the years since the *Whren* decision, Washington has experimented with multiple rules that provide differing levels of protection against pretextual stops. In 1999, the Washington Supreme Court held in *State v. Ladson* that the state constitution barred police from conducting pretextual traffic stops. However, in 2012, the court eased this restriction on pretextual stops in *State v. Arreola*.

Exploiting a dataset of 8,257,527 traffic stops conducted by the Washington State Patrol from 2008 through 2015, we carry out difference-in-differences and triple-difference analyses to assess whether the *Arreola* decision increased traffic stops among drivers of color relative to white drivers. We find that the *Arreola* decision is associated with a statistically significant increase in traffic stops of drivers of color relative to white drivers. Further, we find this increase in traffic stops of drivers of color is concentrated during

* Stephen Rushin is an Associate Professor at Loyola University Chicago School of Law. Ph.D., J.D., University of California, Berkeley. Griffin Edwards is an Associate Professor at the University of Alabama, Birmingham. Ph.D., Emory University. This paper benefited from workshops at Vanderbilt Law School, Florida International University College of Law, and Loyola University Chicago. We thank our colleagues from across the country who provided us with helpful feedback and ideas related to this paper, especially Christy Lopez, Tracey Maclin, and Jordan Blair Woods. We also thank the editorial staff at the *Stanford Law Review*, especially Brett Oliver Parker and Hannah K. Song.

An Empirical Assessment of Pretextual Stops and Racial Profiling
73 STAN. L. REV. 637 (2021)

daytime hours, when officers can more easily ascertain a driver's race through visual observation.

These insights suggest that judicial decisions like *Whren* and *Arreola* increase the probability of racial profiling by police officers. We conclude by discussing the implications of these findings for the literature on police accountability.

Table of Contents

Introduction640

I. The Fourth Amendment and Pretextual Stops.....645

 A. *Whren v. United States*.....646

 B. Scholarly Criticism of *Whren*.....649

 C. State Departures from *Whren*.....652

 1. *State v. Ladson*: A ban on pretextual stops.....653

 2. *State v. Arreola*: The introduction of mixed-motive stops.....655

II. Existing Literature.....657

III. The Effects of *Arreola* on Police Behavior.....664

 A. Study Design.....665

 B. Dataset.....667

 C. Trends in the Raw Data.....673

 D. Effects of *Arreola* on Traffic Stops.....683

 E. Effects of Daylight on Traffic Stops.....690

 F. Event Study.....693

 G. Methodological Limitations.....695

IV. Implications for the Law of Policing.....697

 A. Harmful Consequences of *Whren*.....698

 B. Lack of Options for Redress.....700

 C. Decoupling Criminal Investigations and Traffic Enforcement.....702

Conclusion.....705

Appendix.....706

 A. Modeling Choices.....706

 1. Reliance on the number of traffic stops by race by county.....706

 2. Calculating and clustering of standard errors.....707

 3. Parallel-trends assumption.....707

 4. Use of triple-difference regressions and a placebo test.....708

 5. Additional controls and fixed effects.....711

 B. Alternative Models That Include Stops Where Race Is Unidentified as Stops of Nonwhite Drivers.....712

 C. Alternative Models Assuming Delayed Effect of Training in Mixed-Motive Stops.....716

 D. Effects of *Arreola* on Searches.....722

Introduction

In 1996, the U.S. Supreme Court held in *Whren v. United States* that pretextual traffic stops do not violate the Fourth Amendment.¹ As long as a police officer identifies an objective violation of a traffic law, the officer may lawfully stop a motorist—even if the officer’s actual intention is to use the stop to investigate a hunch that, by itself, would not amount to reasonable suspicion or probable cause.² In a unanimous decision, the Court concluded that an officer’s “[s]ubjective intentions play no role in ordinary, probable-cause Fourth Amendment analysis.”³

The scholarly response to *Whren* has been “overwhelmingly critical.”⁴ Modern traffic codes “regulate the details of driving in ways both big and

1. 517 U.S. 806, 809, 819 (1996) (“Here the District Court found that the officers had probable cause to believe that petitioners had violated the traffic code. That rendered the stop reasonable under the Fourth Amendment . . .”).

2. *See id.* at 812-13.

3. *Id.* at 813.

4. Gabriel J. Chin & Charles J. Vernon, *Reasonable but Unconstitutional: Racial Profiling and the Radical Objectivity of Whren v. United States*, 83 GEO. WASH. L. REV. 882, 884 & n.2, 886 (2015) (listing many existing studies supporting the proposition that *Whren* is “notorious for its effective legitimization of racial profiling in the United States”); *see also* David A. Sklansky, *Traffic Stops, Minority Motorists, and the Future of the Fourth Amendment*, 1997 SUP. CT. REV. 271, 274, 278-79 (summarizing *Whren* and providing contemporary context on the importance of the decision, and also noting that the *Whren* decision illustrated a “systematic disregard for the distinctive concerns of racial minorities”); David A. Harris, “Driving While Black” and All Other Traffic Offenses: The Supreme Court and Pretextual Traffic Stops, 87 J. CRIM. L. & CRIMINOLOGY 544, 545-46 (1997) (recognizing that while the *Whren* decision “makes some sense, at least from the point of view of judicial administration,” it ultimately could prove “profoundly dangerous” to the development of a “free society, especially one dedicated to the equal treatment of all citizens,” because it will allow for police to “use the traffic code to stop a hugely disproportionate number of African-Americans and Hispanics”); Tracey Maclin, *Race and the Fourth Amendment*, 51 VAND. L. REV. 333, 344-54 (1998) [hereinafter Maclin, *Race and the Fourth Amendment*] (presenting evidence about the possible link between pretextual stops and racial bias); Andrew D. Leipold, *Objective Tests and Subjective Bias: Some Problems of Discriminatory Intent in the Criminal Law*, 73 CHI.-KENT L. REV. 559, 565-72 (1998) (hypothesizing about how *Whren* may cause racial profiling, providing examples, and theorizing about how existing law may make it difficult for victims of racial profiling to succeed in any challenge); Anthony C. Thompson, *Stopping the Usual Suspects: Race and the Fourth Amendment*, 74 N.Y.U. L. REV. 956, 981, 983-98 (1999) (tracing the way that the Court has attempted to remove race from its consideration of Fourth Amendment issues and arguing that social-science data suggests racially neutral searches may still involve police relying on racial judgments); Tracey Maclin, *Cops and Cars: How the Automobile Drove Fourth Amendment Law*, 99 B.U. L. REV. 2317, 2347-49 (2019) [hereinafter Maclin, *Cops and Cars*] (providing a useful review of Sarah A. Seo’s book on the Fourth Amendment and the American automobile, *Policing the Open Road: How Cars Transformed American Freedom*, and discussing the impact of *Whren* on police behavior). *See generally* SARAH A. SEO, POLICING THE OPEN ROAD: HOW CARS TRANSFORMED AMERICAN FREEDOM (2019)
footnote continued on next page

small, obvious and arcane.”⁵ If an officer follows any motorist long enough, the motorist will eventually “violat[e] some traffic law,” making “any citizen fair game for a stop, almost any time, anywhere, virtually at the whim of police.”⁶ Scholars have suggested that when given this unfettered discretion, police officers will use it in a way that disproportionately targets motorists of color.⁷ And given the high bar that litigants must clear in order to prevail on a selective-enforcement claim under the Equal Protection Clause of the Fourteenth Amendment or its Fifth Amendment analog, the *Whren* decision left individuals of color with few means to challenge the discriminatory use of pretextual stops.⁸ Thus, scholars and activists have long worried that by allowing officers to engage in pretextual stops, *Whren* contributed to

(providing a comprehensive review of how the ubiquity of automobiles in the United States has transformed policing tactics and the law, and also emphasizing the importance of *Whren*).

5. Harris, *supra* note 4, at 545.
6. *Id.*; see also David A. Moran, *The New Fourth Amendment Vehicle Doctrine: Stop and Search Any Car at Any Time*, 47 VILL. L. REV. 815, 831 (2002) (“[T]ake any minor traffic or equipment violation, add a pretextual stop and a custodial arrest for the minor traffic violation, and *voila*, you get a lawful search of the automobile.”).
7. See Maclin, *Race and the Fourth Amendment*, *supra* note 4, at 344-46. Additionally, as one judge has argued, inherent to the use of pretextual stops is the risk that “some police officers will use the pretext of traffic violations or other minor infractions to harass . . . [groups based on factors such as their] race or ethnic origin, or simply appearances that some police officers do not like.” *United States v. Scopo*, 19 F.3d 777, 785-86 (2d Cir. 1994) (Newman, C.J., concurring).
8. See Angela J. Davis, *Race, Cops, and Traffic Stops*, 51 U. MIA. L. REV. 425, 427 (1997) (“The *Whren* Court left African-Americans and Latinos without an effective remedy for discriminatory pretextual traffic stops when it suggested the Equal Protection Clause as the appropriate constitutional basis for challenging these stops.”); Pamela S. Karlan, *Race, Rights, and Remedies in Criminal Adjudication*, 96 MICH. L. REV. 2001, 2010 (1998) (“As far as I can tell, with the exception of two New Jersey state court cases that antedate *Whren*, there are no reported cases in which suppression was the remedy for racially selective enforcement. And prior to *Whren*, the doctrinal handle for the suppression was the Fourth Amendment: the seizures were unreasonable because they were unconstitutional . . .” (footnote omitted)); see also Wesley MacNeil Oliver, *With an Evil Eye and an Unequal Hand: Pretextual Stops and Doctrinal Remedies to Racial Profiling*, 74 TUL. L. REV. 1409, 1477-79 (2000) (discussing a related issue—remedies to racial profiling—and arguing that the failure to correct patterns of racial bias in the face of a U.S. Department of Justice (DOJ) consent decree should potentially lead to evidentiary exclusion).

widespread and unchecked racial profiling⁹ by American police officers.¹⁰

Despite concern about the link between racial profiling and pretextual stops, no academic study to date has empirically evaluated the effect of *Whren* (or similar state cases) on law-enforcement behavior.¹¹ This gap in the literature stems not from a shortage of scholarly interest, but from the limited data available on police behavior and a lack of within-jurisdiction variation in pretextual-stop policies.¹² While many studies have found evidence of police officers engaging in racially biased behavior,¹³ no existing research has been able to empirically link pretextual-stop doctrines like *Whren* to subsequent patterns of racial profiling.¹⁴

-
9. We adopt the definition of racial profiling used in Samuel R. Gross & Debra Livingston, Essay, *Racial Profiling Under Attack*, 102 COLUM. L. REV. 1413, 1415 (2002) (defining racial profiling as occurring whenever an officer stops, questions, arrests, searches, or takes some other investigative action because the officer is operating under the belief that a person's racial or ethnic group makes them more likely than the community at large to commit the kind of offense under investigation).
 10. See *infra* Parts I.B, II (describing scholarly criticism of *Whren* and the existing literature on this topic). In fact, the DOJ was so concerned about the link between pretextual stops and racial profiling that it explicitly barred the Ferguson Police Department from engaging in pretextual stops as part of a broader federal consent decree. Consent Decree ¶ 80, *United States v. City of Ferguson*, No. 16-cv-000180 (E.D. Mo. Apr. 19, 2016), ECF No. 41 (stating that “officers will not conduct pretextual stops except where the actual reason for the stop is to investigate a felony”).
 11. Searches of Google Scholar, Westlaw, and Lexis produced no study that attempted to tackle this empirical question. But these searches produced dozens of empirical studies about the presence of racial profiling in individual agencies or states, often with hypotheses that this profiling was in part the result of laws permitting pretextual stops. For more information on the existing literature, see Part II below.
 12. As explained at greater length in Part III below, an ideal empirical assessment of the link between pretextual stops and racial profiling would require some court or legislature to issue a new rule related to pretextual stops that—to use the language of economics—served as an “exogenous shock” by unexpectedly overturning an existing rule or regulation on the topic. Provided that a jurisdiction kept sufficient data (including the race of those targeted for traffic enforcement) before and after this exogenous shock, researchers could use this data to evaluate whether the introduction of this new legal rule resulted in any corresponding changes in police behaviors. Unfortunately, the *Whren* decision did not present any obvious opportunities for such a controlled experiment, as at the time of the *Whren* decision, very few police departments kept data, including the race of those stopped by police for traffic infractions. See *infra* note 155 and accompanying text.
 13. See *infra* Part II (describing the existing literature, which finds racial profiling to be common among American police departments but fails to link this profiling specifically to court decisions like *Whren*).
 14. Prior studies have acknowledged that data on this proposition has proven “hard to come by.” Leipold, *supra* note 4, at 565; see also David Rudovsky, *Law Enforcement by Stereotypes and Serendipity: Racial Profiling and Stops and Searches Without Cause*, 3 U. PA. J. CONST. L. 296, 304 (2001) (“The failure of most law enforcement agencies to collect
- footnote continued on next page*

Employing a novel analysis of a newly available dataset, this Article is the first to illustrate empirically that judicial doctrines permitting police officers to engage in pretextual traffic stops contribute to a statistically significant increase in racial profiling of minority drivers.¹⁵ We focus our analysis on a series of legal events in State of Washington that presents a rare opportunity to analyze the effects of pretextual stops on police behavior. After the U.S. Supreme Court decided *Whren* in 1996, the Washington Supreme Court established differing levels of protection against pretextual stops in a series of opinions.¹⁶ In 1999, the Washington Supreme Court held in *State v. Ladson* that the state constitution barred police from conducting pretextual traffic stops.¹⁷ Then, in 2012, the court changed course in *State v. Arreola*, concluding that officers could conduct “mixed-motive traffic stop[s],” effectively legalizing the use of tactics akin to pretextual traffic stops.¹⁸ Thus, between 1999 and 2012, Washington effectively barred the use of pretextual stops.¹⁹ Since 2012,

and analyze data concerning car and pedestrian stops, or to conduct comprehensive reviews of the legality of stops and searches . . . has undermined efforts to make sound empirical judgements.”).

15. As best we can tell, the only other study to come close to making this type of claim is CHARLES R. EPP, STEVEN MAYNARD-MOODY & DONALD HAIDER-MARKEL, *PULLED OVER: HOW POLICE STOPS DEFINE RACE AND CITIZENSHIP* (2014). In that book, the authors distinguish between two different styles of traffic enforcement—one focused specifically on responding to traffic violations, and another focused on using traffic stops as pretexts for other types of criminal investigations. *Id.* at 14-15. The authors found that when police are involved in “investigatory stops,” the most important factor in who they stop is often the race and gender of the car’s occupants (Black young men were seemingly the most likely targets of investigatory stops). *Id.* at 66; see also Kathryn M. Young & Joan Petersilia, *Keeping Track: Surveillance, Control, and the Expansion of the Carceral State*, 129 HARV. L. REV. 1318, 1324-25 (2016) (reviewing EPP ET AL., *supra*). Nevertheless, that impressive book was unable to isolate the effect of changes in pretextual-stop doctrines on subsequent police behavior. Thus, to the best of our estimation—and as described in more detail in Part II below—we believe we are the first to conduct this kind of study.
16. See *infra* Parts I.C.1.-2 (describing how Washington moved from permitting pretextual stops after *Whren*, to outlawing their use after *State v. Ladson*, to presumptively allowing something akin to pretextual stops after *State v. Arreola*).
17. 979 P.2d 833, 842 (Wash. 1999) (en banc) (“We conclude the citizens of Washington have held, and are entitled to hold, a constitutionally protected interest against warrantless traffic stops or seizures on a mere pretext to dispense with the warrant when the true reason for the seizure is not exempt from the warrant requirement.”).
18. 290 P.3d 983, 991 (Wash. 2012) (en banc) (defining a “mixed-motive stop” as one where an officer has two separate, independent justifications for a traffic stop—one that gives the officer the legal justification to conduct a traffic stop and one that does not—and ultimately concluding that such mixed-motive stops do “not violate [Washington Constitution] article I, section 7 so long as the police officer making the stop exercises discretion appropriately”).
19. As we discuss in more detail in Part III.A below, the training given to law-enforcement professionals in Washington suggests that, after *Ladson*, officers were cautioned

footnote continued on next page

however, the police in Washington have operated under a narrowed definition of pretextual stops that more closely mirrors the holding in *Whren*.²⁰

We draw on a comprehensive dataset of 8,257,527 traffic stops conducted by the Washington State Patrol between 2008 and 2015 to examine the effect of *Arreola* on police behavior.²¹ The Washington State Patrol employs around 1,100 state troopers, who are primarily responsible for enforcing traffic laws on highways throughout the state.²² By employing a difference-in-differences framework, we find that *Arreola* is associated with a statistically significant increase in traffic stops and searches of nonwhite drivers relative to white drivers.²³ To further bolster our analysis, we use a triple-difference framework to observe the effect of daylight on officer behavior before and after *Arreola*.²⁴ We find that most of the increase in traffic stops of nonwhite drivers after *Arreola* occurred during the daytime, when police officers could more easily ascertain a driver's race.²⁵ This increase in traffic stops of nonwhite drivers during the daytime hours is also statistically significant.²⁶ These results support the hypothesis that judicial approval of pretextual stops contributes to racial profiling.

against pretextual stops (and similar behavior that courts may interpret as pretextual stops). But after *Arreola*, the training materials first introduced the concept of a “mixed-motive” stop, which was not present in any of the previous training updates. *See infra* Part III.A.

20. As we discuss in more detail in Part I.C.2 below, the majority in *Arreola* believed that it had created a new type of stop distinguishable from pretextual stops permitted by *Whren* but barred by *Ladson*. We take the view that even if the conduct permitted by *Arreola* is technically narrower than that permitted by *Whren*, it still represents a substantial increase in discretionary authority given to Washington police officers. We also endorse the concerns expressed by the dissent in *Arreola*, which did “not believe the spirit of *Ladson* [would] survive the court’s opinion” because police were now free to “stop citizens *primarily* to conduct an unconstitutional speculative investigation as long as they [could] claim there was an independent secondary reason for the seizure.” 290 P.3d at 993 (Chambers, J., dissenting).
21. *See infra* Parts III.A-.B (describing the dataset and methodology).
22. *About Us*, WASH. ST. PATROL, <https://perma.cc/V2CH-MZ6Q> (archived Jan. 4, 2021) (describing the number of commissioned and budgeted employees working for the Washington State Patrol and stating that there are around 1,100 commissioned employees and 1,100 civilian employees who handle approximately 3,092 contacts per day and around 1,128,642 contacts per year across the state’s thirty-nine counties).
23. *See infra* Part III.D (providing the regression outputs for these results).
24. *See infra* Part III.E (providing the regression outputs for these results focusing specifically on the relationship between daylight and police behavior). For an example of another study that utilizes such a triple-difference framework, see Jonathan Gruber, *The Incidence of Mandated Maternity Benefits*, 84 AM. ECON. REV. 622, 627 (1994).
25. *See infra* Part III.E.
26. *See infra* Part III.E.

Our analysis has important implications for the study of policing and criminal procedure. Our findings are consistent with one of the most common critiques of the *Whren* decision: that it leads to racial discrimination in policing. If the Washington Supreme Court's decision in *Arreola*, with its somewhat narrower holding than *Whren*, has contributed to a statistically significant increase in the targeting of drivers of color, then *Whren* may have had the same effect on policing all across the country. This increased targeting of drivers of color via pretextual stops is a matter of serious concern, as even routine traffic stops can escalate to more serious encounters involving the use of force, searches, and other coercive police actions.²⁷ More broadly, our findings suggest that legal rules granting police officers increased discretionary authority may create the risk of racially discriminatory law enforcement. This insight provides ammunition for scholarly proposals to decouple criminal investigations from traffic enforcement. It may also strengthen calls for the integration of technology into traffic enforcement so as to limit police discretion.

This Article proceeds in four parts. Part I summarizes the history of judicial regulation of pretextual stops, with a particular focus on the scholarly criticisms of *Whren* and the various sets of rules governing pretextual stops in Washington. Part II evaluates the existing literature on the relationship between pretextual stops and racial profiling. Part III sets out the methodology and results of our difference-in-differences and triple-difference frameworks. Part IV considers the implications of our findings.

I. The Fourth Amendment and Pretextual Stops

The Fourth Amendment protects against unreasonable searches and seizures by the government.²⁸ Police conduct is typically considered a seizure

27. For example, some have alleged that Sandra Bland was a victim of a pretextual stop, which ultimately resulted in her dying in police custody. In response, lawmakers in Texas have proposed banning police from stopping drivers as a pretext to investigate other potential crimes. *See, e.g.,* Jolie McCullough & Cassandra Pollock, *The Texas Lawmakers Who Led the Sandra Bland Act Are Pushing to Reinstate the Police Reforms Stripped from Their Original Bill*, TEX. TRIB. (June 9, 2020, 12:00 PM), <https://perma.cc/7WWH-R29K>; Amel Ahmed, *Sandra Bland, Samuel DuBose and the Rise of "Vehicular Stop and Frisk,"* AL JAZEERA AM. (July 30, 2015, 2:00 PM ET), <https://perma.cc/WCJ8-JTX9> (describing Sandra Bland's stop as a potential pretextual stop and using the death of Samuel DuBose as another example of where a pretextual stop may ultimately lead).

28. U.S. CONST. amend. IV:

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

for Fourth Amendment purposes if, under a totality of the circumstances, a police officer restrains a person's freedom of movement either through the use of force or through some show of authority.²⁹ Traffic stops entail a seizure of a driver "even though the purpose of the stop is limited and the resulting detention is quite brief."³⁰ A traffic stop is ordinarily considered reasonable for Fourth Amendment purposes when a police officer witnesses a traffic infraction and thus has probable cause to believe a traffic infraction has occurred³¹ or when a police officer has reasonable suspicion based on articulable facts that a criminal act is ongoing.³² In the years leading up to the *Whren* decision, federal courts of appeals were split on whether pretextual traffic stops complied with the Fourth Amendment.³³ This circuit split set the stage for *Whren*.³⁴

A. *Whren v. United States*

On June 10, 1993, police officers were patrolling a "high drug area" in Washington, D.C., when they observed two young Black men driving a vehicle in a manner that the officers alleged aroused their suspicions.³⁵ The vehicle sat at a stop sign for "what seemed an unusually long time—more than 20

29. Margaret M. Lawton, *The Road to Whren and Beyond: Does the "Would Have" Test Work?*, 57 DEPAUL L. REV. 917, 920 (2008); *Brendlin v. California*, 551 U.S. 249, 254 (2007) ("A person is seized by the police and thus entitled to challenge the government's action under the Fourth Amendment when the officer, 'by means of physical force or show of authority,' terminates or restraints his freedom of movement" (quoting *Florida v. Bostick*, 501 U.S. 429, 434 (1991))).

30. *Delaware v. Prouse*, 440 U.S. 648, 653 (1979).

31. *Id.* at 659; *see also Whren v. United States*, 517 U.S. 806, 810 (1996) ("As a general matter, the decision to stop an automobile is reasonable where the police have probable cause to believe that a traffic violation has occurred.").

32. *See* Wayne R. LaFave, *The "Routine Traffic Stop" from Start to Finish: Too Much "Routine," Not Enough Fourth Amendment*, 102 MICH. L. REV. 1843, 1846, 1848 (2004) (noting that most state authorities believe that reasonable suspicion is sufficient to justify a traffic stop); *see also* 4 WAYNE R. LAFAVE, SEARCH & SEIZURE: A TREATISE ON THE FOURTH AMENDMENT § 9.3(a) (5th ed. 2012) ("Most courts have assumed . . . that traffic stops as a class are permissible without probable cause if there exists reasonable suspicion, that is, merely equivocal evidence. Such an assumption is to be found in the federal court decisions of the various circuits, as well as in the decisions of most states." (footnote omitted)).

33. Lawton, *supra* note 29, at 922-23 (explaining that most circuits had concluded that pretextual traffic stops did not violate the Fourth Amendment; noting that the Ninth and Eleventh Circuits adopted more stringent tests; and also noting that the Tenth Circuit briefly adopted a reasonableness test before backtracking after finding that the test was "unworkable" and led to "inconsistent" results (quoting *United States v. Botero-Ospina*, 71 F.3d 783, 786 (10th Cir. 1995))).

34. *Id.* at 923.

35. *Whren*, 517 U.S. at 808, 810.

seconds.”³⁶ The officers also observed one of the youthful occupants of the car looking at the lap of another passenger.³⁷ When the police car made a U-turn to further investigate, the vehicle allegedly made a sudden right turn without signaling and drove away at an “unreasonable” speed.³⁸ The officers then pursued the vehicle for a short time before executing a traffic stop.³⁹ One officer observed two large plastic bags of crack cocaine in the hands of a passenger, Michael Whren.⁴⁰ The officers arrested Whren and the car’s driver, James Brown.⁴¹ In a search of the vehicle incident to the arrests, they discovered additional drugs.⁴²

Whren and Brown subsequently faced multiple drug-related charges.⁴³ They tried to suppress the evidence against them by arguing that the officers lacked probable cause or reasonable suspicion to conduct the original traffic stop.⁴⁴

The government maintained that the stop was objectively reasonable.⁴⁵ Regardless of their subjective intentions, the officers claimed they had an objectively reasonable basis on which to conduct a traffic stop because the driver of the vehicle engaged in “several traffic offenses.”⁴⁶

In response, Whren and Brown argued that this objective justification for the traffic stop was pretextual, and thus impermissible under the Fourth Amendment.⁴⁷ The actual reason that the officers pursued and stopped the

36. *Id.* at 808.

