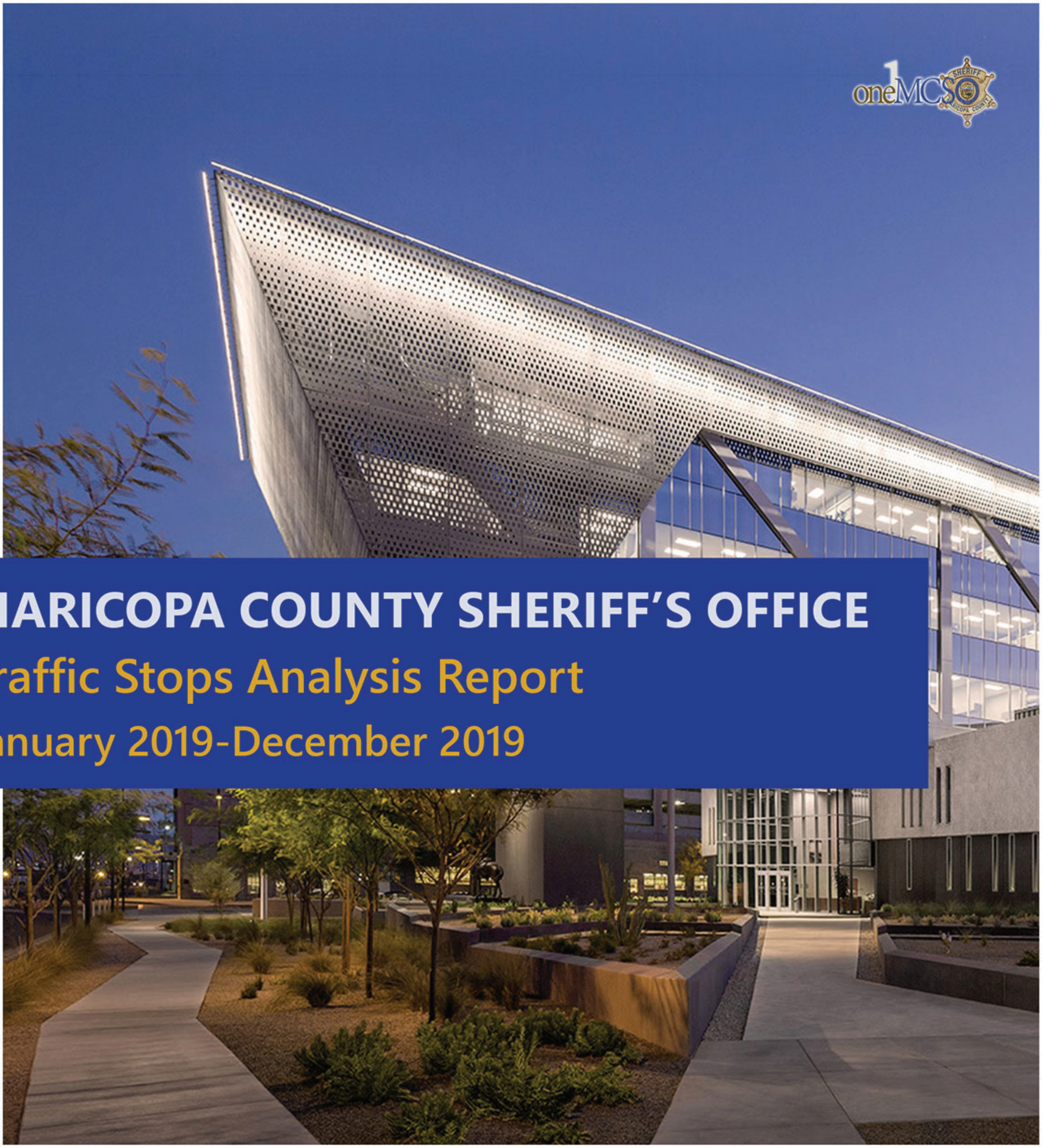


MARICOPA COUNTY SHERIFF'S OFFICE

Traffic Stops Analysis Report

January 2019-December 2019

CNA

This document contains the best opinion of CNA at the time of issue.

CNA Analysis Team:

Zoë Thorkildsen, Project Director; Emma Wohl, Deputy Project Director;
Bridgette Bryson, Analyst; Benjamin Carleton, Analyst; Jennifer Lafferty, Analyst;
Brittany Cunningham, Analyst; James R. Coldren, Jr., Advisor; Hildy Saizow, Advisor

Suggested citation:

Thorkildsen, Z., Bryson, B., Wohl, E., Carleton, B., & Lafferty, J. (2020). *Maricopa County Sheriff's Office Traffic Stops Analysis Report: January 2019–December 2019*. Phoenix, AZ: Maricopa County Sheriff's Office.

Distribution

Distribution unlimited.

May 2020

TABLE OF CONTENTS

TABLE OF CONTENTS

TABLES

FIGURES

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK.

EXECUTIVE SUMMARY[1]

The Maricopa County Sheriff's Office (MCSO), established in 1871, serves and protects the unincorporated areas of Maricopa County, Arizona, and several cities to which the office provides law enforcement services on a contractual basis. Since 2014, the MCSO has worked towards achieving compliance with a federal court order requiring the MCSO to stop its immigration enforcement and refrain from using Hispanic ancestry as a factor in making law enforcement decisions. The MCSO currently operates under two related court orders, respectively titled the First Order and Second Order. As a feature of the First Order, the MCSO must conduct organizational- and individual-level analyses of patrol activity to determine whether racial disparities exist in MCSO traffic stops and outcomes. In November 2016, Sheriff Paul Penzone was elected as Maricopa County Sheriff and took office in January 2017. In 2018, the MCSO contracted with the CNA Institute for Public Research to analyze patrol activity on an annual and monthly basis and support the development of quarterly reports on special topics related to traffic stops. This report examines patterns of patrol activity within the MCSO; it does not analyze or identify individual deputies. The analysis in this report includes traffic stops made by MCSO deputies from the start of January 2019 through the end of December 2019. The MCSO expects to use this report as a knowledge base regarding patrol activity in the office and as a guide for potential interventions, initiatives, and new or revised policies in conjunction with the appointed Monitoring Team and Parties to the Court Orders (namely the Department of Justice and American Civil Liberties Union).

The MCSO uses its Traffic and Criminal Software (TraCS) data system to capture data in the field from traffic stops. Of the 209 variables available through TraCS (which include deprecated legacy variables), we used a