37. *Id.*

38. *Id.*

39. It is worth noting that the circumstances leading up to the traffic stop made it a bit unusual. The officers did not conduct an ordinary traffic stop, but instead followed the car for a period of time before pulling up beside it at a red light. *Id.* One officer then walked up to the side of the vehicle and asked for the driver to put the car in park. *Id.* At that point, he observed the drugs and made the arrests. *Id.* at 809. This effectively operated as a traffic stop for Fourth Amendment purposes, even if the circumstances were atypical. *Id.*

40. *Id.* at 809.

41. *Id.*

42. *Id.*

43. Specifically, they “were charged in a four-count indictment with violating various federal drug laws, including 21 U.S.C. §§ 844(a) and 860(a).” *Id.*

44. *Id.*

45. Brief for the United States at 7-8, 39-42, *Whren*, 517 U.S. 806 (No. 95-5841), 1996 WL 115816 (arguing for the use of an “objective assessment” of the actions of an officer and maintaining that the stop in this case was lawful (quoting *Maryland v. Macon*, 472 U.S. 463, 470-71 (1985))).

46. *Id.* at 39.

47. *Whren*, 517 U.S. at 810. As the Court explained, the petitioners argued that “in the unique context of civil traffic regulations” probable cause is not enough. Since, they

footnote continued on next page

vehicle was to investigate an unsubstantiated hunch.⁴⁸ The officers lacked reasonable suspicion or probable cause to stop the vehicle based on this hunch alone.⁴⁹ Thus, Whren and Brown argued, the officers impermissibly relied on a pretextual justification.⁵⁰ The petitioners also emphasized that the scope of the traffic code was so broad that, by following any driver long enough, a police officer could “invariably” identify some “technical violation” that could objectively justify a stop.⁵¹ Further, the petitioners argued that allowing officers to engage in pretextual stops would mean that officers “might decide which motorists to stop based on decidedly impermissible factors, such as the race of the car’s occupants.”⁵²

In a unanimous decision, the Supreme Court held that pretextual stops do not violate the Fourth Amendment.⁵³ Because Whren and Brown conceded on appeal that the officers had probable cause to believe that they had violated the traffic code, the Court’s analysis turned on whether the apparently pretextual nature of the stop turned an otherwise constitutional seizure into a violation of the Fourth Amendment. Walking through a series of major Fourth Amendment cases, the Court rejected the petitioners’ argument that the “constitutional reasonableness of traffic stops depends on the actual

contend, the use of automobiles is so heavily and minutely regulated that total compliance with traffic and safety rules is nearly impossible, a police officer will almost invariably be able to catch any given motorist in a technical violation.

Id. (quoting Reply Brief for the Petitioners at 1, *Whren*, 517 U.S. 806 (No. 95-5841), 1996 WL 164375).

48. *See id.*

49. *See id.*

50. The petitioners believed that their opposition to pretextual stops was consistent with *Florida v. Wells*, 495 U.S. 1 (1990), where the Court held that police cannot use an inventory-search program as a “ruse for a general rummaging in order to discover incriminating evidence,” as well as subsequent cases that analyzed whether inventory searches were merely pretexts for other unconstitutional behavior. *Whren*, 517 U.S. at 811 (quoting *Wells*, 495 U.S. at 4).

51. *Whren*, 517 U.S. at 810. It is also worth noting that, as Tracey Maclin has observed, the petitioners argued that police-department policy also barred

plainclothes officers from making routine traffic stops. Essentially, the defendants contended that a traffic stop cannot be constitutionally reasonable when officers violate their own departmental rules. The Court responded that this argument—to equate violation of departmental rules with constitutional wrongs—would make the traffic code “a dead letter at the option of the police department.”

Maclin, *Cops and Cars*, *supra* note 4, at 2348 (quoting Transcript of Oral Argument at 8, *Whren*, 517 U.S. 806 (No. 95-5841), 1996 WL 195296).

52. *Whren*, 517 U.S. at 810. Thus, the petitioners urged the Court to adopt a test that would require future courts to ask whether the officer, “acting reasonably, would have made the stop for the reason given.” *Id.*

53. *Id.* at 807, 819.

motivations of the individual officers involved.”⁵⁴ The Court also rejected the argument that the expansiveness of modern traffic codes meant that an objective test would give police officers unreasonably broad authority:

[W]e are aware of no principle that would allow us to decide at what point a code of law becomes so expansive and so commonly violated that infraction itself can no longer be the ordinary measure of the lawfulness of enforcement. And even if we could identify such exorbitant codes, we do not know by what standard (or what right) we would decide, as petitioners would have us do, which particular provisions are sufficiently important to merit enforcement.⁵⁵

The Court did acknowledge that “selective enforcement of the law based on considerations such as race” violates the Constitution.⁵⁶ But the Court concluded that victims of this type of selective enforcement must prove their claims under the Equal Protection Clause of the Fourteenth Amendment (or its analog in the Fifth Amendment’s Due Process Clause⁵⁷), not under the Fourth Amendment.⁵⁸ Thus, *Whren* stands for the proposition that police officers are permitted to engage in pretextual traffic stops—that is, stops justified by technical violations of the law but executed primarily so that the officer can investigate an unsubstantiated hunch (a hunch that, by itself, would not create constitutionally adequate suspicion).

B. Scholarly Criticism of *Whren*

Legal scholars have leveled three primary charges against *Whren*. First, some scholars have argued that *Whren* disproportionately harms drivers of color.⁵⁹ For example, Bennett Capers has written that *Whren* “essentially green-lighted the police practice of singling out minorities for pretextual traffic stops in the hope of discovering contraband” because *Whren* allowed

54. *Id.* at 811-13.

55. *Id.* at 818-19.

56. *Id.* at 813.

57. See *Bolling v. Sharpe*, 347 U.S. 497, 498-99 (1954).

58. *Whren*, 517 U.S. at 813.

59. See, e.g., Phyllis W. Beck & Patricia A. Daly, *State Constitutional Analysis of Pretext Stops: Racial Profiling and Public Policy Concerns*, 72 TEMP. L. REV. 597, 597 (1999) (“The primary concern with pretext stops is that they facilitate racial profiling, the process of singling out drivers based on their race.”); Abraham Abramovsky & Jonathan I. Edelstein, *Pretext Stops and Racial Profiling After Whren v. United States: The New York and New Jersey Responses Compared*, 63 ALB. L. REV. 725, 726 (2000) (“In other words, the *Whren* Court validated one of the most common methods by which racial profiles are put into effect—the pretext stop.”); Kami Chavis Simmons, *Beginning to End Racial Profiling: Definitive Solutions to an Elusive Problem*, 18 WASH. & LEE J. C.R. & SOC. JUST. 25, 29 (2011) (describing the holding of *Whren* and concluding that “highly discretionary [investigatory] stops permit racial bias, either explicit or implicit, to go unchecked and unpunished”).

police to “use race as an ‘unofficial’ proxy for suspicion.”⁶⁰ David Harris has argued that *Whren* leads police officers to “use the traffic code to stop a hugely disproportionate number of African-Americans and Hispanics.”⁶¹ Devin Carbado has bluntly concluded that after *Whren*, “at least under the Fourth Amendment, racial-profiling claims are not constitutionally cognizable” because “race matters in the Fourth Amendment context only to the extent that a police officer’s conduct is overtly racially coercive.”⁶² Kevin Johnson has written a detailed account of how *Whren*, alongside other major Supreme Court decisions, has “made legal challenges to profiling *more*, not *less*, difficult, thereby implicitly encouraging police officers to rely on racial profiles in law enforcement.”⁶³ And according to Anthony Thompson and others, psychological evidence suggests that race plays an integral role in police officers’ perceptions and subsequent behavior, which compounds the racial-profiling concerns associated with pretextual stops.⁶⁴

Second, scholars like Albert Alschuler,⁶⁵ Angela Davis,⁶⁶ and Pamela Karlan⁶⁷ have described how, by forcing litigants to bring all challenges to pretextual stops under the Equal Protection Clause rather than the Fourth Amendment, *Whren* functionally leaves victims of racial profiling with few remedies because of the “substantial hurdles” equal-protection claimants face.⁶⁸ Typically, an equal-protection claimant must prove that “a police officer intentionally discriminated against him based on his race,”⁶⁹ which is nearly

60. I. Bennett Capers, *Rethinking the Fourth Amendment: Race, Citizenship, and the Equality Principle*, 46 HARV. C.R.-C.L. L. REV. 1, 33-34 (2011).

61. Harris, *supra* note 4, at 546.

62. Devon W. Carbado, *(E)racing the Fourth Amendment*, 100 MICH. L. REV. 946, 1033, 1044 (2002).

63. Kevin R. Johnson, *How Racial Profiling in America Became the Law of the Land: United States v. Brignoni-Ponce and Whren v. United States and the Need for Truly Rebellious Lawyering*, 98 GEO. L.J. 1005, 1007 (2010).

64. See Thompson, *supra* note 4, at 983-91 (describing the social-science literature on how race inevitably influences police perceptions of potential suspects).

65. See Albert W. Alschuler, *Racial Profiling and the Constitution*, 2002 U. CHI. LEGAL F. 163, 168, 193 (claiming that “[t]he Court appeared to treat the Fourth Amendment and the Equal Protection Clause as hermetically sealed units whose principles must not contaminate one another,” and connecting this to the “difficulty of devising effective injunctive remedies for unlawful profiling”).

66. See Davis, *supra* note 8, at 435-38 (describing in detail the hurdles to recovery for victims of racial profiling after *Whren*, when read in conjunction with other cases).

67. See Karlan, *supra* note 8, at 2003-05 (describing how *Whren*, when read alongside *United States v. Armstrong*, makes it particularly difficult to obtain relief in the case of racial profiling).

68. Davis, *supra* note 8, at 427.

69. *Id.* at 436.

impossible to prove in the event of a pretextual stop—particularly given widespread evidence of police perjury.⁷⁰ In fact, Andrew Leipold⁷¹ and Tracey Maclin⁷² have worried about the effects of the *Whren* decision in light of the evidence of police officers’ willingness to lie on the stand in order to build cases against criminal defendants.

Third, scholars like Gabriel Chin and Charles Vernon have concluded that the *Whren* case was wrongly decided as a constitutional matter.⁷³ Chin and

70. See, e.g., Morgan Cloud, Essay, *The Dirty Little Secret*, 43 EMORY L.J. 1311, 1315 (1994) (stating that perjury by police officers “occurs most frequently when officers are testifying about searches and seizures and witness interrogations,” and noting that “[p]olice perjury about these topics is often the product of rules imposing penalties for illegal police practices”). See generally Andrew J. McClurg, *Good Cop, Bad Cop: Using Cognitive Dissonance Theory to Reduce Police Lying*, 32 U.C. DAVIS L. REV. 389 (1999) (discussing police lying); Carol A. Chase, *Rampart: A Crying Need to Restore Police Accountability*, 34 LOY. L.A. L. REV. 767 (2001) (advocating for reforms designed to address police lying).

71. See Leipold, *supra* note 4, at 562 (“Put bluntly, if police perjury is as common as some suspect, the likelihood of discovering an improper motive through the judicial process is slim indeed.” (footnote omitted)).

72. As Tracey Maclin has argued, “[o]ne need not accept that perjury is a pervasive problem in every police department to recognize that perjury (or the potential for perjury) may play a central role in how pretextual traffic stops are carried out.” See Maclin, *Race and the Fourth Amendment*, *supra* note 4, at 379-86 (arguing that “[p]olice often commit perjury” to “deny black and Hispanic motorists their substantive rights under the Fourth Amendment,” providing a detailed summary of the problem of police perjury, and discussing how *Whren* may exacerbate this issue).

73. See Chin & Vernon, *supra* note 4, at 887; see also Diana Roberto Donahoe, Essay, “*Could Have, Would Have: What the Supreme Court Should Have Decided in Whren v. United States*,” 34 AM. CRIM. L. REV. 1193, 1194-95 (1997) (criticizing the test adopted by the *Whren* Court, which leads to “arbitrary, unconstitutional searches and seizures,” and ultimately offering an alternative proposal). Admittedly, this quick summary of the scholarly backlash to *Whren* does not cover all articles and essays written on the topic. Numerous other scholars have also done important work in this area. We regret that we are unable to discuss all of these important works in the detail they deserve. For additional scholarship, see, for example, Alberto B. Lopez, *Racial Profiling and Whren: Searching for Objective Evidence of the Fourth Amendment on the Nation’s Roads*, 90 KY. L.J. 75, 80 (2001-2002) (providing a historical account of the *Whren* decision and situating it within the broader debate about racial profiling); Daniel B. Yeager, *The Stubbornness of Pretexts*, 40 SAN DIEGO L. REV. 611, 617-28 (2003) (providing an account of *Whren* and situating the case within previous doctrine); LaFave, *supra* note 32, at 1859 (“The totality of the Court’s analysis in *Whren* is, to put it mildly, quite disappointing. By misstating its own precedents and mischaracterizing the petitioners’ central claim, the Court managed to trivialize what in fact is an exceedingly important issue regarding a pervasive law-enforcement practice.”); and Jeffrey Fagan & Mukul Bakhshi, *New Frameworks for Racial Equality in the Criminal Law*, 39 COLUM. HUM. RTS. L. REV. 1, 8-9 (2007) (linking the differential treatment by police that African Americans and other racial and ethnic minorities experience to selective police enforcement). For an excellent and thorough summary of the scholarly critiques of *Whren*, see Lawton, *supra* note 29, at 928-32. These critiques have been echoed in prominent campaigns by civil rights organizations like the American Civil Liberties Union (ACLU), which alleged

footnote continued on next page

Vernon have argued that the Court should regulate pretextual stops in a manner similar to its previous regulation of abuses of prosecutorial discretion.⁷⁴ Chin and Vernon also believe that the Court's desire for Fourth Amendment objectivity leaves ample room for the regulation of the subjective motivations of police officers in executing traffic stops.⁷⁵ Yet despite the widespread concern about *Whren*, few states have enacted limitations on the use of pretextual stops, as discussed in more depth in the next Subpart.

C. State Departures from *Whren*

States have generally not strayed far from the core holding of *Whren*. State constitutions sometimes include their own versions of the Fourth Amendment, limiting the ability of state law enforcement to engage in unreasonable searches or seizures.⁷⁶ Many of these state constitutional provisions are broader than their federal counterpart, limiting the ability of state law enforcement to engage in conduct that might otherwise be permitted under the U.S. Constitution.⁷⁷ And state legislatures are also free to pass legislation limiting the ability of police officers in their states to engage in pretextual stops. Nevertheless, very few state supreme courts or state legislatures have established more stringent limitations on pretextual stops than those articulated in *Whren*. A 2016 analysis by Margaret Lawton suggests that at least two states—New Mexico and Washington—have established some limitations on pretextual stops via judicial rulings.⁷⁸

that *Whren* would lead police to target those “Driving While Black.” Carbado, *supra* note 62, at 1035-40 (describing and quoting from an ACLU pamphlet); *see also* David A. Harris, *Driving While Black: Racial Profiling on Our Nation's Highways*, ACLU (June 1999), <https://perma.cc/FRE4-HFBZ>.

74. Chin & Vernon, *supra* note 4, at 902.

75. *See id.* at 904-12.

76. *See infra* Part I.C.1 (describing the application of the State of Washington's version of the Fourth Amendment to regulate police behavior).

77. *See, e.g.*, Jack L. Landau, *Should State Courts Depart from the Fourth Amendment? Search and Seizure, State Constitutions, and the Oregon Experience*, 77 MISS. L.J. 369, 394-95 (2007) (explaining how Oregon began departing from the Fourth Amendment and “remain[s] free . . . to interpret [its] own constitutional provision regarding search and seizure and to impose higher standards on searches and seizures under [its] own constitution than are required by the federal constitution” (quoting *State v. Caraher*, 653 P.2d 942, 947 (Or. 1982) (en banc))).

78. Margaret M. Lawton, *State Responses to the Whren Decision*, 66 CASE W. RESV. L. REV. 1039, 1040-41 (2016) (discussing these two state departures from *Whren*). Rhode Island has passed a law that explicitly outlaws racial profiling, but does not directly address the issue of pretextual stops. *See Comprehensive Community-Police Relationship Act of 2015*, 31 R.I. GEN. LAWS ch. 21.2 (2021). For more on New Mexico's rule, see Michael Sievers, Note, *State v. Ochoa: The End of Pretextual Stops in New Mexico?*, 42 N.M. L. REV. 595, 595-96 (2012). For more on Washington's rule, see Parts I.C.1-2 below.

Among these two states, Washington is distinctive in its experimentation with different rules regulating pretextual stops over time. The Washington Supreme Court first acted to prohibit pretextual stops in *State v. Ladson*.⁷⁹ Then, thirteen years later in *State v. Arreola*, the same court backtracked by redefining the term “pretextual” to apply to a relatively narrow set of factual circumstances.⁸⁰ As a result, Washington is unique among American states in its variation of the rules governing pretextual stops by law-enforcement officers. The Subparts that follow walk through this recent history of judicial regulation of pretextual stops in Washington.

1. *State v. Ladson*: A ban on pretextual stops

The Washington Supreme Court first considered the constitutionality of pretextual traffic stops in 1999 in *State v. Ladson*.⁸¹ That case originated in a traffic stop conducted by Officer Jim Mack of the Lacey Police Department and Detective Cliff Ziesmer of the Thurston County Sheriff’s Department in October 1995.⁸² While working on a gang patrol, Officers Mack and Ziesmer became suspicious of a car driven by a Black man named Richard Fogle.⁸³ The officers recognized Fogle as the suspect from an “unsubstantiated street rumor” involving drugs.⁸⁴ This rumor did not give the officers the necessary reasonable suspicion required to execute a traffic stop.⁸⁵ Nevertheless, the officers followed Fogle’s vehicle until they noticed that his license-plate sticker had recently expired.⁸⁶ The officers did not deny that they used this expired license-plate sticker as a pretext to justify stopping Fogle’s vehicle so that they could investigate the unsubstantiated rumor.⁸⁷

After discovering that Fogle had a suspended license, the officers arrested him and searched his car incident to arrest.⁸⁸ They also ordered Fogle’s passenger, a Black man named Thomas Ladson, out of the car and patted him

79. See *infra* Part I.C.1.

80. See *infra* Part I.C.2.

81. 979 P.2d 833 (Wash. 1999) (en banc).

82. *Id.* at 836.

83. *Id.*

84. *Id.* (noting that the trial court found that “Officer Mack’s suspicions about Fogle’s reputed drug dealing was his motivation in finding a legal reason to initiate the stop of Fogle’s vehicle” (quoting the lower-court record)).

85. See *id.* (describing how the officers “tailed the Fogle vehicle looking for a legal justification to stop the car,” going as far as “shadow[ing] the vehicle while it refueled at a local filling station”).

86. *Id.*

87. *Id.* (“The officers do not deny the stop was pretextual.”).

88. *Id.*

down.⁸⁹ After finding a small handgun, the officers placed Ladson under arrest, at which point the officers found \$600 in cash and some small baggies of marijuana while searching Ladson's person.⁹⁰ The state charged Ladson "with unlawful possession of a controlled substance with intent to deliver while armed with a deadly weapon, and possession of a stolen firearm."⁹¹ Ladson moved to suppress the evidence obtained during the traffic stop, arguing that it was the result of a pretextual stop in violation of article I, section 7 of the Washington Constitution.⁹² Article I, section 7 states that "[n]o person shall be disturbed in his private affairs, or his home invaded, without authority of law."⁹³ While *Whren* established a permissive standard for pretextual stops, the Washington Supreme Court observed that article I, section 7 of the Washington Constitution had previously been found to be more protective than the U.S. Constitution.⁹⁴ Thus, the question raised by *Ladson* was whether the more protective Washington Constitution prohibited pretextual traffic stops, even if the U.S. Constitution did not.

In a 5-4 decision, the court held that the use of pretextual stops violated the Washington Constitution.⁹⁵ As the majority explained:

[T]he problem with a pretextual traffic stop is that it is a search or seizure which cannot be constitutionally justified for its true reason (i.e., speculative criminal investigation), but only for some other reason (i.e., to enforce traffic code) which is at once lawfully sufficient but not the *real* reason.⁹⁶

Permitting pretextual stops would effectively be prioritizing "form over substance" and would represent a "triumph of expediency at the expense of reason."⁹⁷ Thus, the court reasoned, stops based on a bare suspicion of wrongdoing, like that of Mr. Ladson, were "inherently unreasonable" even if an officer were able to identify some minor, pretextual justification.⁹⁸

The court closed by giving guidance to officers and trial courts for deciding whether a traffic stop was pretextual and thus unconstitutional. It

89. *Id.*

90. *Id.*

91. *Id.*

92. *Id.* at 836-37 (describing Ladson's argument that "the state constitution provides broader protection than does the Fourth Amendment in the area of pretextual traffic stops").

93. *Id.* at 837 (quoting WASH. CONST. art. I, § 7).

94. *Id.*

95. *Id.* at 842-43 (concluding that evidence obtained via such a pretextual stop must also be suppressed).

96. *Id.* at 838 (emphasis added).

97. *Id.*

98. *Id.* at 839 (citing cases where the court had previously expressed concern about pretextual stops).

instructed that courts ought to examine the “totality of the circumstances,” including both the “subjective intent of the officer” and the “objective reasonableness of the officer’s behavior.”⁹⁹ This remained the law from July 1, 1999, until the Washington Supreme Court sharply changed course thirteen years later.

2. *State v. Arreola*: The introduction of mixed-motive stops

In December 2012, in *State v. Arreola*, the Washington Supreme Court again considered the constitutionality of pretextual traffic stops under article I, section 7 of the state constitution.¹⁰⁰ *Arreola* was presented not as a direct challenge to the core holding of *Ladson*, but as a clarification of the definition of the term “pretextual.”¹⁰¹ The case arose out of the traffic stop of Gilberto Chacon Arreola by Officer Tony Valdivia in Mattawa, Washington, on October 10, 2009.¹⁰² Officer Valdivia received a tip about a possible drunk driver in a vehicle similar to Arreola’s.¹⁰³ Officer Valdivia followed Arreola for around thirty to forty-five seconds without observing any behaviors consistent with intoxicated driving.¹⁰⁴ Then he noticed that the vehicle had an altered exhaust system, which technically violated Washington traffic code.¹⁰⁵ Officer Valdivia then executed a traffic stop.¹⁰⁶ The trial court found that Officer Valdivia’s primary motivation for pulling over Arreola’s vehicle was to investigate the drunk-driving tip.¹⁰⁷ Nevertheless, Officer Valdivia insisted that the altered exhaust system was another “actual reason for the stop.”¹⁰⁸ And

99. *Id.* at 843.

100. 290 P.3d 983, 986 (Wash. 2012) (en banc).

101. *Id.* (“The issue in this case is whether a traffic stop motivated primarily by an uncorroborated tip, but also independently motivated by a reasonable articulable suspicion of a traffic infraction, is unconstitutionally pretextual under article I, section 7 of the Washington State Constitution and *State v. Ladson*.” (citation omitted)).

102. *Id.* at 986-87.

103. *Id.* at 986.

104. *Id.* at 986-87.

105. *Id.*

106. *Id.* at 987 (“Still without any signs of intoxicated driving, Officer Valdivia then activated his overhead lights and pulled over the car.”).

107. *Id.* (noting that Officer Valdivia continued to claim that the tip was not the only reason for the stop and clarifying that “he would sometimes commence a traffic stop for an altered muffler because, as a member of the community, he appreciates concerns about the excessive noise that such mufflers emit”).

108. *Id.* (noting that Officer Valdivia further justified his decision by explaining that while he would not normally “go out of his way to chase down a car with an altered muffler, he often would commence a traffic stop if already on the road and behind such a vehicle, so long as conducting the stop would not hinder a more pressing investigation”).

he claimed that while it was the uncorroborated tip that led him to follow Arreola's vehicle, he "would have stopped the vehicle, once following it, even if he wasn't suspicious of a DUI."¹⁰⁹

Thus, the issue in *Arreola* was whether Officer Valdivia's actions constituted an impermissible "pretextual" stop within the meaning of *Ladson*.¹¹⁰ Ultimately, the court narrowed the core holding of *Ladson* significantly, holding that Officer Valdivia's behavior constituted a lawful "mixed-motive" stop rather than an impermissible "pretextual" stop.¹¹¹ To delineate between constitutionally permissible mixed-motive stops and impermissible pretextual stops, the court contrasted the facts in *Ladson* and *Arreola*. According to the majority, the officer in *Ladson* had admitted to the court that he was relying on "a false reason" intended to disguise his "real motive."¹¹² The officer in *Ladson* likely "would not have conducted the stop had there been no street rumor," meaning that the officer "abused his discretion by conducting the stop without deeming it reasonably necessary to enforce license plate tab regulations."¹¹³ In contrast, the officer in *Arreola* testified that the muffler violation was an "actual" and independent justification for the traffic stop apart from the unsubstantiated tip.¹¹⁴ In the majority's view, this made the stop in *Arreola* a permissible mixed-motive stop distinguishable from the pretextual stop in *Ladson*.¹¹⁵

The dissent in *Arreola* argued that the majority opinion fundamentally redefined and narrowed *Ladson*'s prohibition on pretextual stops so as to make it virtually unrecognizable.¹¹⁶ The dissent worried that the *Arreola* majority relied on a tenuous distinction between the term "real" and the term "primary."¹¹⁷ The majority opinion said that a police officer may lawfully conduct a traffic stop where the "primary" motivation is a desire to investigate a hunch, but officers are still barred from conducting pretextual stops where the investigation of a hunch is the "real" reason for the stop.¹¹⁸ How, then,

109. *Id.* (quoting the lower-court record).

110. *Id.* at 986.

111. *Id.* at 991 ("We hold that a traffic stop is not unconstitutionally pretextual so long as investigation of either criminal activity or a traffic infraction (or multiple infractions), for which the officer has a reasonable articulable suspicion, is an actual, conscious, and independent cause of the traffic stop.")

112. *Id.* (emphasis omitted) (quoting *State v. Ladson*, 979 P.2d 833, 843 n.11 (Wash. 1999) (en banc)).

113. *Id.*

114. *Id.* at 987.

115. *Id.* at 991.

116. *Id.* at 992-93 (Chambers, J., dissenting).

117. *Id.* at 993.

118. *Id.*

should courts distinguish between “real” and “primary” motivations? In reality, the dissent argued, this distinction is practically meaningless; the rule adopted by the court effectively enabled the police in Washington to engage lawfully in pretextual stops.¹¹⁹

Reasonable readers may disagree on the real-world implications of the Washington Supreme Court’s decision in *Arreola*. Nevertheless, the decision substantially narrowed the holding of *Ladson* by giving police permission to engage in mixed-motive traffic stops that, at minimum, resemble pretextual stops. Thus, *Arreola* represented an expansion of law-enforcement power to execute discretionary traffic stops against motorists in Washington.

II. Existing Literature

An extensive and growing body of literature suggests that police treat drivers of color differently than white drivers.¹²⁰ These studies commonly find that police are more likely to subject drivers of color to stops, searches, and other coercive actions compared to white drivers.¹²¹ In many of these cases, differences in driving behavior do not explain this differential treatment.¹²² All of this suggests that police in a wide number of jurisdictions may consider a driver’s race (either implicitly or explicitly) in making traffic-enforcement decisions. Some studies point to *Whren*—and the broad discretion and deference given to police generally—as contributing to this type of racial profiling.¹²³ But no study to date has empirically tested the link between *Whren* (and its state law equivalents) and racially biased behavior by police officers.

Shortly after the *Whren* decision, David Rudovsky wrote a detailed summary of the then-existing universe of studies on racial profiling by police in traffic stops.¹²⁴ At that point, studies of the New Jersey State Police,¹²⁵

119. *Id.* (explaining that police officers were now “free to stop citizens *primarily* to conduct an unconstitutional speculative investigation as long as they can claim there was an independent secondary reason for the seizure”).

120. *See infra* notes 124-29 and accompanying text.

121. *See infra* notes 124-29 and accompanying text.

122. *See infra* notes 124-29 and accompanying text (citing many studies showing that underlying rates of offending likely do not explain differences observed in law-enforcement behavior).

123. *See, e.g., infra* notes 152-53 and accompanying text.

124. Rudovsky, *supra* note 14, at 299-306 (providing a detailed summary of the then-existing literature on racial profiling).

125. PETER VERNIERO & PAUL H. ZUBEK, INTERIM REPORT OF THE STATE POLICE REVIEW TEAM REGARDING ALLEGATIONS OF RACIAL PROFILING 26-28, 111 (1999), <https://perma.cc/87SU-TCLV> (finding that 77.2% of all searches in the data involved Black or Hispanic drivers and generally evaluating the evidence for racial profiling in traffic stops).

Illinois State Police,¹²⁶ Philadelphia Police Department,¹²⁷ New York City Police Department (NYPD),¹²⁸ and Boston Police Department¹²⁹ all showed evidence of racial profiling by law enforcement. Indeed, Rudovsky's summary from two decades ago foreshadowed a research field that has since grown substantially. Today, racial-profiling research is a major field of study, with academics,¹³⁰ government agencies,¹³¹ and nonprofits¹³² all regularly producing studies. These studies have attempted to document the presence of racial profiling through a relatively common set of methodological approaches. They often collect data on the frequency of police stops and searches of white and nonwhite drivers.¹³³ They then generally compare these stop and search rates with some baseline to determine whether police are treating drivers of color differently than we would expect given the underlying population breakdowns, rates of traffic-code violations, or other reference points.¹³⁴

126. Harris, *supra* note 73 (finding that although Latinos made up less than 8% of the Illinois population and less than 3% of the motorists in the state, they made up 30% of the motorists stopped by drug-interdiction officers).

127. Rudovsky, *supra* note 14, at 301 (pointing out how in predominantly white police districts of Philadelphia, African Americans were roughly ten times more likely to be stopped through either vehicle or pedestrian stops than one would expect based on their representation in the underlying population).

128. *Id.* at 302 (explaining how a 1999 analysis by the New York Attorney General of 175,000 pedestrian stops found that African Americans were approximately six times more likely to be stopped than whites, and that these imbalances were still statistically significant after adjusting for crime rates by race).

129. *Id.* at 302-03 (describing how Boston Police Department officers engaged in racially biased stops and searches of minority individuals).

130. *See, e.g., infra* notes 140-48 and accompanying text.

131. *See, e.g., supra* note 125 and accompanying text; *Vehicle Stops Report: 2020 Vehicle Stop Report Changes*, MO. ATT'Y GEN., <https://perma.cc/HS2U-GSMC> (archived Feb. 22, 2021) (providing annual reports on the racial breakdown of vehicle stops in Missouri). *See generally* ANNUAL REPORT 2020, CAL. RACIAL & IDENTITY PROFILING BD. (2020), <https://perma.cc/RA4T-JQ3Y> (providing a comprehensive analysis of evidence of racial bias in policing behavior in California).

132. *See, e.g., supra* note 126 and accompanying text; R.I. AFFILIATE ACLU, *THE PERSISTENCE OF RACIAL PROFILING IN RHODE ISLAND: A CALL FOR ACTION 3-6* (2007), <https://perma.cc/HT4T-TJ7R> (presenting evidence of possible racial bias by Rhode Island police); ACLU & ACLU OF FLA., *RACIAL DISPARITIES IN FLORIDA SAFETY BELT LAW ENFORCEMENT 1-3* (2016), <https://perma.cc/2FD2-VH6V> (finding evidence of racial disparities in seat-belt enforcement in Florida).

133. *See, e.g., supra* notes 125-29 and accompanying text.

134. *See, e.g., infra* notes 135-42 and accompanying text.

In doing so, many studies have struggled with a common methodological limitation: the so-called “benchmark” problem.¹³⁵ Lorie Fridell, in a report authored in conjunction with the Police Executive Research Forum and funded by the U.S. Department of Justice (DOJ) Office of Community Oriented Policing Services, described “the benchmarking challenge” as follows:

Jurisdictions collecting police-citizen contact data are calling upon social science to determine whether there is a cause-and-effect relationship between a driver’s race/ethnicity and vehicle stopping behavior by police. In analyzing the data, researchers have attempted to develop comparison groups to produce a “benchmark” against which to measure their stop data. If an agency determines that, say, 25 percent of its vehicle stops are of racial/ethnic minorities, to what should this be compared? In other words, what percentage would indicate racially biased policing? This is the question at the core of benchmarking.¹³⁶

Another example may best illustrate this problem. In *Floyd v. City of New York*, a group of plaintiffs argued that the NYPD was engaged in a pattern of stops and frisks that failed to meet the constitutional standards articulated in *Terry v. Ohio*.¹³⁷ To demonstrate a pattern of unconstitutional misconduct, the plaintiffs pointed to the vast overrepresentation of Black and Latino young men among the population of those subjected to *Terry* stops relative to the city’s overall population.¹³⁸ But the NYPD denied wrongdoing and argued that Black and Latino young men would be subject to a disproportionate number of *Terry* stops even without racial discrimination. It claimed that there were higher rates of criminal activity among those racial groups, and that the NYPD had made a tactical choice to allocate more officers to higher-crime

135. Jeffrey Grogger & Greg Ridgeway, *Testing for Racial Profiling in Traffic Stops from Behind a Veil of Darkness*, 101 J. AM. STAT. ASS’N 878, 878 (2006) (“[T]he key empirical problem in testing for racial profiling [is] measuring the risk set, or the ‘benchmark,’ against which to compare the racial distribution of traffic stops.”). David A. Harris provides a detailed and careful examination of how prior researchers have dealt with the benchmark issue. He notes that many early researchers simply used the “easiest,” the “most widely available,” or the “cheapest” benchmark data available—often census information. David A. Harris, *U.S. Experiences with Racial and Ethnic Profiling: History, Current Issues, and the Future*, 14 CRITICAL CRIMINOLOGY 213, 230 (2006). But these benchmarks fail to consider potential differences in the underlying behavior of the populations studied. Census data may also be problematic for studies of racial profiling in traffic stops because the racial breakdown of a community may not match the racial breakdown of those driving within that community. *See id.* at 213, 229-33 (describing the benchmarking issue and providing a detailed assessment of the literature on this issue).

136. LORIE FRIDELL, *BY THE NUMBERS: A GUIDE FOR ANALYZING RACE DATA FROM VEHICLE STOPS* 7 (2004), <https://perma.cc/WZ8A-2KK2>.

137. *Floyd v. City of New York*, 959 F. Supp. 2d 540, 556, 558-59 (S.D.N.Y. 2013) (describing the circumstances that led to the litigation); *Terry v. Ohio*, 392 U.S. 1 (1968).

138. *Floyd*, 959 F. Supp. at 583-89 (describing the “competing benchmarks” used by each side during the litigation).

communities where more Blacks and Latinos lived relative to other racial groups.¹³⁹ So in deciding whether Blacks and Latinos were overrepresented among those targeted for stops and frisks, the NYPD asserted that the right comparison point was not the proportion of New York City's population that Black and Latino young men composed, but rather the rates at which various races appeared in suspect descriptions from crime victims.

Identifying the appropriate "benchmark" for comparison is critical in evaluating whether the resulting statistical disparities in police behavior are the result of racial profiling by law-enforcement officers or the result of genuine differences in underlying behavior.¹⁴⁰ This benchmark problem has complicated efforts by litigants, including the DOJ, to make legally sufficient showings of racial bias in court proceedings against local law-enforcement agencies.¹⁴¹ As one researcher known for her skepticism of racial-profiling studies critically remarked on traffic enforcement, "[u]ntil someone devises an adequately sophisticated benchmark that takes into account population patterns on the roads, degrees of law breaking, police deployment patterns, and the nuances of police decision making, stop data are as meaningless as they are politically explosive."¹⁴² To address these benchmarking challenges, traffic studies have taken a number of different methodological approaches.

Some studies have simply compared the rate at which police stop and search the vehicles of drivers of color with the underlying population of a

139. *Id.* at 584 ("The City's experts, by contrast, used a benchmark consisting of the rates at which various races appear in suspect descriptions from crime victims—in other words, 'suspect race description data.' The City's experts assumed that if officers' stop decisions were racially unbiased, then the racial distribution of stopped pedestrians would be the same as the racial distribution of the criminal suspects in the area." (footnote omitted) (quoting the defendant's expert report)).

140. For a broader discussion of the benchmark problem in the context of traffic stops, see Grogger & Ridgeway, *supra* note 135, at 878 (concluding that benchmarks based on residential population may result in a poor or imprecise estimate of the population of drivers on the road violating traffic laws).

141. The benchmark problem has been problematic in cases where the DOJ has attempted to prove racial bias in traffic stops under 34 U.S.C. § 12601, which allows the Attorney General to sue local law-enforcement agencies for engaging in a pattern or practice of unlawful or unconstitutional conduct. *See, e.g.,* United States v. Johnson, 122 F. Supp. 3d 272, 331-38 (M.D.N.C. 2015) (describing why the district court ultimately concluded that Dr. John Lamberth's use of an observational study to establish a benchmark for racial-profiling analysis was inadmissible under *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993), and, regardless, lacked credibility in the eyes of the court).

142. Brian L. Withrow & Howard Williams, *Proposing a Benchmark Based on Vehicle Collision Data in Racial Profiling Research*, 40 CRIM. JUST. REV. 449, 451 (2015) (quoting HEATHER MAC DONALD, ARE COPS RACIST? HOW THE WAR AGAINST THE POLICE HARMS BLACK AMERICANS 22 (2003)).

geographical area¹⁴³ or with the racial distribution of licensed drivers.¹⁴⁴ Other studies have compared the rate at which police stop and search drivers of color with the rate at which those same drivers appear to violate traffic laws, relying on systematic field observations or self-reported surveys to establish a benchmark.¹⁴⁵ Still others have developed a benchmark by relying on vehicle-

-
143. See, e.g., Michael R. Smith & Matthew Petrocelli, *Racial Profiling? A Multivariate Analysis of Police Traffic Stop Data*, 4 POLICE Q. 4, 8-17 (2001) (comparing the number of Black individuals stopped in Richmond, Virginia, with their percentage of the driving-eligible population and finding apparent evidence of racial profiling); FRIDELL, *supra* note 136, at 75-113 (describing the benchmark challenge and identifying “adjusted census data” as one possible way to establish a benchmark).
144. William R. Smith, Donald Tomaskovic-Devey, Matthew T. Zingraff, H. Marcinda Mason, Patricia Y. Warren, Cynthia Pfaff Wright, Harvey McMurray & C. Robert Fenlon, *The North Carolina Highway Traffic Study 6* (U.S. Dep’t of Just., Working Paper No. 204021, 2004), <https://perma.cc/ALS9-H8F5> (“Somewhat to our surprise, we found empirical evidence to the effect that there is also racial variation by time of day in the distribution of drivers on the highways of North Carolina. African Americans are more likely to be driving in the evening and early morning hours *relative to their distribution in the licensed driver population.*” (emphasis added)).
145. For example, John Lamberth has conducted numerous studies using this methodology in jurisdictions like Maryland, New Jersey, and Alamance County, North Carolina. See, e.g., *Johnson*, 122 F. Supp. 3d at 304-07; Report of John Lamberth, Ph.D. at 2-4, *Wilkins v. Md. State Police*, No. 93-cv-00468 (D. Md. Nov. 8, 1996); John Lamberth, *Driving While Black*, WASH. POST (Aug. 16, 1998), <https://perma.cc/9BSF-36N4> (describing Lambert’s use of this methodology for benchmarking in litigation against the New Jersey State Police). A number of other researchers have similarly used field-observation benchmarks to demonstrate a pattern of racial bias by police officers in traffic stops, commonly finding evidence of racial profiling. For an example of a study that has relied on Lamberth’s benchmarking process, see Samuel R. Gross & Katherine Y. Barnes, *Road Work: Racial Profiling and Drug Interdiction on the Highway*, 101 MICH. L. REV. 651, 664 (2002). In that study, Gross and Barnes examined Maryland State Police data from stops and searches conducted on Interstate 95 between 1995 and the mid-2000s. *Id.* at 658. At the time of their study, Gross and Barnes claimed that Maryland State Police were the only major agency to make this kind of data publicly available for review. *Id.* (“There is only one American jurisdiction for which detailed data on racial profiling in highway searches are available for a considerable period. Since January 1995, Maryland State Police . . . troopers have been under court order to file a report on every incident in which they stop and search a motor vehicle, including information on the race of the driver, the basis for the search, and the type and quantity of the drugs recovered, if any.”). By analyzing a dataset of 2,146 searches that occurred on a portion of Interstate 95 from Baltimore to the Delaware border, Gross and Barnes found that the Maryland State Police engaged in racial profiling by stopping and searching cars with Black and Hispanic drivers more often than cars with white drivers. *Id.* at 658-60; see also Patricia Warren, Donald Tomaskovic-Devey, William Smith, Matthew Zingraff & Marcinda Mason, *Driving While Black: Bias Processes and Racial Disparity in Police Stops*, 44 CRIMINOLOGY 709, 712 (2006) (noting that “[s]urveys offer an alternative to official records,” as they “allow for data that are unlikely to be available in official reports to be collected and analyzed,” and describing how the authors use “self-reported driving behavior and past vehicle stops to statistically control for law-breaking behavior”).

collision data, arguing that this collision data provide better insight into the racial breakdown of drivers on the road than population data alone.¹⁴⁶ Another group of studies has attempted to circumvent the benchmark problem not by focusing on the comparative rate of stops and searches of drivers of color relative to a benchmark, but instead by comparing differential rates at which police search the vehicles of drivers of color and the rate at which these searches result in the collection of contraband.¹⁴⁷ And an emerging set of studies has adopted a “veil-of-darkness” methodology that compares the rate at which police stop white and nonwhite drivers at day and at night, under the assumption that evidence of racial profiling will be most evident during daylight hours when police can more easily ascertain the race of nearby drivers.¹⁴⁸

-
146. See, e.g., Geoffrey P. Alpert, Michael R. Smith & Roger G. Dunham, *Toward a Better Benchmark: Assessing the Utility of Not-at-Fault Traffic Crash Data in Racial Profiling Research*, JUST. RSCH. & POL'Y, Spring 2004, at 43, 56-63 (arguing that traffic-crash data can be a useful tool for establishing a benchmark and using Miami-Dade County, Florida, as an example for implementing this methodology); Withrow & Williams, *supra* note 142, at 464 (arguing that this type of benchmark is more reliable than alternatives).
147. See generally George E. Higgins, Gennaro F. Vito & William F. Walsh, *Searches: An Understudied Area of Racial Profiling*, 6 J. ETHNICITY CRIM. JUST. 23, 28-29, 37 (2008) (using data from more than 36,000 traffic stops in Louisville, Kentucky, to show that race appeared to influence the likelihood of searches taking place after a traffic stop). As another example, Frank R. Baumgartner and others have shown through an analysis of 18 million traffic stops in North Carolina between 2002 and 2013 that Black drivers, particularly younger Black men, are disproportionately likely to be searched and arrested incident to a traffic stop. See Frank R. Baumgartner, Derek A. Epp, Kelsey Shoub & Bayard Love, *Targeting Young Men of Color for Search and Arrest During Traffic Stops: Evidence from North Carolina, 2002-2013*, 5 POL. GRPS. & IDENTITIES 107, 112, 117-19 (2017) [hereinafter Baumgartner et al., *Young Men of Color*] (conducting the analysis described and discussing the hit rate for these searches). Baumgartner and others have found similar patterns in data from the Texas Department of Public Safety from 2003 through 2014. FRANK R. BAUMGARTNER, BRYAN D. JONES, JULIO ZACONET, COLIN WILSON & ARVIND KRISHNAMURTHY, *RACIAL DISPARITIES IN TEXAS DEPARTMENT OF PUBLIC SAFETY TRAFFIC STOPS, 2002-2014*, at 1-2 (2015), <https://perma.cc/E55C-NV7R>.
148. Grogger & Ridgeway, *supra* note 135, at 878-79 (using a veil-of-darkness methodology to examine whether the Oakland Police Department engaged “in racial profiling when selecting particular vehicles to stop”). For more discussions and examples of the veil-of-darkness methodology, see Joseph A. Ritter & David Bael, *Detecting Racial Profiling in Minneapolis Traffic Stops: A New Approach*, CURA REP., Spring/Summer 2009, at 11, 11 (using a veil-of-darkness methodology to analyze racial profiling in Minneapolis); Robert E. Worden, Sarah J. McLean & Andrew P. Wheeler, *Testing for Racial Profiling with the Veil-of-Darkness Method*, 15 POLICE Q. 92, 93, 105 (2012) (using the veil-of-darkness methodology to find no evidence of racial bias in Syracuse, New York); and William C. Horrace & Shawn M. Rohlin, *How Dark Is Dark? Bright Lights, Big City, Racial Profiling*, 98 REV. ECON. & STAT. 226, 227, 231 (2016) (redefining the parameters for a veil-of-darkness analysis in Syracuse, New York, and finding evidence that Black drivers were being stopped 15% more during daylight compared to darkness hours). It is also worth noting that police administrative records can conceal the presence of
- footnote continued on next page*

Admittedly, this brief survey only scratches the surface of the evidence of racial profiling that surfaced in the years before and after the *Whren* decision.¹⁴⁹ Even so, a couple of important lessons emerge. First, scholars using a wide variety of methodological techniques have found evidence of racial profiling in police agencies across the country. This is not to say that all police departments are equal, or that all departments demonstrate troubling patterns of racial bias. There are around 18,000 police departments in the United States, each with its own unique policies, procedures, and culture.¹⁵⁰ The nation is not policed by one police department, but by thousands of decentralized agencies.¹⁵¹ Nevertheless, the existing literature suggests that racial profiling by police is a relatively common occurrence across a number of American police departments.

Second, while these studies frequently cite *Whren* as one of the causal mechanisms that may be contributing to the prevalence of racial profiling by law enforcement, none of these studies have causally connected *Whren* or similar cases to the patterns they observe. Take, for example, the comprehensive study of North Carolina traffic-stop data conducted by Frank Baumgartner, Derek Epp, Kelsey Shoub, and Bayard Love in 2016.¹⁵² In their discussion of the potential root causes of racial profiling, they argue:

[T]he Supreme Court decided in *Whren v. United States* (1996) that any traffic violation was a legitimate reason to stop a driver, even if the purported violation (e.g. changing lanes without signaling) was clearly a pretext for the officer's desire to stop and search the vehicle for other reasons, such as a general suspicion. There was no requirement that speeding laws, for example, be equitably enforced; if all drivers are speeding, it is constitutionally permissible, said the Justices, to pick out just the minority drivers and enforce the speeding laws selectively. Of course, once a car is stopped, officers are able to conduct a "consent" search when drivers do not object to the officer's request to search the vehicle. The *Whren* decision opened the floodgates to pretextual stops. Thus, tens of thousands of black and

underlying racial bias by law enforcement. See generally Dean Knox, Will Lowe & Jonathan Mummolo, *Administrative Records Mask Racially Biased Policing*, 114 AM. POL. SCI. REV. 619 (2020) (arguing that administrative records of police behavior can often mask the presence of racial bias and outlining alternative methods for addressing this problem).

149. For a list of references to prior research on this topic, see FRIDELL, *supra* note 136, at 423-37.

150. See BRIAN A. REAVES, U.S. DEP'T OF JUST., NCJ 233982, CENSUS OF STATE AND LOCAL LAW ENFORCEMENT AGENCIES, 2008, at 2 (2011), <https://perma.cc/PG66-3J5E> (reporting that there were nearly 18,000 state and local law-enforcement agencies in 2008).

151. Stephen Rushin, *Using Data to Reduce Police Violence*, 57 B.C. L. REV. 117, 139-42 (2016).

152. Baumgartner et al., *Young Men of Color*, *supra* note 147, at 108.

brown drivers have routinely been stopped and searched in an effort to reduce drug use.¹⁵³

While this seems like an intuitive conclusion, existing studies have struggled to prove it empirically. As we explain in the next Part, we believe that our study begins to fill this gap in the existing literature.

III. The Effects of *Arreola* on Police Behavior

In the wake of *Whren*, scholars expressed widespread concern that by green-lighting pretextual traffic stops, the Supreme Court had inadvertently facilitated racial profiling.¹⁵⁴ Demonstrating this proposition quantitatively has proven difficult, however, for two primary reasons. First—at least according to our examination of the existing literature and case law—there is a lack of widespread, historical variation in pretextual stop policies across jurisdictions. Thus, it is not clear that the Supreme Court’s decision in *Whren* acted as an exogenous shock. Instead, it may have actually validated practices already common among many police departments. This means that even if data were widely available on police stops, it is not clear that we could measure the effect of *Whren* on racial profiling, as it may not have actually changed many departments’ policing practices.

Second, empirically evaluating the effects of pretextual-stop doctrines has proven challenging because of a lack of comprehensive data on police behavior. Only recently have some states required police departments to keep data on traffic and pedestrian stops, including the race of those individuals stopped.¹⁵⁵ Such laws remain relatively rare today.¹⁵⁶ When the Court handed

153. *Id.* Although Baumgartner et al. cited *Whren* as a case that may contribute to racial profiling, they did not establish a causal link between pretextual stops and racial profiling.

154. *See supra* Part I.B.

155. *See, e.g.,* Matt Kiefer, *Police in Illinois Will Permanently Have to Record Race, Other Traffic Stop Data in New Bill*, CHI. REP. (May 23, 2019), <https://perma.cc/P9AQ-55UB> (“The Illinois Traffic Stop Study—launched 15 years ago under legislation sponsored by then-State Sen. Barack Obama—requires police officers to record key data points during every traffic stop, including the reason for the stop, the race of the driver and outcomes that may include warnings, tickets and searches.”); John Sides, *What Data on 20 Million Traffic Stops Can Tell Us About “Driving While Black,”* CHARLOTTE OBSERVER (July 17, 2018, 2:55 PM), <https://perma.cc/VS7H-4WPX> (to locate, click “View the live page”) (“North Carolina became the first state to mandate the collection of traffic stops [sic] data in 1999, thanks in large part to efforts by black representatives in the state legislature.”).

156. ACLU of Ill., *Vote Yes on HB 1613*, at 1 (2019), <https://perma.cc/XRM9-Q89L> (“Of the 15 states with data collection laws, 13 are permanent. IL and MD are the only 2 with temporary laws.”).

down *Whren* in 1996, such data was scarcely available,¹⁵⁷ and the federal government has never kept national data on police traffic stops or the race of those stopped by law enforcement.¹⁵⁸ Combined, this lack of data and the lack of jurisdictional variation have meant that scholars can only hypothesize about the potentially harmful effects of *Whren* on racial minorities.

As described in Part I, at least two states—New Mexico¹⁵⁹ and Washington¹⁶⁰—have acted to limit police use of pretextual stops. Of these two states, only Washington—with its progression from *Ladson* to *Arreola*—has both the doctrinal variation and available data necessary to test the effect of pretextual-stop doctrines on the racial disparities in traffic stops by police officers. The next Subpart lays out our research model.

A. Study Design

Table 1 illustrates the legality of pretextual stops and other similar forms of traffic stops in Washington over time.

Table 1
Legality of Pretextual Stops in Washington over Time (1999-Present)

1999-2012	2012-present	2013
Pretextual stops unconstitutional under <i>Ladson</i>	Mixed-motive stops permissible under <i>Arreola</i>	Officers trained in the use of mixed-motive stops

As Table 1 shows, the Washington Supreme Court banned the use of pretextual stops between 1999 and 2012. Then, in December 2012, the Washington Supreme Court authorized a form of mixed-motive stops that closely resembles pretextual stops.¹⁶¹ However, we are primarily interested not in the date that the Washington Supreme Court handed down *Arreola*, but in the dates on which officers in Washington received training in the use of mixed-motive stops.

157. *Sides*, *supra* note 155 (quoting Baumgartner and his coauthors for the proposition that North Carolina was the first state with a comprehensive data-collection law passed in 1999—three years after *Whren*).

158. *Rushin*, *supra* note 151, at 117-18 (describing how the federal government keeps very few statistics on police behavior, including major subjects like the number of individuals killed by law enforcement each year).

159. *See Sievers*, *supra* note 78, at 595-601 (describing the New Mexico departure from *Whren*).

160. *See supra* Parts I.C.1-2 and accompanying text.

161. *State v. Arreola*, 290 P.3d 983, 991 (Wash. 2012) (en banc).

To gain some insight into these matters, we turn to data from the Washington State Criminal Justice Training Commission (WSCJTC), an organization “created in 1974 to establish standards and provide training to criminal justice professionals, including peace officers.”¹⁶² Washington is unique among most American states in that officers across the state receive consistent training via the WSCJTC.¹⁶³ Additionally, under Washington law, all law-enforcement officers in the State of Washington must undergo twenty-four hours of mandatory in-service training each year.¹⁶⁴ The WSCJTC provides information to law-enforcement officers and agencies on new court rulings through the publication of monthly law-enforcement digests.¹⁶⁵ And each year, the WSCJTC updates a comprehensive Law Enforcement Legal Update Outline, which summarizes all case law “on arrest, search, seizure, and other topical areas of interest to law-enforcement officers; plus a chronology of independent grounds rulings under Article I, Section 7 of the Washington Constitution.”¹⁶⁶

In the years between the *Ladson* and the *Arreola* decisions, these training materials advised officers about the ban on pretextual stops.¹⁶⁷ Even in cases from other states in which litigants had not directly challenged stops as pretextual—like the case in which a gang unit executed a traffic stop and asked about the driver’s possible gang affiliation—the Commission went out of its way to warn officers that “there is [a] substantial chance that Washington courts would find [the stop] to be pretextual under the Washington

162. *About the Commission: Training the Guardians of Democracy*, WASH. ST. CRIM. JUST. TRAINING COMM’N, <https://perma.cc/98MG-H2KJ> (archived Jan. 6, 2021).

163. *Id.*

164. *Training & Education: Mandated 24-Hour In-Service Audit*, WASH. ST. CRIM. JUST. TRAINING COMM’N, <https://perma.cc/XRX5-8339> (archived Jan. 6, 2021) (citing the statutory requirements established by Washington law). The in-service training requirement went into effect in 2006, and thus covers our primary time period of interest (the period immediately after *Arreola*). *Id.*

165. *Resources: Law Enforcement Digests*, WASH. ST. CRIM. JUST. TRAINING COMM’N, <https://perma.cc/DRK2-VBJK> (archived Jan. 6, 2021).

166. *Id.* (noting that the most recent version of this annual document went live on July 1, 2020, and stating that the document is updated “at least once each year”).

167. See JOHN R. WASBERG, WASH. STATE CRIM. JUST. TRAINING COMM’N, LAW ENFORCEMENT LEGAL UPDATE OUTLINE 10-11 (2020), <https://perma.cc/K5R9-WVPR> (noting that the Law Enforcement Legal Update Outline first incorporated a warning about the *Ladson* holding in September 1999 and framing *Ladson* as a prohibition on pretextual stops); WASH. STATE CRIM. JUST. TRAINING COMM’N, LAW ENFORCEMENT DIGEST: AUGUST 2012, at 7 (2012), <https://perma.cc/WJ4E-BDRH> (citing *Ladson* and explaining that if a federal case had been “tried under the Washington constitution, article I, section 7, the pretext stop . . . would have tainted the evidence obtained in subsequent police action at the scene of the stop”).

constitution.”¹⁶⁸ After the Washington Supreme Court decided *Arreola*, discussion of the case first appeared in the monthly law-enforcement digest approximately three months later, in March 2013.¹⁶⁹ It then presumably appeared in the annual Law Enforcement Legal Update Outline later that year in July 2013 (and it remains in the most recent version of this document, published in July 2020).¹⁷⁰

Admittedly, we cannot precisely pinpoint the moment when all or most troopers in the Washington State Patrol became familiar with the mixed-motive-stop doctrine established by *Arreola*. The existing evidence on training leads us to hypothesize that *Arreola* may have had some immediate effect upon the behavior of officers aware of the decision in December 2012. But we think it is possible that the full effect of *Arreola* may not have been felt until as long as a full year or more after the decision—that is, after the WSCJTC and the Washington State Patrol had fully incorporated the decision into their training materials, and after all officers had presumably completed their statutorily required twenty-four hours of annual in-service training for 2013. As discussed in later Subparts, we employ our models under various assumptions related to the time at which officers learned how to employ mixed-motive stops.

B. Dataset

The Stanford Open Policing Project has made available online extensive amounts of data on police traffic stops (among other datasets) from departments across the country.¹⁷¹ We draw on data provided by this database for the Washington State Patrol.¹⁷² The Washington State Patrol is the

168. WASH. STATE CRIM. JUST. TRAINING COMM’N, LAW ENFORCEMENT DIGEST: MARCH 2009, at 3-4 (2009), <https://perma.cc/U249-6QVS>.

169. See WASBERG, *supra* note 167, at 11. We ascertained the date of *Arreola*’s first appearance because the Law Enforcement Legal Update Outline operates as a historical document that notes the month and year in which an opinion was added. In the section on pretextual stops, the outline notes the holding of *Arreola* and states that it first appeared in the March 2013 issue of the law-enforcement digest. *Id.* at i, 11, 63.

170. *Id.* at i, 11, 63 (noting that the 2019 and 2020 versions were published on July 1 of those years).

171. Data, STAN. OPEN POLICING PROJECT, <https://perma.cc/86JV-NLFU> (archived Jan. 6, 2021) (providing datasets for Washington, including for the Washington State Patrol, Seattle, and Tacoma); see also Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff & Sharad Goel, *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE: HUM. BEHAV. 736, 736-37 (2020) (presenting the analysis that resulted from data collection by Stanford Open Policing Project researchers).

172. Data, *supra* note 171.

primary state policing agency for Washington and employs around 1,100 troopers whose primary responsibilities include “provid[ing] a safe motoring environment for all Washingtonians” on the “17,524 miles of the state’s highways” as well as interstates.¹⁷³

The dataset we examine includes 8,257,527 stops made by troopers of the Washington State Patrol from December 2008 through December 2015.¹⁷⁴ It includes data on the date, time, and location of each stop.¹⁷⁵ It also includes data on the race, age, and sex of each driver, as well as data on whether the officer conducted a search after the stop, whether this search resulted in the collection of any contraband, whether the officer issued a citation, whether the officer issued a warning, and whether the officer performed a frisk of any suspect.¹⁷⁶ The availability of data from multiple years before and after December 2012 allows us to examine the effects of the *Arreola* decision.

The Washington State Patrol reports the race of drivers using five different racial classifications: White, Black, Hispanic, Asian/Pacific Islander, and Other. If an officer fails to select one of these racial identifiers, then the database categorizes the race of the driver as unavailable, or “n/a.”¹⁷⁷ This raises an immediate methodological challenge: How should we handle cases where officers fail to list the race of the driver? This is a particularly challenging question because an officer’s decision to not report the race of a driver during a traffic stop may not be random. It may even be done in an effort to avoid the detection of a pattern of stopping drivers of color.¹⁷⁸ Media reports¹⁷⁹ and prior scholarly examinations¹⁸⁰ of the Washington State Patrol

173. *About Us*, *supra* note 22; *Driver*, WASH. ST. PATROL, <https://perma.cc/9LF4-S9K8> (archived Jan. 6, 2021).

174. *Data*, *supra* note 171.

175. *Id.*

176. *Id.* We would have preferred to have used data from Seattle and Tacoma as well, since these agencies also provided datasets to the Stanford Open Policing Project. But we were unable to use data from these two agencies, as they did not include information on the driver’s race or whether searches were conducted. *Id.*

177. Around 26% of stops by the Washington State Patrol fail to list the race of the driver.

178. *See generally* Elizabeth Luh, *Not So Black and White: Uncovering Racial Bias from Systematically Misreported Trooper Reports* (Apr. 30, 2020) (unpublished manuscript), <https://perma.cc/9WX2-5ZLB> (finding compelling evidence that officers in Texas systematically misreported race to avoid bias detection).

179. *See, e.g.*, Jason Buch & Joy Borkholder, *Native American Drivers Are More Likely to Be Searched by Washington State Patrol*, CROSSCUT (Dec. 19, 2019), <https://perma.cc/RD3Z-W8RF> (“Some troopers were disciplined for refusing to accurately record data, according to a 2015 training presentation to State Patrol cadets . . .”).

180. NICHOLAS P. LOVRICH, MICHAEL J. GAFFNEY, CLAYTON MOSHER, MITCHELL PICKERILL & TRAVIS C. PRATT, WASH. STATE UNIV., *ANALYSIS OF TRAFFIC STOP DATA COLLECTED BY THE WASHINGTON STATE PATROL: ASSESSMENT OF RACIAL AND ETHNIC EQUITY AND BIAS IN STOPS, CITATIONS, AND SEARCHES USING MULTIVARIATE QUANTITATIVE AND MULTI-*
footnote continued on next page

have noted that officers have faced discipline and accusations of wrongdoing for failing to properly document the race of drivers. Based on the available research on intentional racial misidentification of drivers by police officers during traffic stops, we believe it is important to account for the very real possibility that officers may use these “n/a” cases to systematically conceal stops of drivers of color. Thus, as we explain in more detail in the next Subpart, we chose to run all of our analyses both with and without the inclusion of these “n/a” cases in our definition of nonwhite drivers.¹⁸¹

Both approaches produced substantially similar results. Regardless of whether we include “n/a” cases in or exclude “n/a” cases from our definition of nonwhite drivers, we still find that the introduction of *Arreola* is associated with a statistically significant increase in traffic stops of these drivers. We also find *Arreola* to be associated with statistically significant increases in stops of Black drivers, Hispanic drivers, and drivers of “other” races relative to white drivers, even when we exclude both “n/a” cases and all other nonwhite races from our analysis. Put simply, whether we aggregate all nonwhite groups or we analyze them separately, our results do not change much.

In analyzing this dataset, we faced another methodological challenge: Should we look at the raw number of stops of each racial group, or should we attempt to convert these raw stop numbers into rates of stops per capita? We made the purposeful choice *not* to convert the dataset from raw stop numbers into rates of stops per capita. If we converted these numbers into per capita rates, we would need to know two pieces of information: (1) the number of stops of each racial subgroup by county (or some other geographical or jurisdictional area); and (2) the number of drivers of each racial group on roads policed by the Washington State Patrol in that county (or geographical or jurisdictional area). While we have self-reported data from the Washington State Patrol on (1), we lack data for (2). As previously addressed in Part II,

METHOD QUALITATIVE RESEARCH TECHNIQUES 5 (2005), <https://perma.cc/G2TY-4CGY> (explaining the authors’ decision to conduct an analysis of stop records by comparing records to photographs from the Department of Licensing offices, and acknowledging that the authors’ decision to conduct this analysis was in part motivated by media reports that some troopers may be “systematically” misclassifying the race of minority drivers).

181. Stated another way, we first ran all of our models throughout the paper with the “n/a” cases removed entirely from the dataset. This approach focuses our analysis exclusively on cases where state troopers selected one of the racial identifiers. While this sidesteps the “n/a” problem, we also recognize that it may fail to consider a number of stops of drivers of color. Because officers may attempt to use the “n/a” classification to conceal unusually frequent stops of drivers of color, we alternatively characterized “n/a” cases as being drivers of color and reran all of our analyses. Both methodologies produced largely similar, statistically significant results. Where the inclusion or exclusion of “n/a” cases affected the results, we explain these differences. Alternative specifications of our models are available in Parts B and C of the Appendix.

scholars have widely debated the best way to address this so-called benchmark or “denominator” problem.¹⁸²

One way to address the benchmark problem would be to rely on U.S. Census population estimates by county by race by month, under the assumption that residents of a given county are a close proxy for the drivers on state and interstate highways in that county policed by the Washington State Patrol. But this assumption has been proven false in multiple prior studies, as documented in detail by Geoffrey Alpert, Michael Smith, and Roger Dunham.¹⁸³ In one such study, Joel Miller, Paul Quinton, and Nick Bland used mounted video cameras to document the ethnicity of drivers in cities in England. They found that the population and demographic profile of the drivers in those cities often varied substantially from the census-level residential population.¹⁸⁴ Howard Greenwald reached a substantially similar conclusion in his analysis of Sacramento, California, in 2001, finding that in some areas “minority drivers, observed as a percentage of total observations, far exceeded their proportions in the corresponding census population,” while in other areas, “the reverse was true and minorities were significantly underrepresented relative to the census.”¹⁸⁵ In Denver between 2001 and 2002, Deborah Thomas found that only around half of all traffic stops by the Denver Police Department involved residents of Denver, meaning that using census data to calculate stop rates would have been “wildly inaccurate.”¹⁸⁶ Similar disparities existed when Alpert, Smith, and Dunham compared the observed race of drivers at various intersections in Miami-Dade County, with the smallest block and tract census data.¹⁸⁷ And in Plano, Texas, a large suburb of Dallas, researchers found that 79% of all traffic stops within city limits

182. See *supra* Part II; see also Samuel Walker, *Searching for the Denominator: Problems with Police Traffic Stop Data and an Early Warning System Solution*, JUST. RSCH. & POL’Y, Spring 2001, at 63, 71-73 (discussing how it can be difficult for social-science researchers to establish a proper baseline in racial-profiling cases).

183. Alpert et al., *supra* note 146, at 45.

184. Joel Miller, Paul Quinton & Nick Bland, *Measuring Stops and Searches: Lessons from U.K. Home Office Research*, JUST. RSCH. & POL’Y, Fall 2002, at 143, 151-52 (“The research shows that measures of available populations are very different from resident populations (as measured by the 1991 census).”).

185. Alpert et al., *supra* note 146, at 45; see also HOWARD P. GREENWALD, FINAL REPORT: POLICE VEHICLE STOPS IN SACRAMENTO, CALIFORNIA 38-39 (2001), <https://perma.cc/XS7F-6BFV> (comparing census population figures with the observed races of passing drivers, finding that there is a significant disparity between these measures, and further finding that around one-third of all drivers on Sacramento roads are nonresidents). Because of this, the study urges future researchers to act with “extreme caution” before using census data as a denominator in calculating stop rates.
Id.

186. Alpert et al., *supra* note 146, at 45.

187. *Id.*

involved drivers who were not residents of the city, suggesting the population demographics of drivers may differ from the demographics of the residential population.¹⁸⁸

This mismatch between residential population and the population of drivers on roadways may be even more substantial in the case of state and interstate highways that run through smaller, rural counties to connect major population centers, tourist attractions, national or state parks, and major employers. Take, for example, the highly trafficked, 175-mile corridor of Interstate 5 between Seattle, Washington, and Portland, Oregon, which also includes other major population hubs like Olympia and Tacoma.¹⁸⁹ This corridor is so highly trafficked by commuters, businesspeople, and tourists that lawmakers are examining the introduction of a high-speed railway to accommodate up to 3 million trips per year by 2040.¹⁹⁰ If and until that happens, many of these travelers navigate Interstate 5, which takes them through six different counties: King, Pierce, Thurston, Lewis, Cowlitz, and Clark.¹⁹¹ Given the frequency of travel between Seattle, Tacoma, Olympia, and Portland along Interstate 5, there is a high probability that individuals stopped by the Washington State Patrol along Interstate 5 in rural counties like Lewis or Cowlitz are actually residents of larger urban counties, like King or Pierce, or are out-of-state drivers. And given that counties like King have drastically different demographic profiles from counties like Lewis and Cowlitz, use of county population as a proxy for driver population may skew the stop data and future analysis in an unpredictable and unreliable way.¹⁹²

188. PLANO POLICE DEP'T, 2019 PLANO POLICE DEPARTMENT RACIAL PROFILING REPORT 5-6 (2019), <https://perma.cc/YTB2-YXVF> (noting this figure and then developing an estimate of the demographic profile of drivers, which differs from the demographic profile of the city's residential population).

189. Gregory Scruggs, *The Case for Portland-to-Vancouver High-Speed Rail*, BLOOMBERG CITYLAB (Dec. 4, 2019, 7:15 AM PST), <https://perma.cc/FRD6-D43E> (identifying the significance of the Interstate 5 corridor that includes Portland and Seattle, and stating that "[o]nly 175 miles separate Portland from Seattle"). For a detailed feasibility study outlining the importance of this corridor and the kinds of individuals who travel along Interstate 5, see generally WASH. STATE DEP'T OF TRANSP., *ULTRA-HIGH-SPEED GROUND TRANSPORTATION BUSINESS CASE ANALYSIS: FINAL REPORT* (2019), <https://perma.cc/B4L5-KQAQ>.

190. Scruggs, *supra* note 189 (citing the 3 million figure).

191. WASH. STATE DEP'T OF TRANSP., *supra* note 189, app. C at 4 (listing the counties along this corridor and including the counties listed above); see also *Washington Road Map*, MAPPERY, <https://perma.cc/E6H7-XGYB> (archived Feb. 22, 2021).

192. In fact, a study conducted by the Washington State University faculty recognized this fact in a report issued in 2005. In explaining why census resident-population estimates were likely an unreliable benchmark in calculating rates of traffic stops by race, they explained, "[i]t is also important to note that certain areas of the state (particularly the Interstate-5 corridor running from the Canadian border to the Oregon border) patrolled by the WSP have a high proportion of out-of-state drivers, and it is probable

footnote continued on next page

Alternatively, inclusion of state population numbers to generate stop rates fails to capture the fact that individuals of color are often less likely to have driver's licenses,¹⁹³ and a substantial number of the travelers on state and interstate highways are out-of-state residents.¹⁹⁴ The use of state population numbers also would not account for the fact that the Washington State Patrol only patrols a small fraction of the state's public roadways.¹⁹⁵ The driver population on the particular segment of roadways policed by the Washington State Patrol may systematically differ from the state's driver-age population.

Because of these issues, the National Institute of Justice has bluntly concluded that "social scientists now disregard comparisons to the census for assessing racial bias."¹⁹⁶ Alpert, Smith, and Dunham have argued that "[e]vidence is mounting . . . that the census population of an area under study does not accurately represent the driving population available to be stopped," and thus "the use of census data largely ha[s] been discredited."¹⁹⁷ And Greg Ridgeway and John MacDonald have explained that "[i]t is quite conceivable that the residential population in many neighborhoods has little resemblance

that these drivers are more likely to be members of racial minority groups than resident in-state drivers." LOVRICH ET AL., *supra* note 180, at 13.

193. See VANESSA M. PEREZ, AMERICANS WITH PHOTO ID: A BREAKDOWN OF DEMOGRAPHIC CHARACTERISTICS 3 (2015), <https://perma.cc/96G6-MXAQ> (showing the disparities in the likelihood of having a government-recognized photograph ID among different races).
194. See, e.g., LOVRICH et al., *supra* note 180, at 13 (describing in Washington the "high proportion of out-of-state drivers" along Interstate 5 in particular and how that may be an important consideration in judging stops made by the Washington State Patrol).
195. The exact percentage of roadways policed by the Washington State Patrol depends on whether one uses lane miles or centerline miles. Using the former approach, it appears that the agency is responsible for around 11.2% of the state's lane miles, or 18,699 of 167,542. See *About Us*, *supra* note 22; *Highway Statistics 2018*, FED. HIGHWAY ADMIN., U.S. DEP'T OF TRANSP., <https://perma.cc/P7WR-9B7E> (last updated Oct. 31, 2019) (finding 167,542 lane miles in Washington). Using the latter approach, the Washington State Patrol appears to be responsible for policing around 8.7% of the state's centerline miles, or 7,000 of 80,338. Fed. Highway Admin., *Highway Statistics Series: Washington 1* (2015), <https://perma.cc/6ZAF-BXUG> (providing the 80,338 figure); WASH. STATE PATROL, 2017 ANNUAL REPORT 14 (2018), <https://perma.cc/V3FW-7BZV> (estimating that the Washington State Patrol oversees traffic enforcement on around 7,000 centerline miles).
196. *Racial Profiling and Traffic Stops*, NAT'L INST. JUST. (Jan. 9, 2013), <https://perma.cc/CM7C-U9WK>.
197. Alpert et al., *supra* note 146, at 45; Geoffrey P. Alpert, Roger G. Dunham & Michael R. Smith, *Investigating Racial Profiling by the Miami-Dade Police Department: A Multimethod Approach*, 6 CRIMINOLOGY & PUB. POL'Y 25, 32 (2007) (explaining that while census population data had been popular because it is "readily available at little or no cost," researchers "quickly reasoned that the static nature of the census did not represent the fluid nature of those who drove in the same areas").

to the patterns of people on the street.”¹⁹⁸ As discussed earlier, numerous scholars and experts have proposed alternative denominators for calculating stop rates like no-fault accident records¹⁹⁹ and visual-observation studies.²⁰⁰ But we have neither the access to these extensive no-fault accident records nor the resources to conduct a visual-observation study of drivers on the relatively small percentage of roads in Washington policed by the Washington State Patrol.

Ultimately, the inclusion of rates of stops (rather than the number of stops) is necessary in a study like ours only if the population of those policed by the Washington State Patrol on state and interstate highways has changed in some meaningful way between 2008 and 2015. Given the relatively short period of time we study, we think it is unlikely that the underlying population of drivers on the roads policed by the Washington State Patrol changed substantially. At minimum, we think this assumption is more defensible than the assumption that county or state population (or adjusted census population, or some other artificially constructed denominator) is an adequate substitute for the actual driver population along highways and interstates. Thus, we analyze stop levels rather than stop rates. The next Subpart begins our analysis of this stop data by examining raw trends and acknowledging additional methodological limitations.

C. Trends in the Raw Data

Before exploring the results of our more sophisticated modeling, it is helpful to examine the trends in the raw data. As a preliminary matter, it is worth calculating the change in stops before and after *Arreola*. To do this, Figures 1 through 7 graph the average number of traffic stops per county of white drivers (represented with a dashed line) as compared to various nonwhite drivers in the aggregate and by individual racial subgroups (represented with a solid line) in Washington in the years immediately before and after *Arreola*. To compare trends between the two groups, we plot the average number of stops per month per county of nonwhite drivers on the

198. Greg Ridgeway & John MacDonald, *Methods for Assessing Racially Biased Policing*, in RACE, ETHNICITY, AND POLICING: NEW AND ESSENTIAL READINGS 180, 182 (Stephen K. Rice & Michael D. White eds., 2010) (further explaining that “benchmarking with census data does not help us isolate the effect of racial bias from differential exposure and differential offending,” and also noting that “[c]ensus estimates provide only the racial distribution of residents and not how these numbers vary by time of day, business attractors such as shopping centers, daily traffic patterns involving commuters, and so forth”).

199. See, e.g., Alpert et al., *supra* note 197, at 34.

200. See *supra* note 145 and accompanying text (providing a detailed summary of visual-observation studies and other similar methodologies).

left-hand *y*-axis, and the average number of stops per month per county of white drivers on the right-hand *y*-axis. In some of these figures, we group together all nonwhite drivers as compared to white drivers. In other figures, we compare the trends in the number of stops of individual racial subgroups (for example, Black drivers, Hispanic drivers, and Asian drivers) to trends in the number of stops of white drivers.²⁰¹

As discussed in Subpart A above, we also recognize that the effects of *Arreola* may not have been immediate. It may have taken time for supervisors and legal counsel to train officers on this new standard, and it may have taken time for officers to understand fully how to employ mixed-motive stops. To address this uncertainty, we draw three different vertical lines to represent three different points at which the *Arreola* decision may have started to affect a substantial number of officers: (A) a line at the date of the *Arreola* decision in December 2012; (B) a line at the date that the *Arreola* decision first appeared in the WSCJTC law-enforcement digest in March 2013; and (C) a line at December 2013, signifying the date by which we would expect all or most officers to have completed their twenty-four-hour annual in-service training requirement after the *Arreola* decision.²⁰²

201. The trends in Figures 1-7 represent the average number of stops across all counties per month over the time window of our dataset. Though there are multiple acceptable ways to generate these graphs using any of many available statistical software packages, we used the *lpoly* command in Stata, which under the conditions of a zero-degree polynomial and a bandwidth set to one generates average trends across time.

202. See *supra* note 164 and accompanying text.

Figure 1

Average Stops per Month: White Drivers and Nonwhite Drivers in Washington,
Excluding Stops Where Race Is Unidentified (2008-2015)

Figure 2
Average Stops per Month: White Drivers and Nonwhite Drivers in Washington,
Including Stops Where Race Is Unidentified (2008-2015)

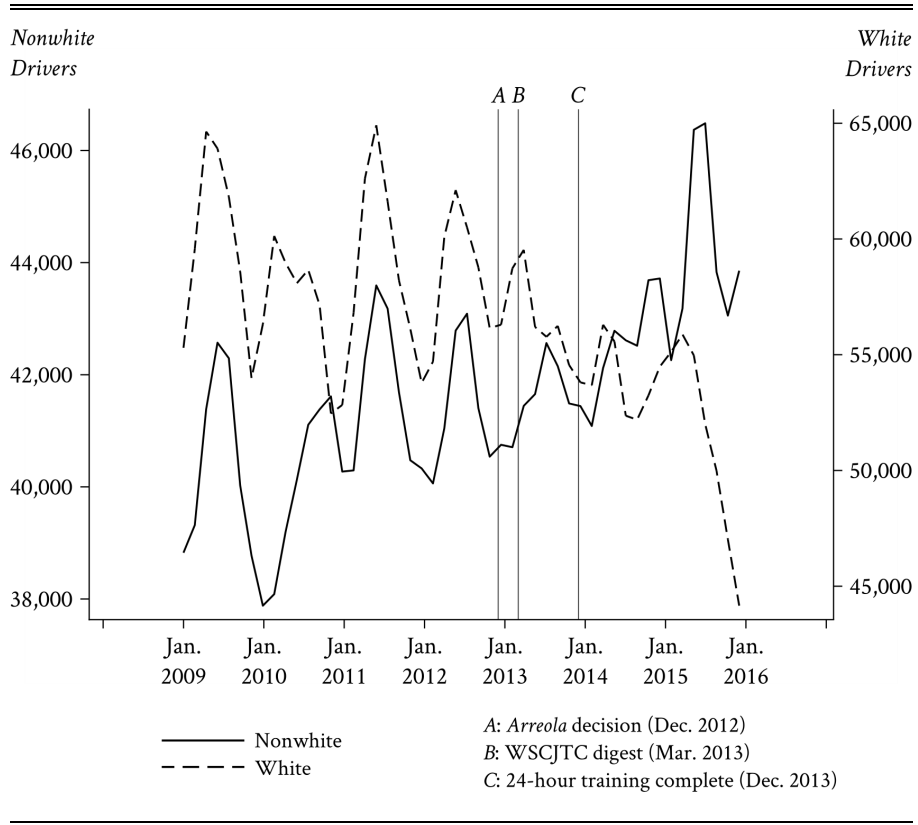


Figure 3
Average Stops per Month: White Drivers and Hispanic Drivers in Washington
(2008-2015)

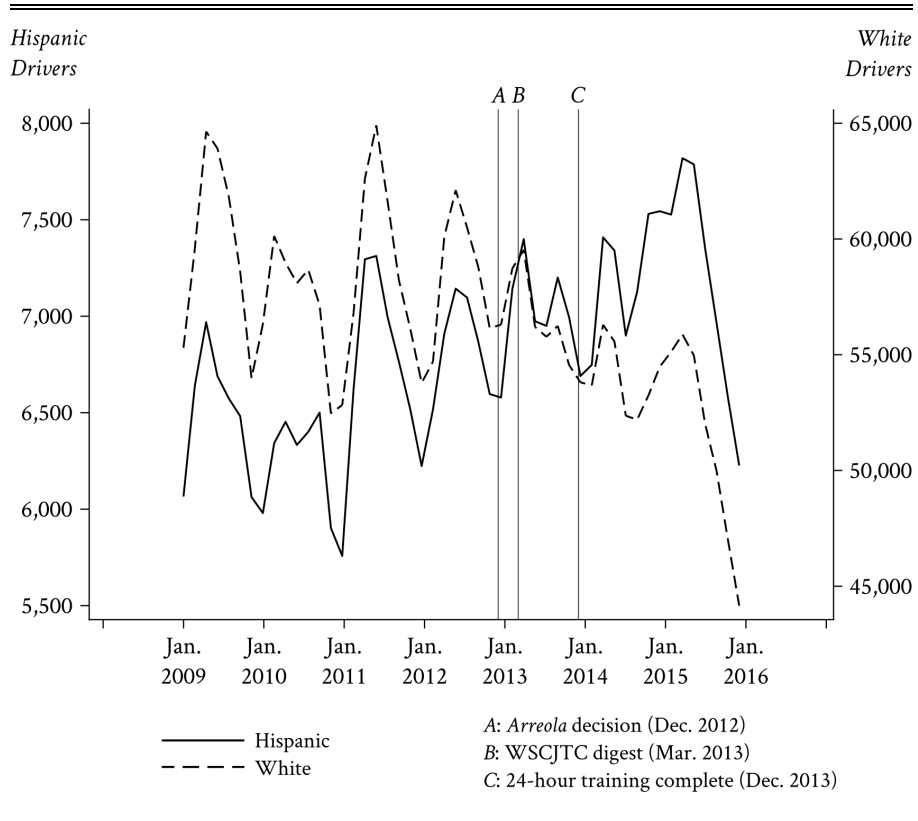


Figure 4
Average Stops per Month: White Drivers and Black Drivers in Washington
(2008-2015)

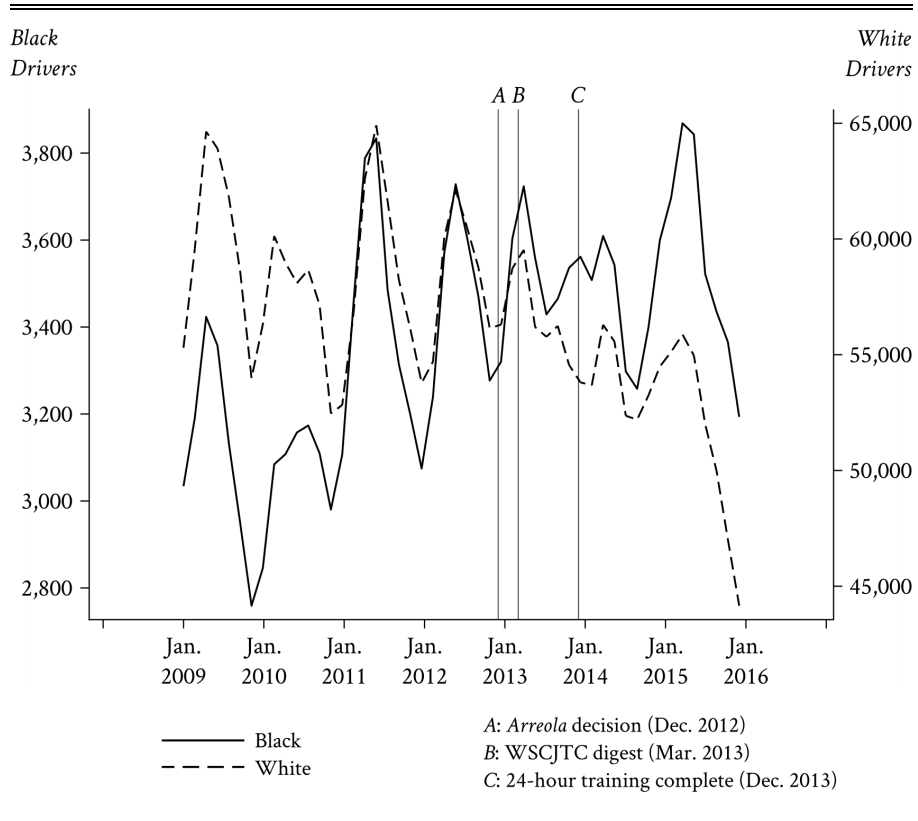


Figure 5
Average Stops per Month: White Drivers and Asian Drivers in Washington
(2008-2015)

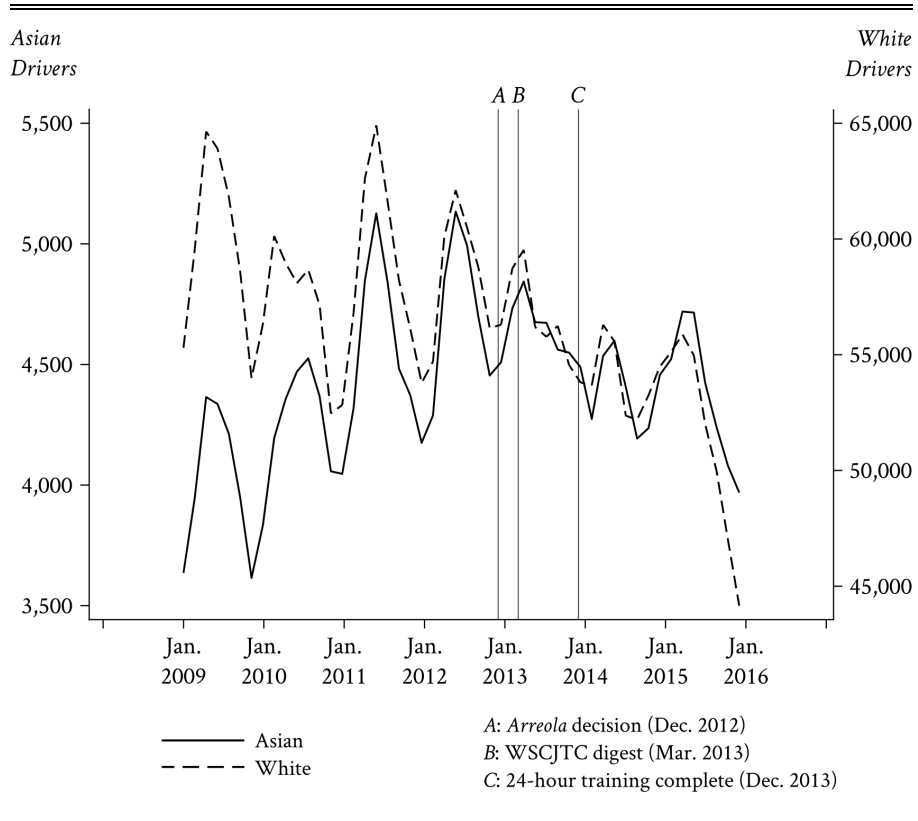


Figure 6
 Average Stops per Month: White Drivers and Drivers of
 Other Nonwhite Races in Washington (2008-2015)

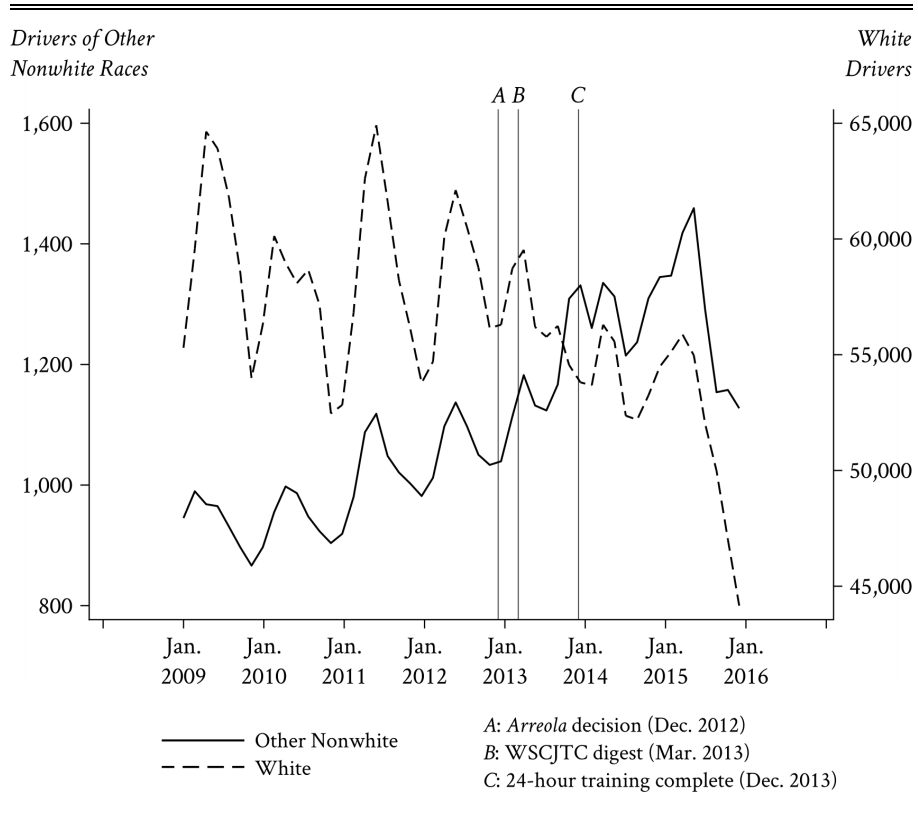
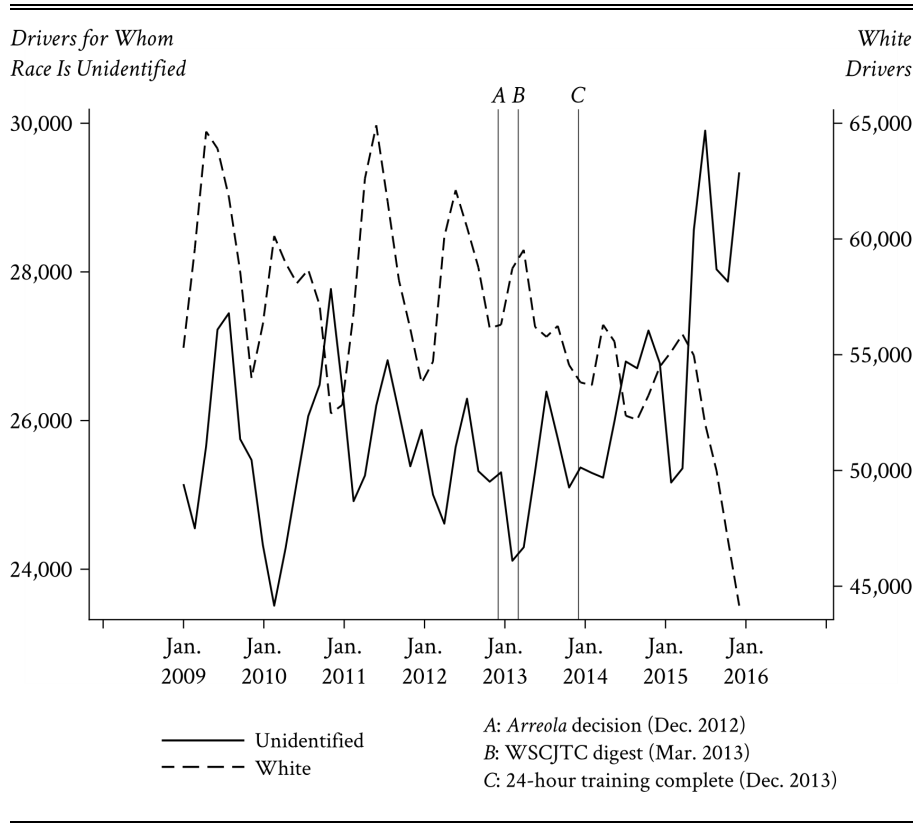


Figure 7
 Average Stops per Month: White Drivers and Drivers for Whom
 Race Is Unidentified in Washington (2008-2015)



As the data in Figure 1 illustrates, when we aggregate all nonwhite drivers together, we see that the number of stops of these drivers appears to grow relative to the stops of white drivers in the years after *Arreola*. When we break down our analysis by race, we see stronger evidence of increasing numbers of stops of Hispanic drivers after *Arreola* in Figure 3, with comparatively less evidence of such an increase among Asian drivers in Figure 5. There is also some evidence in Figures 1 and 2 that if *Arreola* had an effect on the behavior of officers, this effect may have grown somewhat over time—perhaps consistent with the theory that the effect would increase as officers were fully trained and understood how to use mixed-motive stops. We also see an increase in the number of stops of drivers with unidentified races in the years after *Arreola* in Figure 7, which could be consistent with the hypothesis that troopers increasingly failed to document the race of drivers of color when executing mixed-motive stops after *Arreola*.²⁰³ Thus, these general trend lines provide at least some possible support for the racial-profiling hypothesis.²⁰⁴

One challenge we faced in analyzing this dataset is the change in recreational-marijuana laws during this same time period. Washington legalized recreational marijuana around the same time that *Arreola* was decided.²⁰⁵ With marijuana becoming legal,²⁰⁶ police may have changed their behavior, perhaps by changing the frequency of stops or searches incident to stops. But even after legalization, questions remained about whether the smell of marijuana or the suspected presence of marijuana in a vehicle could serve as the basis for an automobile search.²⁰⁷ After all, driving under the influence of marijuana remains illegal, as does possession of large amounts of marijuana.²⁰⁸

203. See *supra* notes 178-80 and accompanying text.

204. To be clear, neither this evidence, nor any of the evidence in this Article, can prove any purposeful targeting of drivers by race. Any disproportionate effect of mixed-motive stops on drivers of color may be the result of implicit bias rather than any explicit bias. See generally Jennifer L. Eberhardt, Phillip Atiba Goff, Valerie J. Purdie & Paul G. Davies, *Seeing Black: Race, Crime, and Visual Processing*, 87 J. PERSONALITY & SOC. PSYCH. 876 (2004) (empirically testing the presence of implicit bias in visual observation); Kimberly Barsamian Kahn, Phillip Atiba Goff, J. Katherine Lee & Diane Motamed, *Protecting Whiteness: White Phenotypic Racial Stereotypicality Reduces Police Use of Force*, 7 SOC. PSYCH. & PERSONALITY SCI. 403 (2016) (finding that racial appearance may influence police use of force).

205. See Christina Ng, Abby Phillips & Clayton Sandell, *Colorado, Washington Become First States to Legalize Recreational Marijuana*, ABC NEWS (Nov. 6, 2012, 8:49 AM), <https://perma.cc/8DNS-TD4F>.

206. An Act Relating to Marijuana, 2013 Wash. Sess. Laws 28, 29, 42-44.

207. See Kip Hill, *Smell of Pot Can Be Basis for Police Search*, SPOKESMAN-REV. (July 27, 2014), <https://perma.cc/N4LR-GDL2>.

208. Michael Rubinkam, *In Era of Legal Pot, Can Police Still Search Cars Based on Odor?*, PBS NEWSHOUR (Sept. 13, 2019, 12:00 PM EST), <https://perma.cc/SV5P-A5SG> (“Judges have also ruled that marijuana odor can be used in conjunction with other factors to support
footnote continued on next page”).

So while it seems possible that the legalization of recreational marijuana may have influenced the number of stops and searches executed by the Washington State Patrol, officers can still search vehicles in some cases based on a suspicion of marijuana both before and after legalization.²⁰⁹ The number of searches of both white and nonwhite drivers declined throughout this period, which may be an indication that marijuana legalization had some impact on police behavior during traffic stops.²¹⁰

While marijuana legalization may have had some impact on the behavior of officers after a traffic stop has been initiated, it is less clear how the legalization of marijuana impacted the initial decision by officers to execute traffic stops. We nevertheless take several methodological approaches in the hopes of disentangling the effect of *Arreola* on traffic stops from any possible effect of marijuana legalization. The next Subpart walks through these more sophisticated methodologies.

D. Effects of *Arreola* on Traffic Stops

To evaluate the effects of *Arreola* on the number of police traffic stops of drivers of color relative to white drivers, we conduct both an ordinary difference-in-differences analysis and a series of more sophisticated regressions.²¹¹ This methodology calculates the difference in police behavior toward a treatment group and compares that difference to a baseline difference from a control group. For example, consistent with the hypothesis advanced by many scholars and civil rights activists after *Whren*, it seems plausible that

a search. If the smell is overpowering, for example, an officer might conclude the motorist has a quantity of cannabis far in excess of what's allowed. Driving under the influence of marijuana is illegal in all 50 states, so police are free to search the car of a driver who shows signs of impairment.”).

209. Additionally, at least one court has heard an appeal to determine whether a state police officer in a jurisdiction with legalized recreational marijuana may use the federal prohibition on marijuana as probable cause to justify an automobile search. See Orin S. Kerr, *Can a State Police Officer Search a Car Based on Probable Cause of a Federal Marijuana Crime?*, REASON (Feb. 17, 2020, 2:31 AM), <https://perma.cc/SYY2-HYZJ>; *United States v. Martinez*, 811 F. App'x 396, 397-98 (9th Cir. 2020).

210. In the Appendix below, we provide various figures showing trends in the number of searches of white and nonwhite drivers. See *infra* Appendix, Part D.

211. For examples of prior studies using the difference-in-differences methodology, see Michael Lechner, *The Estimation of Causal Effects by Difference-in-Difference Methods*, 4 *FOUND. & TRENDS ECONOMETRICS* 165, 168-69 (2010) (providing a review of literature on the use of difference-in-differences in empirical studies); and Elizabeth A. Stuart, Haiden A. Huskamp, Kenneth Duckworth, Jeffrey Simmons, Zirui Song, Michael E. Chernew & Colleen L. Barry, *Using Propensity Scores in Difference-in-Differences Models to Estimate the Effects of a Policy Change*, 14 *HEALTH SERVS. & OUTCOMES RSCH. METHODOLOGY* 166, 167 (2014) (explaining that social scientists employ difference-in-differences methods to evaluate the effect of changes in laws).

by authorizing pretextual-like stops in *Arreola*, the Washington Supreme Court may have facilitated racial profiling.²¹² If this hypothesis is true, we would expect the introduction of *Arreola* (the independent variable) to result in a change in police treatment of drivers of color (the dependent variable). And if this hypothesis is true, we would expect *Arreola* to have less of an effect on police treatment of white drivers.

So, to conduct a simple difference-in-differences analysis, we use changes in the number of stops of white drivers before and after *Arreola* as our baseline (or control group), and we compare this difference to the change in the number of stops of drivers of color (our treatment group). To formally calculate the difference-in-differences estimate of *Arreola*'s effect on nonwhite drivers relative to white drivers, we calculate the following differences:

Model 1

$$\beta = (Stops_{post-Arreola}^{nonwhite} - Stops_{pre-Arreola}^{nonwhite}) - (Stops_{post-Arreola}^{white} - Stops_{pre-Arreola}^{white})$$

In conducting these difference-in-differences estimates, we averaged results across all nonwhite racial identifiers, resulting in an average difference-in-differences of around 120 additional stops per county per month of nonwhite racial subgroups relative to white drivers, excluding “n/a” cases (as shown in Table 2). We ran separate analyses for each nonwhite racial group. For example, running this analysis for Black drivers resulted in a difference-in-differences of approximately 119 additional stops relative to white drivers per county per month after *Arreola*, as shown in Table 3. And running this same analysis for Hispanic drivers resulted in an additional 127 stops relative to white drivers per county per month, also shown in Table 3. This number remains relatively stable (around 121) if we include “n/a” cases as a separate class of nonwhite drivers, as shown in Table A.2 in the Appendix. This result suggests that relative to changes in the number of stops of white drivers over the same time period, the number of stops for each nonwhite racial subgroup of drivers (Black, Hispanic, Asian/Pacific Islander, and other) appears to have increased by an average of around 120 stops per county per month after *Arreola*. This initial test suggests that *Arreola* may have had a larger effect on drivers of color than white drivers. Nevertheless, it does not allow us to make any causal claims, nor does it allow us to estimate the statistical significance of our findings because it does not account for confounding factors.

To bolster our analysis, we employ a multiple-regression technique common for studies that employ a difference-in-differences framework.²¹³

212. See *supra* Part I.B.

213. See, e.g., Griffin Edwards, Stephen Rushin & Joseph Colquitt, *The Effects of Voluntary and Presumptive Sentencing Guidelines*, 98 TEX. L. REV. 1, 28-32 (2019) (using difference-
footnote continued on next page

Regressions allow us to both estimate standard errors and include other measurable factors that may be influencing traffic stops, such as driver age, officer race, officer gender, stop location, and time since *Arreola*. Formally, we estimate:

Model 2

$$O_{ikt} = \alpha + \partial nonwhite_{ik} + \varphi Arreola_t + \beta nonwhite * Arreola_{it} + \theta X + \varepsilon$$

The difference-in-differences approach is formalized in regression format by the inclusion of a dummy variable for the affected group, or *nonwhite*, which varies by racial group *i* and county *k*; a time dummy variable that flags all months and years post-*Arreola*, which varies by time *t*; and an interaction between the two, or *nonwhite * Arreola*, which varies by racial group *i* and time *t*. Additionally, depending on the specification, we include county-level fixed effects and controls for driver age, officer race, and officer gender, all encapsulated in vector *X*. The standard errors for each estimate are two-way clustered by county and year, and all inferences are based on conventional two-tailed tests at normal significance cutoffs.²¹⁴

This model allows us to estimate more precisely the relationship between *Arreola* and any subsequent changes while controlling for alternative explanatory variables. Table 2 presents the first results from this difference-in-differences modeling, focusing specifically on the estimated effect of *Arreola* on the number of stops conducted by Washington State Patrol troopers of nonwhite drivers relative to white drivers. The results represent the change in the average total number of stops of each nonwhite racial subgroup (for example, Black drivers, Hispanic drivers, Asian drivers, and drivers whose race was categorized as “other”) relative to white drivers per county per month since *Arreola*. In this model, a positive number indicates an increase in the relative number of stops per month and a negative number indicates a decline in the relative number of stops per month. Table 2 removes all stops where race is unidentified from the dataset. We reran this analysis including these unidentified race cases and produced substantially similar results, which are available in the Appendix.²¹⁵

in-differences frameworks and multiple-regression techniques to estimate the effect of changes in sentencing guidelines in Alabama on judicial behavior); Stephen Rushin & Griffin Edwards, *De-Policing*, 102 CORNELL L. REV. 721, 754-58 (2017) [hereinafter Rushin & Edwards, *De-Policing*] (using this same methodological approach to assess the influence of federal intervention on crime rates).

214. For a longer discussion of the modeling choices including controls, fixed effects, and standard errors, see Appendix, Part A below.

215. See *infra* Appendix, Part B; *infra* Appendix, Table A.2.

Table 2
Effect of *Arreola* on Stops of Nonwhite Drivers Relative to White Drivers,
Excluding Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of nonwhite drivers	120.019** (39.059)	105.896*** (11.953)	115.434** (37.580)
Mean driver age	—	-22.335** (6.389)	-8.718** (2.761)
Officer race	—	-1,098.019* (488.923)	109.179 (132.102)
Officer gender	—	-822.652 (433.011)	76.804 (144.741)
<i>R</i> -squared	0.288	0.326	0.536

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 15,470. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

As seen in Table 2, our model estimates that stops of nonwhite racial subgroups of drivers per county per month relative to white drivers increased by a statistically significant margin in the years after *Arreola*. These results are statistically significant when we add in both controls and county-level fixed effects. Group-level fixed effects are essentially a set of dummy variables for each group—in our case, each county—in the dataset.²¹⁶ Fixed-effect models are a common strategy used in empirical work to capture unobserved, idiosyncratic, time-invariant factors that vary across the group being studied.²¹⁷ This gives us some confidence that *Arreola* may be contributing to an uptick in the number of stops of drivers of color relative to white drivers.

Table 2 also reports the results for our set of controls, which include the mean driver age, officer race, and officer gender. These latter two control variables are bounded between zero and one in our model (with unconditional averages of 93% and 96% respectively in the dataset). Thus, in operationalizing the officer gender and officer race variables, our model estimates the effects of shifting from zero (all nonwhite or all female officers) to one (all white or all male officers). In interpreting the results of this regression, it may therefore be useful to scale these numbers down to represent a more realistic shift in the racial or gender makeup of the Washington State Patrol. Say that the percentage of white officers conducting traffic stops shifted by one percentage point (for example, from 92% to 93%). Based on our model, we would expect this one-percentage-point shift in the underlying racial makeup of the police force to be associated with a reduction of 10.98 traffic stops per county per month. Once we include fixed effects in this model, this shift in the racial makeup of the force becomes nearly indistinguishable from zero (1.098 additional stops per county per month).

The results from Table 2 assume that officers began employing the new mixed-motive stop doctrine permitted by *Arreola* immediately after the Washington Supreme Court issued its holding in December 2012. But as discussed in Part C of the Appendix, it seems theoretically plausible that the effects of *Arreola* may have been delayed until officers received training in how to employ these new mixed-motive stops.²¹⁸ To test this hypothesis, we ran additional specifications of our model that altered the date of the *Arreola* decision to match: (1) the date the opinion first appeared in the WSCJTC

216. See, e.g., JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE, *MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST'S COMPANION* 221-27 (2009); Brett Parker, Note, *Is Death Different to Federal Judges? An Empirical Comparison of Capital and Noncapital Guilt-Phase Determinations on Federal Habeas Review*, 72 STAN. L. REV. 1655, 1683 n.147 (2020).

217. For more fixed-effects information, see Appendix, Part A.5 below. For an example of another study using fixed effects, see Edwards et al., *supra* note 213, at 29 (describing the use of fixed effects in a prior study).

218. See *supra* Part III.A.

monthly training bulletin (March 2013);²¹⁹ and (2) the date by which all officers in the state should have completed their annual in-service training requirement post-*Arreola* (December 2013).²²⁰ To the extent that *Arreola* is driving the change in the number of stops of drivers of color relative to white drivers, the results of these additional tests suggest that this effect may have been greatest after officers had time to understand how to employ mixed-motive stops. The results of these additional tests are available in the Appendix.²²¹

We also disaggregated our data and ran separate regressions comparing changes in the number of monthly traffic stops per county per month for individual racial groups (in our case, Black drivers, Hispanic drivers, Asian drivers, and drivers of other nonwhite races) relative to white drivers. We find that even if we limit our analysis to individual racial subgroups, the effect remains statistically significant. Table 3 reproduces these findings, focusing specifically on Black and Hispanic drivers—two groups most commonly cited as being victimized by racial profiling.²²²

219. See *supra* note 169 and accompanying text.

220. See *supra* note 164 and accompanying text.

221. See *infra* Appendix, Part C.

222. See, e.g., Baumgartner et al., *supra* note 147, at 108 (specifically mentioning the possibility of pretextual stops targeting “black and brown drivers”); Rudovsky, *supra* note 14, at 301 (mentioning the targeting of African American drivers); Harris, *supra* note 73 (finding possible targeting of Latino drivers).

Table 3
Effect of *Arreola* on Stops of Black and Hispanic Drivers Relative to White Drivers
(2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of Hispanic drivers	127.710** (38.217)	133.953*** (24.514)	129.026*** (32.755)
Change in stops of Black drivers	119.019** (39.270)	108.585*** (27.876)	116.383** (36.368)
Mean driver age	—	-22.044** (6.700)	-7.003** (2.556)
Officer race	—	-1,110.949* (486.933)	77.780 (124.491)
Officer gender	—	-863.607* (432.726)	19.992 (151.347)
<i>R</i> -squared	0.290	0.328	0.538

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 15,470. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

As shown in Table 3, we find the greatest relative increase in traffic stops among Hispanic drivers after *Arreola*. But we nevertheless see a statistically significant uptick in the number of stops of Black drivers relative to white drivers. These results also remain statistically significant after the introduction of controls and county-level fixed effects.

We also find evidence that the treatment of nonwhite drivers after a traffic stop, relative to white drivers, may have shifted after *Arreola* in a manner consistent with the racial-profiling hypothesis. These results are discussed in more length in Part D of the Appendix. Combined, these analyses suggest that once police were given more discretionary authority under the Washington pretextual-stop doctrine, they may have used this authority to disproportionately target drivers of color for additional stops.

There is another way to test whether officers may be targeting nonwhite drivers more after *Arreola*. If police responded to *Arreola* by targeting nonwhite drivers for additional scrutiny, we would expect this result to be more evident during the daytime than at night. Prior racial-profiling studies have operated under the belief that racial profiling happens when police officers are able to ascertain a driver's race, usually through visual observation, and then use this observation in deciding whether to execute a traffic stop.²²³ Presumably, police officers will be able to determine the race of a suspect more easily during the daytime than at night. Thus, if *Arreola* is driving the apparent change in the treatment of drivers of color by Washington state troopers from 2013 to 2015, we would expect this effect to be concentrated in the daytime hours rather than at night. The next Subpart employs this veil-of-darkness methodology²²⁴ as a robustness check of our findings.

E. Effects of Daylight on Traffic Stops

If the changes we observe in the number of traffic stops of nonwhite drivers after *Arreola* are truly the result of increased racial profiling, we would expect the changes to be most evident during daylight hours, when it is easiest for a police officer to discern the race of the driver.²²⁵ Thus, if the hypothesized link between racial profiling and pretextual stops is present, we would expect this apparent effect to be strongest during daylight hours and weakest during the darkest hours of the night and morning.

To test this theory, we run triple-difference regressions. These regressions mirror the difference-in-differences regressions from the previous Subpart but

223. See *supra* note 148 and accompanying text.

224. See *supra* note 148 and accompanying text (describing prior usage of the veil-of-darkness methodology).

225. See, e.g., Grogger & Ridgeway, *supra* note 135, at 878 (describing this methodology in detail).

add one more layer of analysis: They compare the difference-in-differences estimates obtained from stops conducted in daylight and at night. Formally, we estimate this regression using the following equation:

Model 3

$$O_{ikt} = \alpha + a_1 \text{nonwhite}_{ik} + a_2 \text{Arreola}_{kt} + a_3 \text{dark}_{kt} + a_4 \text{nonwhite}_{ik} * \text{Arreola}_{ikt} + a_5 \text{nonwhite}_{ik} * \text{dark}_{ikt} + a_6 \text{Arreola}_{kt} * \text{dark}_{ikt} + a_7 \text{nonwhite}_{ik} * \text{dark}_{ikt} * \text{Arreola}_{kt} + \theta X + \varepsilon$$

We define “daylight hours” as the time between sunrise and sunset, and we define “dark hours” as the time between the end of nautical twilight in the evening and the start of nautical twilight the next morning.²²⁶ The remaining window, which we describe as “twilight hours,” covers the periods between nautical twilight and sunrise in the morning and between sunset and the end of nautical twilight in the evening. Since it is not clear whether an officer could visually observe a driver’s race during twilight hours, we exclude the stops made during these hours (which make up only 7% of total stops).²²⁷

Table 4 presents our findings on the differences between stops of drivers of color relative to white drivers during the daytime and nighttime after *Arreola*. This table shows the change in the number of stops per county per month at night (as compared to daytime) of nonwhite drivers (as compared to white drivers). If police are engaged in racial profiling after *Arreola*, we would expect to see police stopping more nonwhite drivers relative to white drivers during the day, and we would expect this imbalance to decrease at night.

226. For the definition of nautical twilight, see Konstantin Bikos, *Nautical Twilight—Nautical Dawn and Dusk*, TIME & DATE, <https://perma.cc/4K7Q-J2ZS> (archived Feb. 22, 2021).

227. See generally Christian Bünnings & Valentin Schiele, *Spring Forward; Don’t Fall Back: The Effect of Daylight Saving Time on Road Safety*, 103 REV. ECON. & STAT. 165 (2021) (providing a further discussion of the various uses of twilight for social-science research).

Table 4
Stops of Nonwhite Drivers at Night Relative to Day Post-*Arreola*,
Excluding Stops Where Race Is Unidentified (2008-2015)

Triple-difference regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Post- <i>Arreola</i> stops of nonwhite drivers at night relative to day, relative to white drivers	-40.346* (17.298)	-45.934*** (8.804)	-41.153** (16.334)
Mean driver age	—	-7.942** (2.290)	-3.350** (1.060)
Percentage of stops by white officers	—	-416.851* (183.158)	42.801 (45.536)
Percentage of stops by male officers	—	-292.555 (157.382)	38.366 (41.263)
<i>R</i> -squared	0.284	0.306	0.493

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 28,676. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4 finds evidence consistent with the hypothesis that police may be targeting nonwhite drivers for additional scrutiny after *Arreola*. Police appear particularly likely to target drivers of color during daylight hours. And at night, the number of stops of drivers of color per county per month relative to white drivers decreases by about forty relative to daytime. These results are statistically significant with and without the introduction of controls and fixed effects. They remain statistically significant when we include stops where race is unidentified in our definition of nonwhite drivers, as seen in the Appendix.²²⁸

F. Event Study

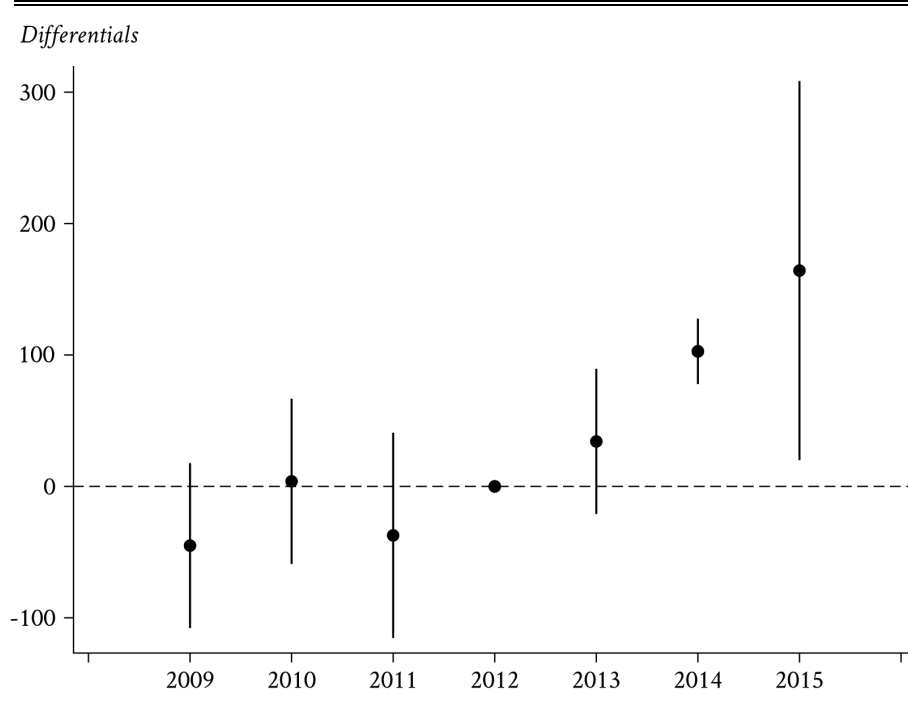
As a final test of the effect of *Arreola* on officer behavior, we conduct an event study. We use this methodology to determine whether the purported change in officer behavior towards nonwhite drivers can be fairly attributed to the *Arreola* decision from December 2012 (and the subsequent training that occurred in 2013) rather than other contemporaneous events.²²⁹ Essentially, this methodology calculates the difference in traffic-stop patterns between white and nonwhite drivers over time. The lines that extend above and below each data point represent confidence intervals. If the entirety of this line is above or below zero, then we can say at some level of confidence that the differential between trends in white and nonwhite traffic stops is statistically significant. By contrast, if this confidence interval extends both above and below zero, then we cannot say that the results are statistically significant. Put differently, if *Arreola* is driving any subsequent changes in stops of nonwhite drivers relative to white drivers, we would expect the differentials to become statistically significant only after the court issued its opinion in *Arreola* in December 2012 and officers received training in this new technique in 2013. Figure 8 presents the results of this analysis.²³⁰

228. As is sometimes customary with triple-difference regressions, we report in the Appendix the results of the “placebo” difference-in-differences regressions (that is, comparing stops for just white drivers during the day and at night). However, we urge caution in reading too much into these results for reasons outlined more fully in the Appendix, Parts A.4 and B.

229. The use of this methodology for this purpose is consistent with prior studies. See generally Melissa S. Kearney & Phillip B. Levine, *Media Influences on Social Outcomes: The Impact of MTV’s 16 and Pregnant on Teen Childbearing*, 105 AM. ECON. REV. 3597 (2015) (using this approach to attribute changes in teen pregnancy to an MTV television show on teenage pregnancy); Griffin Edwards, Erik Nesson, Joshua J. Robinson & Frederick Vars, *Looking Down the Barrel of a Loaded Gun: The Effect of Mandatory Handgun Purchase Delays on Homicide and Suicide*, 128 ECON. J. 3117, 3134-35 (2018) (using such an event study in a methodologically similar way).

230. The year 2012 acts as our comparison in the event study since it was the last full year without treatment. In these event-study-style regressions, we necessarily must use one
footnote continued on next page

Figure 8
Event Study of Differentials in Stops of White Drivers and Nonwhite Drivers,
Excluding Stops Where Race Is Unidentified (2009-2015)



year as a baseline for comparison. Because we estimate no effect in 2012 it has no confidence bounds. We signify that in Figure 8 with the dot at zero.

In the year leading up *Arreola*, the differentials in the trends of white and nonwhite traffic stops are statistically insignificant. These differentials only become statistically significant after the *Arreola* decision. This is true regardless of whether we include stops where race is unidentified in our analysis.²³¹

G. Methodological Limitations

While we believe that our results provide evidence consistent with the hypothesis that *Arreola* resulted in a disproportionate number of stops of drivers of color, it is important to recognize the limitations of our dataset and methodology. First, this study focuses specifically on the Washington State Patrol and does not cover all law-enforcement behavior in Washington. The Washington State Patrol has a somewhat different set of law-enforcement priorities than many municipal police and sheriffs' departments.²³² These differences in responsibilities may call into question whether our findings are generalizable to other law-enforcement agencies in Washington or the rest of the United States. Despite this potential limitation, we still believe that the Washington State Patrol is a particularly useful agency in which to study the effects of judicial regulation of traffic-code enforcement because of the enormous volume of traffic stops the agency conducts across the state.²³³ But we acknowledge that any differences in job responsibilities between state troopers and municipal police officers or sheriff's deputies could limit the generalizability of our findings.

Second, our analysis is limited to a single law-enforcement agency. This limitation was unavoidable. The Washington State Patrol is the only law-enforcement agency that appears to have kept this kind of extensive traffic-stop data in any of the two states (New Mexico and Washington) that

231. See *infra* Appendix, Figure A.1.

232. See *About Us*, *supra* note 22. For a detailed description of the law-enforcement priorities of this agency, see *Crime*, WASH. ST. PATROL, <https://perma.cc/L5YN-JX7W> (archived Jan. 8, 2021) (describing the agency's responsibility for "[v]essel and terminal safety," certain investigation services, investigations of missing children and most wanted criminals, and criminal and collision records). These responsibilities may be compared to local law-enforcement agencies, which unlike their state counterparts, may be focused more on local policing issues. See, e.g., *About Us: About Policing*, CITY OF SEATTLE, <https://perma.cc/B4W2-7SWC> (archived Feb. 22, 2021) (describing the mission and jurisdiction of the Seattle Police Department, including parking enforcement, SWAT, the management of a 9-1-1 center, and a canine unit); *Crime Prevention and Safety*, CITY OF VANCOUVER, WASH., <https://perma.cc/WYB5-BCY9> (archived Feb. 22, 2021) (describing the crime prevention efforts of the police department in Vancouver, Washington, including responding to graffiti, construction-site-theft prevention, home safety, and internet safety).

233. See *supra* notes 21-22 and accompanying text.

experimented with different rules for pretextual stops.²³⁴ Seattle and Tacoma have kept data on traffic stops since 2005 and 2007, respectively.²³⁵ But neither jurisdiction has provided the Stanford Open Policing Project with data on the race of drivers stopped by police officers.²³⁶ And as far as we can tell, no jurisdiction in New Mexico keeps consistent, publicly available data on traffic stops sufficient for this type of rigorous analysis.²³⁷ The depth and extensiveness of the Washington State Patrol dataset, though, helps alleviate some of the concerns about our focus on a single jurisdiction.

Third, our model cannot control for other political forces at work around the time of the *Arreola* decision. As mentioned previously, the legalization of marijuana occurred almost simultaneously to *Arreola*.²³⁸ Legalization poses a threat to our empirical strategy if there is reason to believe that police officers changed their behavior in response to legalized marijuana differently across racial groups. It seems likely that marijuana legalization changed law-enforcement behavior in some way. Provided their behavior changed the same across races, our results remain valid. To be more specific, as long as the resulting change in police behavior affected racial groups equally, then it would not result in any changes in our difference-in-differences or triple-difference models. Based on visual inspection of the trends in stop patterns by racial group as seen in Figure 2 and the formal analysis of pretrends in Figure 8,²³⁹ we fail to find any evidence to suggest disproportionate responses by race in the months and years leading up to marijuana legalization. That is, we suspect that most law-enforcement agencies in Washington knew that recreational marijuana may become legalized. The downward trend in stops and searches of vehicles prior to the legalization date may reflect the anticipated legalization.²⁴⁰

234. See *supra* note 78 and accompanying text.

235. *Data, supra* note 171.

236. *Id.* (showing no data on driver race for these jurisdictions).

237. *Id.* (showing in the list of available datasets that no jurisdiction in New Mexico has provided data to the Project).

238. See *supra* notes 205-06 and accompanying text.

239. Simon Freyaldenhoven, Christian Hansen & Jesse M. Shapiro, *Pre-event Trends in the Panel Event Study Design*, 109 AM. ECON. REV. 3307, 3307 (2019) (explaining that “[a] common diagnostic approach in such settings is to look at whether the policy change appears to have an effect on the outcome before it actually occurs” and that “[t]he presence of such pre-event trends or ‘pre-trends,’ is taken as evidence against the strict exogeneity of the policy change”).

240. Note again in Figure 3 how closely the pretrends in white and nonwhite stop counts track one another. If there were reason to believe that officers, say, decreased the rate at which they stopped white drivers, but not nonwhite drivers, due to marijuana, that should be apparent in the buildup to the new law when officers began to anticipate the change. We see no such divergence in trends.

Additionally, important political changes occurred around the same time as *Arreola*. Washington voters elected a new governor²⁴¹ and attorney general,²⁴² and Republicans took control of the state senate.²⁴³ Any of these changes may suggest political endogeneity—meaning that any underlying change in politics (rather than *Arreola*) could be driving our results. While we cannot discount this possibility, we believe that the evidence is still mostly consistent with the hypothesis that permissive pretextual-stop doctrines may contribute to more stops of drivers of color, particularly since it appears that the effect of *Arreola* may strengthen over time—a trend that may correspond with the gradually increasing dissemination of the new rule through training. Nevertheless, we cannot discount the possibility that other factors are also contributing to the changes in police behavior.

Finally, as with any empirical study of this type, readers should view our study for what it is: “one data point in what will hopefully be a growing literature” on the effect of pretextual-stop doctrines on police behavior.²⁴⁴ An ideal study of the effect of cases like *Whren* and *Arreola* on officer behavior would employ complete datasets from a wide range of jurisdictions that keep data in a sufficiently similar manner (so as to allow for cross-jurisdictional comparisons). Unfortunately, such data is not yet available. Thus, our results will likely need to be replicated, hopefully in new locations as more jurisdictional variation emerges among states and municipalities.

IV. Implications for the Law of Policing

Our findings have important implications for the law of policing. First, and primarily, our results are consistent with the predictions made by many scholars after *Whren*.²⁴⁵ The data from Washington suggest that legal rules giving police officers increased discretion to conduct pretextual or mixed-motive traffic stops may contribute to inequality by facilitating racial profiling. More generally, our analysis suggests that rules granting police discretion in traffic stops may lead to more traffic stops of drivers of color, with some likely escalating to more serious encounters. Second, our findings

241. Jim Brunner, *McKenna Concedes; Inslee to Be Governor*, SEATTLE TIMES (updated Nov. 10, 2012, 1:08 PM), <https://perma.cc/36VL-KZTP>.

242. *Washington's Attorneys General—Past and Present*, WASH. ST. OFF. ATT'Y GEN., <https://perma.cc/59GV-DKM6> (archived Jan. 8, 2021) (showing that Bob Ferguson was elected as Attorney General in November 2012 and took office in 2013).

243. *Party Control of Washington State Government*, BALLOTPEDIA, <https://perma.cc/T89S-VM23> (archived Jan. 8, 2021) (showing that Republicans took control of the state senate in 2013 and held power until 2017).

244. Rushin & Edwards, *De-Policing*, *supra* note 213, at 772.

245. *See supra* Part I.B.

are particularly troubling given that the victims of racial profiling during traffic stops often have limited means of seeking redress for the harms they suffer. With these negative consequences in mind, our results may bolster two emerging reform proposals: (1) the idea that traffic-code enforcement should be decoupled from the investigation of more serious criminal offenses; and (2) the notion that we should remove discretion in traffic enforcement through the integration of technological enforcement tools.

A. Harmful Consequences of *Whren*

As discussed in Part II above, a large and growing body of literature has found suspicious patterns in traffic-stop data in communities around the United States.²⁴⁶ These studies suggest that police in a significant number of jurisdictions may consider a driver's race—either consciously or subconsciously—in executing traffic stops.²⁴⁷ What has remained somewhat less clear, though, is the extent to which deferential judicial decisions like *Whren* contribute to this pattern of apparent racial profiling. Why are police officers in jurisdictions across the country enforcing traffic laws more harshly against minority drivers than white drivers? Is it because of explicit racial bias? Implicit bias? A lack of existing controls? Or perhaps some combination of all of these factors? Our data helps resolve that controversy. It provides evidence that decisions endorsing pretextual traffic stops may be one contributor to racial profiling by police officers. If *Arreola*, with its somewhat narrower holding than *Whren*, has potentially contributed to an increase in stops and searches of drivers of color relative to white drivers across Washington, it stands to reason that *Whren* may have similarly facilitated racial profiling. By giving police officers a license to act on their hunches or suspicions via pretextual or mixed-motive stops, both *Whren* and *Arreola* may lead to more officers treating drivers of color differently because of implicit—or explicit—bias.

This finding may, in turn, guide policymakers looking to prevent racial profiling. Our data suggests that police-reform advocates concerned about racial bias in policing should consider lobbying for legislative enactments that provide additional protections against pretextual stops. With *Whren* decided a little over two decades ago, it seems unlikely that the Supreme Court will reconsider its holding anytime soon. But this does not prevent states from using their legislative powers to enact limitations on police authority to conduct pretextual stops.

246. See *supra* notes 120-48 and accompanying text (describing a wide variety of studies demonstrating apparent patterns of racial bias in jurisdictions across the country).

247. See *supra* notes 120-48 and accompanying text.

If moving from *Ladson* to *Arreola* contributed to a statistically significant increase in apparent racial profiling by Washington state troopers, this would suggest that the *Ladson* decision did exert at least some influence on police behavior.²⁴⁸ It may have suppressed some stops of drivers of color and may have reduced the willingness of police to engage in racial profiling. This realization is important, as it suggests that *Whren's* holding was not merely symbolic. Had the Court ruled differently—for example, by developing a rule similar to that introduced by Washington in the *Ladson* case—it conceivably could have influenced police behavior in a way that reduced racial bias by officers. Such a change could have had major implications for the lives of millions of Americans. As Charles R. Epp, Steven Maynard-Moody, and Donald Haider-Markel have previously argued, traffic stops by police “matter” because “[n]o form of direct government control comes close to these stops in sheer numbers, frequency, proportion of the population affected, and, in many instances, the degree of coercive intrusion.”²⁴⁹ Police conduct an estimated 18 million traffic stops every year.²⁵⁰ These stops “convey powerful messages about citizenship and equality.”²⁵¹ Thus, states could theoretically use the *Ladson* holding as one model for enacting stricter regulations of pretextual stops. And based on Washington’s experience, it seems possible that stricter regulation of pretextual stops could have widespread implications for the relationship between police and communities of color across the country.

Relatedly, our findings may have important implications for the study of police violence. In the years since the protests in Ferguson, Missouri, in 2014, media outlets and civil rights groups have attempted to document the frequency of civilian deaths at the hands of American law-enforcement officials.²⁵² Databases like those maintained by the *Washington Post*,²⁵³ the

248. See *supra* Parts III.D-F.

249. EPP ET AL., *supra* note 15, at 2.

250. *Id.*

251. *Id.*; see also Joe Soss & Vesla Weaver, *Police Are Our Government: Politics, Political Science, and the Policing of Race-Class Subjugated Communities*, 20 ANN. R. POL. SCI. 565, 567, 569 (2017) (discussing the “social control” face of American government and its impact on communities of color).

252. Jamiles Lartey, *US Police Killings Undercounted by Half, Study Using Guardian Data Finds*, GUARDIAN (Oct. 11, 2017, 7:00 AM EDT), <https://perma.cc/H9CQ-DCLH> (discussing a study on the number of individuals killed by police and noting that these kinds of projects were “intended to help remedy the lack of reliable data on police killings, a lack that became especially visible after the 2014 unrest in Ferguson”).

253. *Fatal Force*, WASH. POST, <https://perma.cc/L3UV-9WWH> (last updated Feb. 4, 2021) (providing estimates for the number of individuals shot and killed by police from 2015 through 2021).

Guardian,²⁵⁴ *Fatal Encounters*,²⁵⁵ and *Mapping Police Violence*²⁵⁶ have attempted to document not just the number of killings by police, but also the circumstances that contribute to these deaths. One topic, though, has received somewhat less attention in the growing literature on police violence: how traffic stops serve as the starting point for many violent interactions between police and civilians, including those interactions that ultimately result in officers utilizing deadly force. No existing database provides an easy way to search for police killings that happened after police officers executed a traffic stop. Nevertheless, a quick analysis of the *Guardian* database from 2016 suggests that a substantial number of such incidents began with traffic stops.²⁵⁷ For example, approximately 8% or 9% of all police killings in November and December 2016 happened subsequent to a police traffic stop.²⁵⁸ Thus, it seems possible that by contributing to more routine traffic stops of drivers of color, pretextual-stop doctrines may expose these individuals to a greater likelihood of coercive behavior and ultimately police violence.

B. Lack of Options for Redress

Our findings are also concerning because victims of racial profiling during traffic stops currently have few options for redress. In *Whren*, the Court emphasized that targeting a driver for a traffic stop because of their race violates the Equal Protection Clause of the Constitution.²⁵⁹ But as a practical matter, it remains difficult for a victim of such racial profiling to obtain relief. Pretextual stops based on a driver's race may not result in the discovery of contraband. If police uncover no evidence of criminal wrongdoing, then one of

254. *The Counted: People Killed by Police in the US*, GUARDIAN, <https://perma.cc/URF8-2KPE> (archived Jan. 8, 2021) (providing detailed estimates of the number of individuals killed by police in 2015 and 2016 in the United States). The *Guardian* first made the data available in June 2015, but it was updated regularly thereafter to reflect updated estimates of police killing for both 2015 and 2016. *Id.*

255. FATAL ENCOUNTERS, <https://perma.cc/C9C6-GR2E> (archived Jan. 8, 2021) (providing an extensive database on police killings in the United States over many years, and serving as a major source for other similar databases).

256. MAPPING POLICE VIOLENCE, <https://perma.cc/Q3KL-TQR5> (archived Jan. 8, 2021) (collecting and categorizing an extensive amount of data on the number of individuals killed by law enforcement over the years, with a particular focus on the ways that this violence disproportionately affects Black individuals).

257. *The Counted: People Killed by Police in the US*, *supra* note 254 (to locate, navigate to data for 2016).

258. *See id.* We calculated these figures by manually evaluating whether each killing in the *Guardian's* database appeared to be connected with a routine traffic stop, given the descriptions of the circumstances surrounding the killing provided by the website.

259. *Whren v. United States*, 517 U.S. 806, 813 (1996).

the most important deterrents to police misconduct—the exclusionary rule—is of little practical use.²⁶⁰

Additionally, the compensable damage that an individual suffers from a single unlawful traffic stop is often minimal, making it highly unlikely that potential plaintiffs will take advantage of their right to seek civil damages against police officers who violate their constitutional rights under 42 U.S.C. § 1983.²⁶¹ If a victim of racial profiling hopes to use § 1983 to secure injunctive relief rather than civil damages, the Court’s holding in *City of Los Angeles v. Lyons* makes it difficult for the victim to establish standing in federal court because of their inability to demonstrate a likelihood of future harm.²⁶² And, of course, it can be particularly difficult to prove in a civil court or in an internal disciplinary hearing that a police officer was motivated by a driver’s race, creating significant evidentiary issues.²⁶³ Thus, as previous scholars have persuasively argued,²⁶⁴ it is extremely challenging for victims of racial profiling in traffic enforcement to receive relief under the current police regulatory system.²⁶⁵ This realization, alongside our findings of racial

260. See Rachel A. Harmon, *Promoting Civil Rights Through Proactive Policing Reform*, 62 STAN. L. REV. 1, 10-11 (2009) (explaining that “the scope of the exclusionary rule is inevitably much narrower than the scope of illegal police misconduct,” and further explaining that “[m]any kinds of misconduct” by police do not result in the collection of evidence that may be used or excluded from a later criminal proceeding).

261. See *id.* at 9-10 (discussing how “inexpensive settlement[s]” resulting from § 1983 suits may reduce the incentive for departmental reform); Jason Mazzone & Stephen Rushin, *From Selma to Ferguson: The Voting Rights Act as a Blueprint for Police Reform*, 105 CALIF. L. REV. 263, 276 (2017) (“The absence of punitive damages—a remedy designed to deter unlawful behavior—means any resulting judgment (or threat thereof) may be insufficient to alter police practices, even assuming available compensatory damages are sufficient to prompt victims to bring lawsuits in the first place. In essence, in many instances it is not worth the trouble even to initiate the suit.” (footnote omitted)).

262. 461 U.S. 95, 105-06, 109 (1983) (holding that in order to have standing to pursue injunctive relief against the Los Angeles Police Department to ban the use of a chokehold that caused him harm, Mr. Lyons needed to prove a substantial likelihood of future harm—something he couldn’t do in this case, because he was unlikely to be victimized by the same chokehold procedure again in the future).

263. Intuitively, this difficulty stems from the fact that police will often hide their actual discriminatory intent behind seemingly race-neutral explanations. For an example, see the discussion of this problem in the *Floyd* case. *Supra* notes 137-39 and accompanying text.

264. See *supra* notes 65-72 and accompanying text.

265. To be clear, our evidence alone will not make it any easier for a litigant to succeed in these cases. At best, our data merely provide evidence of the disparate impact of police behavior on communities of color, which we believe the Washington Supreme Court facilitated in its holding in *Arreola*. We think this data alone should be sufficient to worry lawmakers and potentially inspire policy change to limit the scope of police authority in making traffic stops. Nevertheless, we cannot prove the intentional racial discrimination on the part of Washington police that would be required to satisfy the standard articulated by the U.S. Supreme Court in *Washington v. Davis* for claims under
footnote continued on next page

inequality pervading pretextual traffic stops, may reinforce the need for states and localities to enact regulations that go beyond those articulated in *Whren*. Additionally, given the lack of available options for aggrieved individuals to seek redress, other actors within the criminal-justice system may consider using their authority to discourage law-enforcement use of pretextual traffic stops. For example, San Francisco District Attorney Chesa Boudin announced in February 2020 that his office intended to end the charging of criminal cases involving contraband obtained during pretextual stops.²⁶⁶

C. Decoupling Criminal Investigations and Traffic Enforcement

Finally, our results may support emerging scholarly calls for the decoupling of criminal investigations and traffic enforcement.²⁶⁷ The pretextual stops that occurred in major cases like *Whren* and *Ladson* happened when police officers tasked with the enforcement of more serious criminal offenses used a technical traffic violation to justify the investigation of a hunch or suspicion. For example, in *Ladson*, Officers Mack and Ziesmer were not actually concerned about whether the driver had an expired registration sticker.²⁶⁸ As members of a local gang-patrol unit, they suspected that the driver was trafficking drugs.²⁶⁹ Similarly, in *Whren*, the officers were patrolling an area known for drug trafficking, seemingly in anticipation of uncovering evidence of drug crimes.²⁷⁰ In each case, the officers were able to

the Equal Protection Clause, which can serve as the basis for § 1983 claims. 426 U.S. 229, 239 (1976); see also *Village of Arlington Heights v. Metro. Hous. Dev. Corp.*, 429 U.S. 252, 267-68 (1977) (establishing factors that courts can consider in evaluating whether sufficient evidence exists to prove an Equal Protection Clause violation). Thus, *Whren*, *Arreola*, and other comparable opinions may contribute to widespread racial profiling that cannot be easily addressed by police accountability mechanisms.

266. See Evan Sernoffsky, *DA Chesa Boudin Sets New Policies on SF Police Stops, Gang Enhancements, Three Strikes*, S.F. CHRON. (updated Feb. 28, 2020, 4:00 AM), <https://perma.cc/W76M-5K5M>.

267. See, e.g., Jordan Blair Woods, *Decriminalization, Police Authority, and Routine Traffic Stops*, 62 UCLA L. REV. 672, 751, 756-59 (2015) [hereinafter Woods, *Decriminalization*] (offering as one possible reform the removal of police officers from the enforcement of traffic laws, particularly decriminalized traffic offenses, and transferring that authority to state actors without traditional police powers); Jordan Blair Woods, *Traffic Without the Police*, 73 STAN. L. REV. (forthcoming May 2021) (manuscript at 4), <https://perma.cc/3ZEN-V3K7> (arguing for the removal of police officers from certain traffic-enforcement responsibilities).

268. *State v. Ladson*, 979 P.2d 833, 836 (Wash. 1999) (en banc) (“The officers do not deny the stop was pretextual.”).

269. *Id.* (“The officers explained they do not make routine traffic stops while on proactive gang patrol although they use traffic infractions as a means to pull over people in order to initiate contact and questioning.”).

270. See *Whren v. United States*, 517 U.S. 806, 808 (1996) (describing the area under patrol as a “high drug area”).

conduct a pretextual stop because the law empowered them *both* to make traffic stops *and* to make arrests for other criminal offenses. Some non-U.S. jurisdictions have experimented with decriminalizing some traffic offenses and transferring traffic enforcement to units whose only responsibility is to enforce the traffic code, not to investigate and respond to criminal acts more broadly.²⁷¹ This kind of decoupling of traffic enforcement from other police work may result in more evenhanded enforcement, and it would presumably eliminate the use of traffic enforcement as a pretext for other criminal investigations. At least one American city—Berkeley, California—has already taken steps to remove traffic enforcement from the purview of its local police department.²⁷²

Opponents of such a proposal may understandably argue that enforcement of traffic laws exposes non-law-enforcement officers to unreasonable risks of physical harm. Policing, they may argue, is a dangerous job, even if an officer is primarily engaged in traffic stops. But compelling new evidence suggests that many may overestimate the risk of injuries to police officers engaged in routine traffic enforcement. Analyzing over 200 law-enforcement agencies in Florida over a ten-year period, Jordan Blair Woods found that the risk of violence in traffic enforcement was extremely low.²⁷³ Roughly one in every 6.5 million routine traffic stops results in the felonious killing of an officer, and one in every 361,111 stops results in an assault causing serious injury.²⁷⁴ This finding suggests, at a minimum, that traffic enforcement may not be so dangerous as to necessitate the involvement of traditional police personnel.²⁷⁵

Alternatively, our findings may strengthen arguments for reducing police discretion in traffic enforcement through the integration of emerging technology. Elizabeth Joh has persuasively argued that traffic-enforcement

271. See Woods, *Decriminalization*, *supra* note 267, at 756 (citing New Zealand as an example of a jurisdiction that experimented with such an approach between 1936 and 1992).

272. Rachel Sandler, *Berkeley Will Become 1st U.S. City to Remove Police from Traffic Stops*, FORBES (updated July 15, 2020, 8:22 PM EDT), <https://perma.cc/Y8AQ-BJ2U>.

273. Jordan Blair Woods, *Policing, Danger Narratives, and Routine Traffic Stops*, 117 MICH. L. REV. 635, 640, 660-62, 668 (2019).

274. *Id.* at 640, 683.

275. While reassigning traffic-enforcement responsibilities might limit the ability of law-enforcement officers to use traffic stops as a crime-fighting tool, and would potentially come at a significant financial cost, it would all but eliminate the current incentive for police officers to use traffic enforcement as a pretext for broader criminal investigations. Steven Maynard-Moody & Michael Musheno, *Social Equities and Inequities in Practice: Street-Level Workers as Agents and Pragmatists*, 72 PUB. ADMIN. REV. S16, S21 (2012) (“[O]ne of the primary and most institutionalized ‘crime-fighting’ tools of modern proactive policing is the investigatory stop of drivers and pedestrians.”).

technologies could eliminate the need for most discretionary traffic stops.²⁷⁶ Whren, she argued, has made challenges to police discretion “[i]mpracticable.”²⁷⁷ Instead, a number of technologies could, in effect, partially supplant ordinary, discretionary traffic enforcement by police officers: red-light cameras,²⁷⁸ speed cameras,²⁷⁹ and automatic license-plate readers.²⁸⁰ As Joh explains, such automated enforcement technologies could more fairly and consistently perform moving-violation enforcement, criminal-record checks, vehicle-defects checks, and drunk-driving enforcement.²⁸¹ And Joh has argued that automated enforcement could avoid police interactions that are “humiliating or discriminatory.”²⁸² Obviously, these traffic-enforcement technologies may still create significant risks of inequality in how they are developed, in the algorithms they employ, in the data they create, and in where they are utilized.²⁸³ Andrew Guthrie Ferguson has written extensively on the risks associated with these types of advanced policing technologies.²⁸⁴ To be

276. Elizabeth E. Joh, *Discretionless Policing: Technology and the Fourth Amendment*, 95 CALIF. L. REV. 199, 216-25 (2007).

277. *Id.* at 212-13.

278. See, e.g., Richard A. Retting, Susan A. Ferguson & A. Shalom Hakkert, *Effects of Red Light Cameras on Violations and Crashes: A Review of the International Literature*, 4 TRAFFIC INJ. PREVENTION 17, 19, 22 (2003) (finding that red-light-camera enforcement results in a possible decrease in violations and a possible decrease in injury crashes, although injury-crash estimates vary from one study to the next).

279. See, e.g., Richard Tay, *Speed Cameras: Improving Safety or Raising Revenue?*, 44 J. TRANSP. ECON. & POL'Y 247, 248-49 (2010) (finding that the installation of speed cameras in Edmonton, Alberta, resulted in reductions in injury crashes, and suggesting that cameras may have been a deterrent to unlawful speeding).

280. See, e.g., Jason Potts, *Research in Brief: Assessing the Effectiveness of Automatic License Plate Readers*, POLICE CHIEF, Mar. 2018, at 14, 14, <https://perma.cc/EZT3-4V98> (describing a study showing that automatic license-plate readers can increase the ability of police to detect stolen cars).

281. See *id.* at 222 tbl.1 (listing these common reasons for police exercising their discretion to make traffic stops and finding that they would be candidates for automated enforcement).

282. *Id.* at 224.

283. See ANDREW GUTHRIE FERGUSON, *THE RISE OF BIG DATA POLICING: SURVEILLANCE, RACE, AND THE FUTURE OF LAW ENFORCEMENT* 3, 5 (2017) (explaining how data-driven policing disproportionately targets and affects marginalized communities).

284. For a broader discussion of the many risks posed by emerging police technologies, see generally *id.* A number of other commentators have also written detailed accounts of the potential for abuse of these emerging police technologies. See, e.g., Bryce Clayton Newell, *Local Law Enforcement Jumps on the Big Data Bandwagon: Automated License Plate Recognition Systems, Information Privacy, and Access to Government Information*, 66 ME. L. REV. 397, 399-400 (2014) (exploring legal and policy divides surrounding, and some of the potential drawbacks of, automated license-plate readers); Joh, *supra* note 276, at 226-33 (describing various objections to automated enforcement of traffic laws via technological tools). The Policing Project has also done extensive research on the need

footnote continued on next page

clear, our study alone does not support a wholesale move from human to technological enforcement of traffic codes. Nevertheless, our data is consistent with the hypothesis that police may abuse the discretion given to them by *Whren* and similar state cases. To the extent that technological enforcement of traffic codes may limit opportunities to exercise such discretion, it is possible that a careful and well-regulated technological-enforcement regime could produce more equitable outcomes.

Conclusion

For decades, scholars have worried that *Whren* would invite racial profiling in routine traffic enforcement. This hypothesis seemed both intuitive and consistent with the large body of literature on the ways that race affects police decisionmaking.²⁸⁵ Nevertheless, the existing body of research has been unable to evaluate this hypothesis empirically. Our study provides strong support for this hypothesis. The judicial authorization of mixed-motive stops in Washington—which closely resemble the kind of pretextual stops at issue in the *Whren* decision—was associated with a statistically significant increase in stops and searches of drivers of color relative to white drivers in the state. Most of this increase occurred during daylight hours, when police could most readily determine the race of drivers. These findings are consistent with scholarly claims that *Whren* and state court equivalents “permit racial bias, either explicit or implicit, to go unchecked and unpunished.”²⁸⁶ Ultimately, these findings should serve as a sobering reminder that legal rules granting police discretion, even if they make “sense . . . from the point of view of judicial administration,”²⁸⁷ may come at the cost of inequality in our justice system.

to balance privacy concerns against the potential benefits of these technological tools. See *Surveillance Technology*, POLICING PROJECT: N.Y.U. SCH. L., <https://perma.cc/6G6J-BFSM> (archived Jan. 8, 2021) (providing numerous resources on these types of technologies).

285. See, e.g., L. Song Richardson, *Implicit Racial Bias and Racial Anxiety: Implications for Stops and Frisks*, 15 OHIO ST. J. CRIM. L. 73, 75-81 (2017) (discussing the existing body of literature showing that implicit bias and racial anxiety affect police behaviors and perceptions of potential suspects).

286. Simmons, *supra* note 59, at 29.

287. Harris, *supra* note 4, at 545.

Appendix

In this Appendix, we explain in more detail many of our methodological choices. We also provide additional details on versions of our models.

A. Modeling Choices

In an empirical undertaking, there are a host of modeling decisions that need be made. In this Part, we lay out the decisions we made and the justifications for them.

1. Reliance on the number of traffic stops by race by county

Our main analysis aggregates the number of traffic stops per month per county. Aggregating data, though, comes at some empirical costs. For instance, if we were looking at homicides in the United States, and we aggregated the count of homicides by state, an additional five homicides reported in a large state like California or Texas would be less significant than an increase of five homicides in a less populous state like Wyoming. One possible empirical approach is to calculate per capita rates by dividing the outcome by the population and then taking the natural logarithm of that rate. The natural-log transformation is a convenient way to help smooth the data without biasing the outcome variable, and it allows the coefficients to be interpreted as semi-elasticities or percent changes. For reasons laid out previously—namely, the lack of any reliable measure of the underlying population or demographics of those on the specific roads policed by the Washington State Patrol²⁸⁸—we chose to leave the data in count form, meaning that each observation is the count of stops per county per month per racial group. While there still might be some smoothing benefits associated with taking the natural log of the count data, doing so would come at the cost of addressing the occurrence of zero outcomes in the dataset (that is, county-month pairs with zero stops for a particular racial group). Mathematically, we cannot take the natural log of a zero, meaning that we would have to deploy a different transformation. While there are alternative transformations, such as the inverse hyperbolic sine transformation, or alternative regression strategies, such as fixed-effects Poisson regressions, we opt for ordinary-least-squares regressions as they are computationally more feasible, rely on fewer modeling assumptions, and produce largely similar results.²⁸⁹

288. See *supra* Part III.B.

289. See Jeffrey M. Wooldridge, *Multiplicative Panel Data Models Without the Strict Exogeneity Assumption*, 13 *ECONOMETRIC THEORY* 667, 669-72 (1997) (discussing the assumptions involved in Poisson regressions).

2. Calculating and clustering of standard errors

Accurately calculating the standard errors of an estimated coefficient is critical to valid statistical inference. As Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan have argued, difference-in-differences estimates may report systematically biased standard errors. To correct this inaccuracy, they proposed a standard-error-clustering technique to adjust for this sort of bias based on the group level where serial correlation occurs, which usually is within the group.²⁹⁰ Further expansion of this strand of research has provided theoretical justification for multiway clustering in situations where serial correlation in the error terms might exist across multiple groupings—in our case, county and year.²⁹¹ Given this context, we cluster the standard errors of all our results by county and by year (though the results are insensitive to virtually any plausible level and combination of clustering we could imagine). We ultimately settled on two-way clustering by county and year because it consistently provided the most conservative estimates.

3. Parallel-trends assumption

Unbiased difference-in-differences estimations require the existence of parallel pretrends. Essentially, we want to ensure that both our groups (here, stops of white and nonwhite drivers) were trending in a similar direction prior to the event in question. Otherwise, any divergence in trends after *Arreola* may simply be a continuation of this underlying difference in existing trends. Researchers typically operationalize the parallel-trends assumption in several ways. The first is a simple visual inspection of the trends in question. That is, the two lines should move roughly together prior to the new law's passage. As stated previously, visual inspection of Figures 1 and 2 suggest that traffic stops of white and nonwhite drivers were generally trending together in a predictably cyclical fashion prior to *Arreola*.

Second, we present a formal test of the parallel-trends assumption, which we report in Figure 8 as part of the event study,²⁹² by creating dummy

290. See generally Marianne Bertrand, Esther Duflo & Sendhil Mullainathan, *How Much Should We Trust Differences-in-Differences Estimates?*, 119 Q.J. ECON. 249, 254-58, 273 (2004) (proposing such an approach).

291. See generally A. Colin Cameron, Jonah B. Gelbach & Douglas L. Miller, *Robust Inference with Multiway Clustering*, 29 J. BUS. & ECON. STAT. 238, 238-39 (2011) (discussing inferences with multiway clustering).

292. In the event study reported in Table A.1, we chose to group the pretrends dummy variables by year. Grouping them by month or quarter might possibly give a more detailed picture of the nuances in pretrends, but doing so comes at the cost of diminished statistical power and extremely noisy results. We opted for the less granular, but also less noisy, grouping by year since it appears to best strike the balance between precision and noise.

variables that capture the time periods prior to the legal change interacted with the racial group in question. As we report, we find no evidence to suggest that trends in traffic stops of white and nonwhite drivers diverged in a statistically significant way prior to *Arreola*. Divergence of these trends appears to occur after *Arreola*.

4. Use of triple-difference regressions and a placebo test

In Part III.E, we employed a veil-of-darkness methodology to examine changes in the number of stops of nonwhite drivers relative to white drivers at night relative to daytime after *Arreola*. In Table 4, we found that after *Arreola*, officers stopped more nonwhite drivers relative to white drivers in the daytime relative to nighttime. This result holds regardless of whether we include or exclude stops where race is unidentified in our definition of nonwhite, as discussed in Part B of the Appendix below. To further validate this finding, we present in Table A.1 the results of a placebo test. It is common in triple-difference regressions to report the results of this sort of a placebo test as evidence that we used an acceptable control not otherwise influenced by the policy change in question.²⁹³ If we used an acceptable control, we would expect to see no statistically significant effect of the policy change on our control group. In this placebo test, we run a similar model to that displayed in Table 4, except this time we focus exclusively on changes in the number of stops of exclusively white drivers in the daytime relative to the nighttime hours before and after *Arreola*.

293. For examples of the context in which this sort of presentation makes sense, see generally Nicolai Brachowicz & Judit Vall Castello, *Is Changing the Minimum Legal Drinking Age an Effective Policy Tool?*, 28 HEALTH ECON. 1483, 1486 & tbl.1, 1487 (2019) (showing a null result for a falsification test in Table 1 as a check on the robustness of their findings); Gruber, *supra* note 24, at 632 & tbl.3 (relying on a similar approach in the presentation of data in Table 3).

Table A.1
Placebo Test for Veil-of-Darkness Methodology,
White Drivers Only (2008-2015)

Triple-difference regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of white drivers	38.229 (19.771)	24.656** (9.036)	36.810 (22.384)
Mean driver age	—	-76.101** (21.474)	-5.231 (5.882)
Officer race	—	-1,889.853 (977.854)	-108.568 (192.929)
Officer gender	—	-1,487.960 (877.977)	130.418 (120.438)
<i>R</i> -squared	0.087	0.236	0.806

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 6,546. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

As seen in Table A.1 above, when we add in fixed effects and controls, we fail to find evidence that *Arreola* contributed to a statistically significant change in the frequency of stops of white drivers at night relative to the daytime.

More generally, we believe that the nature of our model and our underlying research question may increase the validity of our findings. Some have criticized prior studies employing veil-of-darkness methodologies because they are not clean experiments.²⁹⁴ In a typical study employing a veil-of-darkness methodology, researchers sometimes conduct a difference-in-differences regression that compares the change in the number of stops of white drivers at night versus day with the change in the number of stops of nonwhite drivers at night versus day. But this sort of model cannot account for differences in driving behavior at certain times of the day that may correlate with race in some manner. And while it may be harder for a police officer to determine a driver's race at night, it is not always impossible.

An example may better illustrate this concern. Imagine a researcher is trying to determine whether the hypothetical Pleasantville Police Department is engaged in racial profiling by conducting a veil-of-darkness test. Pleasantville's population is almost entirely white. But Pleasantville has an automobile-manufacturing plant that operates twenty-four hours per day. During the daytime, this plant employs mostly Black and Hispanic workers. But at night, this plant employs mostly white employees. This may complicate the ability of a researcher to use a veil-of-darkness methodology to prove that the Pleasantville Police Department is engaged in racial profiling. Any increase in the number of stops of Black and Hispanic drivers during the daytime may be the result of an increased number of Black and Hispanic drivers on the road violating traffic laws as they commute to and from the automobile plant, rather than any sort of profiling on the part of the police department. Additionally, depending on the lighting on the streets in Pleasantville, police may be able to identify the race of passing drivers at night. Because of these limitations, a typical difference-in-differences regression using a veil-of-darkness methodology faces limitations.

Our study is different. Because we exploit a policy change, we are able to sidestep some of the ordinary concerns surrounding veil-of-darkness methodologies. Consider again the hypothetical of Pleasantville. Our triple-

294. For an example of a criticism of veil-of-darkness methodologies, see generally Jesse Kalinowski, Matthew B. Ross & Stephen L. Ross, *Endogenous Driving Behavior in Tests of Racial Profiling in Police Traffic Stops* 42 (Univ. of Conn. Dep't of Econ. Working Paper Series, Paper No. 2017-03R, rev. 2020), <https://perma.cc/M28E-MAMV> (hypothesizing that veil-of-darkness tests fail to account for the fact that minority motorists may adjust their driving behavior to account for the heightened scrutiny they expect to receive from police as a result of their race during daylight hours, and that this may undercut the validity of such tests).

difference methodology conducts the same difference-in-differences regression described above, except we then take the difference of those difference-in-differences before and after the *Arreola* decision. Thus, for our purposes, it does not matter if we fail to account for complicated and racially imbalanced traffic patterns surrounding the hypothetical twenty-four-hour automobile plant in Pleasantville. As long as these idiosyncratic differences in driving habits by race are the same before and after *Arreola*, then they should not pose any significant concerns to the validity of our findings. Similarly, in an ordinary veil-of-darkness methodology, researchers cannot necessarily argue that darkness completely eliminates the ability of police to identify the race of passing drivers. For our purpose however, since we are able to calculate two sets of difference-in-differences, even if we were to still find evidence of racial profiling in the evening, this does not necessarily bias the result. As long as police officers have the same approximate ability to identify the race of passing drivers at night before and after *Arreola*, our findings should be valid.²⁹⁵

5. Additional controls and fixed effects

We are limited in the controls we are able to include in each model given data availability. We do, however, include in each model—measured at the county level—the mean age of each driver stopped, the proportion of stops in any given month and county conducted by a male officer, and the proportion of stops in any given month and county conducted by a white officer. For the individual-level regressions, since we are dealing with individual-level data, these controls are the actual age of the driver and the actual race and gender of the officer.

Additionally, each regression that is marked as such contains county-level fixed effects, or a dummy variable for each county that captures the time invariant, unobserved factors that might be influencing driving as well as traffic and stopping patterns. For instance, more urbanized counties in Washington may have a higher concentration of roads, traffic, stops, and

295. Our setup allows and adjusts for this sort of fuzzy baseline group that might also be influenced by the policy change. This is akin to the difference between a strict and fuzzy regression-discontinuity design. See generally David S. Lee & Thomas Lemieux, *Regression Discontinuity Designs in Economics*, 48 J. ECON. LITERATURE 281 (2010) (providing an overview of the use of regression-discontinuity designs). Additionally, our specific triple-difference design is especially attractive in this context because we are able to account for potential bias in the standard errors that occur in these types of models, as mentioned previously, by multidirectional clustering, including clusters on light categorizations (day, night, and twilight). For the sake of consistency, we report all of the standard errors of our triple-difference regressions clustered at the county by year levels, though the results are insensitive and stronger if we also cluster by light exposure.

nonwhite drivers. County-level fixed effects account for these differences across counties.

B. Alternative Models That Include Stops Where Race Is Unidentified as Stops of Nonwhite Drivers

As discussed in more depth in Part III.B, around 26% of the dataset lacks data on the race of the driver.²⁹⁶ And the Washington State Patrol has previously faced claims that officers purposefully failed to properly document the race of drivers of color.²⁹⁷ Thus, we recognize that it is possible that troopers may systematically fail to document the race of drivers of color so as to avoid accusations of racial bias. If this were the case, then evidence of this racial bias may be masked through an increase in stops where race is unidentified after *Arreola*. As we saw in Figure 2, it appears that when we include stops where race is unidentified in our definition of nonwhite drivers, the uptick in stops of nonwhite drivers relative to white drivers after *Arreola* remains evident.

To account more fully for this possibility, we reproduce here many of the tables featured earlier in this Article, but this time we include stops where race is unidentified in our definition of nonwhite drivers. As seen below, regardless of whether we include stops where race is unidentified in our definition of nonwhite drivers, our results remain substantially the same.

296. See *supra* note 177 and accompanying text.

297. See *supra* notes 179-80 and accompanying text.

An Empirical Assessment of Pretextual Stops and Racial Profiling
73 STAN. L. REV. 637 (2021)

Table A.2
Effect of *Arreola* on Stops of Nonwhite Drivers Relative to White Drivers,
Including Stops Where Race Is Unidentified (2008-2015)

County-level fixed effects	Difference-in-differences regressions		
	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of nonwhite drivers	121.787*** (32.387)	110.216*** (9.049)	119.568*** (34.789)
Mean driver age	—	-16.685*** (4.250)	-2.229 (1.332)
Officer race	—	-1,331.197* (627.205)	198.158 (167.439)
Officer gender	—	-1,349.498* (584.298)	-280.420 (182.223)
<i>R</i> -squared	0.201	0.245	0.527

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 18,746. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.3
Stops of Nonwhite Drivers at Night Relative to Day Post-*Arreola*,
Including Stops Where Race Is Unidentified (2008-2015)

Triple-difference regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Post- <i>Arreola</i> stops of nonwhite drivers at night relative to day, relative to white drivers	-39.313** (15.462)	-42.235*** (7.694)	-39.003** (15.098)
Mean driver age	—	-6.903** (1.872)	-1.236* (0.552)
Percentage of stops by white officers	—	-485.961* (225.441)	59.001 (50.885)
Percentage of stops by male officers	—	-451.086* (207.536)	-76.931 (52.016)
<i>R</i> -squared	0.210	0.236	0.476

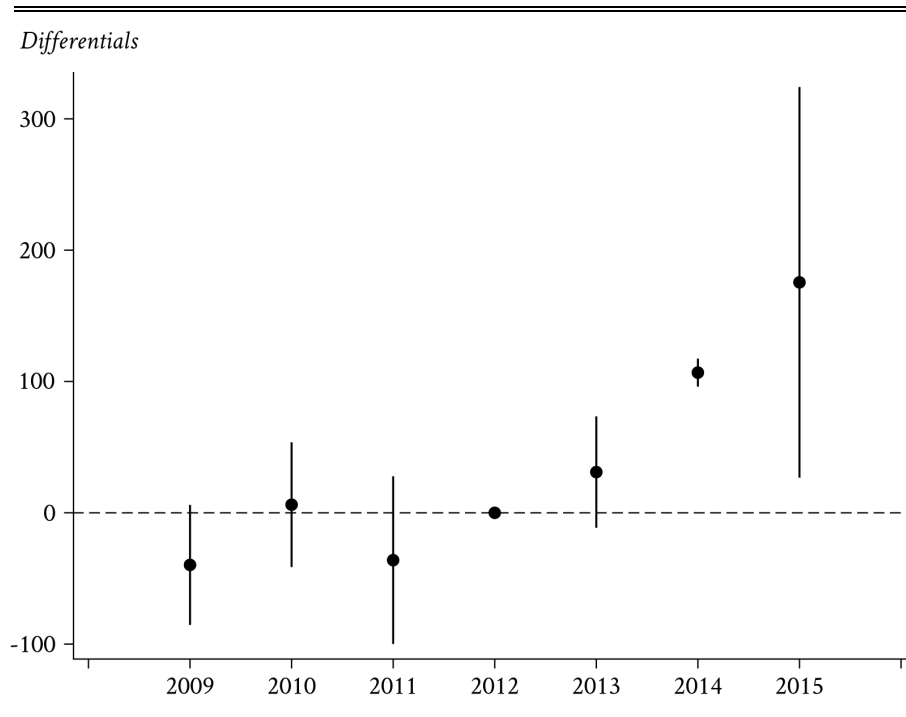
Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. $N = 35,192$. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.4
Effect of *Arreola* on Traffic Stops of Nonwhite Drivers That Led to a Warning,
Including Stops Where Race Is Unidentified (2008-2015)

Triple-difference regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Post- <i>Arreola</i> warnings given to nonwhite drivers relative to white drivers	81.839*** (17.614)	76.674*** (5.163)	79.886** (22.066)
Mean driver age	—	-8.316** (2.390)	-3.593** (1.152)
Percentage of stops by white officers	—	-470.436* (221.049)	9.088 (53.084)
Percentage of stops by male officers	—	-286.205 (160.936)	48.933 (52.216)
<i>R</i> -squared	0.303	0.330	0.503

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 18,746. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Figure A.1
Event Study of Differentials in Stops of White Drivers and Nonwhite Drivers,
Including Stops Where Race Is Unidentified (2009-2015)²⁹⁸



²⁹⁸ As with Figure 8, the year 2012 acts as our comparison in the event study since it was the last full year without treatment. In these event-study-style regressions, we necessarily must use one year as a baseline for comparison. Because we estimate no effect in 2012 it has no confidence bounds. We signify that in Figure A.1 with the dot at zero.

C. Alternative Models Assuming Delayed Effect of Training in Mixed-Motive Stops

As discussed in Part III.A above, it is not clear that *Arreola* resulted in an immediate change in police behavior. Instead, it seems plausible that the decision would result in additional stops of nonwhite drivers only after officers had been trained in the use of mixed-motive stops. In the tables below, we present alternative outputs for our primary model. In Tables A.5 and A.6, we use the date that *Arreola* first appeared in the WSCJTC monthly digest (March 2013) rather than the date of the decision (December 2012). In Table A.5, we *exclude* stops where race is unidentified from our definition of nonwhite; in Table A.6, we *include* stops where race is unidentified in our definition of nonwhite. Alternatively, in Tables A.7 and A.8, we operate under the assumption that the full effect of *Arreola* may have been delayed a full year, to December 2013, when all officers had sufficient opportunity to be exposed to the concept of mixed-motive stops through their annual in-service training requirement. Again, we present these results both excluding (Table A.7) and including (Table A.8) stops where race is unidentified.

Table A.5

Effect of *Arreola* Appearing in the WSCJTC Digest on Stops of Nonwhite Drivers
Relative to White Drivers, Excluding Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of nonwhite drivers	124.679*** (29.157)	110.387*** (9.806)	120.298*** (29.944)
Mean driver age	—	-22.335** (6.380)	-8.718** (2.747)
Officer race	—	-1,098.229* (488.683)	108.906 (132.230)
Officer gender	—	-824.559 (433.671)	75.682 (144.806)
<i>R</i> -squared	0.288	0.326	0.536

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 15,470. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.6

Effect of *Arreola* Appearing in the WSCJTC Digest on Stops of Nonwhite Drivers
Relative to White Drivers, Including Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of nonwhite drivers	127.551*** (22.860)	116.160*** (9.167)	125.342*** (27.102)
Mean driver age	—	-16.686*** (4.245)	-2.229 (1.331)
Officer race	—	-1,331.289* (626.895)	197.945 (167.702)
Officer gender	—	-1,350.621* (585.058)	-280.486 (182.321)
<i>R</i> -squared	0.201	0.245	0.528

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 18,746. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.7
Effect of *Arreola* After Annual Training on Stops of Nonwhite Drivers Relative to
White Drivers, Excluding Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of nonwhite drivers	146.943*** (16.158)	128.364*** (11.517)	138.060*** (18.532)
Mean driver age	—	-22.395** (6.371)	-8.753** (2.734)
Officer race	—	-1,096.589* (487.752)	108.958 (132.824)
Officer gender	—	-832.276 (436.277)	65.751 (144.199)
<i>R</i> -squared	0.288	0.326	0.536

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 15,470. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.8

Effect of *Arreola* After Annual Training on Stops of Nonwhite Drivers Relative to White Drivers, Including Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in stops of nonwhite drivers	153.188*** (9.748)	138.531*** (15.831)	147.570*** (12.475)
Mean driver age	—	-16.741*** (4.242)	-2.255 (1.333)
Officer race	—	-1,329.164* (625.874)	198.810 (168.160)
Officer gender	—	-1,354.505* (587.271)	-285.981 (181.837)
<i>R</i> -squared	0.201	0.245	0.528

Note: Stop data is reported per county per month. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. *N* = 18,746. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

D. Effects of *Arreola* on Searches

Finally, we disaggregate the data and employ the same difference-in-differences technique to measure any possible effect of *Arreola* on the probability of getting searched *after* getting stopped.²⁹⁹ That is, formally:

Model A.1

$$\Pr(\text{Search} | \text{Stop})_{qkt} = \alpha + b_1 \text{nonwhite}_k + b_2 \text{Arreola}_t + b_3 \text{nonwhite} * \text{Arreola}_{qkt} + \theta X + \varepsilon$$

This model predicts the probability of getting searched conditional on getting stopped. Whereas Models 1 and 2 above estimate the change in total stops per county per month, Model A.1 is measured at the individual stop level q in county k at time t . While there may be some concern with the estimation of Model A.1 regarding changes in driving behavior and/or stop practices that do not directly reflect the procedural change that occurred with *Arreola*, reshaping the dataset in this way allows us to control for any changes in driving behavior and/or police stopping behaviors since we are looking at searches of cars that have already been stopped. While Model 1 gives some insight into the overall effect of *Arreola*—that is, the total increase/decrease in stops of nonwhite drivers per county per month—Model A.1 allows us to look more specifically at the probability of getting searched after the stop occurs, which should mitigate changes from unobserved factors like driving behavior.

Tables A.9 and A.10 employ this difference-in-differences modeling to examine the effect of *Arreola* on the probability of nonwhite drivers getting searched relative to white drivers. The results of this model represent the change in the probability of getting searched conditional on being stopped relative to white drivers—so a positive number suggests an increase in the likelihood of getting searched after being stopped relative to white drivers, and a negative number indicates a decrease in the likelihood.

299. Additional models considering the impact of *Arreola* on search outcomes are available upon request. For the sake of brevity, we include only these primary results. Additionally, while there are fewer theoretical reasons why the legalization of recreational marijuana would contribute to increased numbers of traffic stops of particular racial groups, it seems plausible that legalization may have an effect on the frequency of searches and other behavior after traffic stops. Thus, we conduct only a more limited evaluation of this topic in this Article.

Table A.9
Effect of *Arreola* on Probability of Vehicle Searches of Nonwhite Drivers,
Including Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in searches of nonwhite drivers	0.00310*** (0.00038)	0.00298*** (0.00048)	0.00314*** (0.00016)
Driver age	—	-0.00038*** (0.00008)	-0.00035*** (0.00007)
Officer race	—	0.00267 (0.00192)	0.00451 (0.00237)
Officer gender	—	0.00134 (0.00114)	0.00162 (0.00085)
<i>R</i> -squared	0.002	0.003	0.009

Note: Search data is reported on an individual basis. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. $N = 8,257,527$. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.10
Effect of *Arreola* on Probability of Vehicle Searches of Nonwhite Drivers,
Excluding Stops Where Race Is Unidentified (2008-2015)

Difference-in-differences regressions			
County-level fixed effects	<i>No</i>	<i>No</i>	<i>Yes</i>
Change in searches of nonwhite drivers	-0.0059*** (0.0010)	-0.0060*** (0.0011)	-0.0058*** (0.0014)
Driver age	—	0.00065*** (0.00011)	0.00061*** (0.00010)
Officer race	—	0.00459 (0.00258)	0.00686* (0.00311)
Officer gender	—	-0.00054 (0.00136)	-0.00002 (0.00102)
<i>R</i> -squared	0.002	0.005	0.012

Note: Search data is reported on an individual basis. In all but the last row of this table, the first value is the regression coefficient, while the second value (in parentheses) is the standard error of that coefficient. The last row of this table displays the *R*-squared value for each regression. $N = 6,077,266$. Asterisks indicate degrees of significance: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.9 suggests that after *Arreola*, stops of nonwhite drivers may be more likely to result in searches than stops of white drivers. If *Arreola* empowered police to investigate hunches through mixed-motive traffic stops, and if these traffic stops were more likely to target nonwhite drivers, this is consistent with the expected outcome.³⁰⁰ But the results flip when we remove these cases where driver race is unidentified from our analysis, as seen in Table A.10. Thus, on this point, the evidence is somewhat inconclusive. However, as stated previously, we believe including the “n/a” classifications as nonwhite may more accurately capture officers who systematically misidentify (or fail to identify) the race of nonwhite drivers to avoid detection of possible racial profiling.³⁰¹

In total, when we include stops where race is unidentified in the definition of nonwhite in Table A.9, we find that *Arreola* is associated with a 0.3-percentage-point increase in the probability of officers searching the vehicle of a nonwhite driver incident to a traffic stop. While that might not seem like a substantial increase in the likelihood of a search incident to a stop, since only 2.2% of stops resulted in searches, a 0.3-percentage-point increase represents around a 14% increase in the likelihood of getting searched. These results are statistically significant with and without the introduction of controls and fixed effects.

Another way to evaluate this hypothesis is to consider the hit rate of searches of white and nonwhite drivers—that is, the frequency with which a

300. Evidence gathered by prior scholars on how police approach pretextual investigatory stops informs the possible link between pretextual stops and subsequent police searches. Charles R. Epp, Steven Maynard-Moody, and Donald Haider-Markel observed that after the Supreme Court issued its decision in *Whren*, police departments across the country trained officers to use pretextual justifications to conduct so-called “investigatory stops.” EPP ET AL., *supra* note 15, at 36 (describing the institutionalization of the investigatory stop in the 1990s across American police departments). Advocates of these stops argued that they could proactively prevent criminal activity. *Id.* (describing how the International Association of Chiefs of Police “enthusiastically encouraged police departments [across the country] to adopt this practice,” in part because of a belief that it “may be our most effective tool for interdicting criminals” (quoting Earl M. Sweeney, *Traffic Enforcement: New Uses for an Old Tool*, POLICE CHIEF, July 1996, at 45)). A “book-length police training text” by Charles Remsberg, which a leading policing consultant has praised as an authoritative text on the subject, advises officers to follow several steps in employing investigatory stops. *Id.* at 36-37 (citing CHARLES REMSBERG, TACTICS FOR CRIMINAL PATROL: VEHICLE STOPS, DRUG DISCOVERY AND OFFICER SURVIVAL (1995)). First, Remsberg says officers should develop suspicion or curiosity about a driver and identify some legal justification for a traffic stop (often a minor traffic violation). *Id.* at 36. Then, after stopping the driver, the officer should decide whether they can justify a search of the vehicle based on observation and a conversation with the driver. *Id.* at 36-37. When possible, Remsberg also advises officers to seek the consent of the driver to search the vehicle, in hopes of finding evidence of criminal behavior. *See id.* at 37.

301. *See supra* notes 177-81 and accompanying text.

vehicle search leads to the discovery of contraband. To test this hypothesis, we calculate the hit rate of searches across our dataset. We find that prior to *Arreola* searches of vehicles driven by white individuals resulted in the discovery of contraband 19% of the time. By contrast, only 15% of searches of vehicles driven by nonwhite individuals resulted in the discovery of contraband prior to *Arreola*. Admittedly, the use of hit-rate analysis is complicated by the fact that Washington legalized recreational marijuana during the time frame that we study. This may explain the drop in hit rates for vehicle searches for both white and nonwhite drivers. Nonetheless, post-*Arreola* hit rates for nonwhite drivers remained comparatively lower (7%) than hit rates for white drivers (8%).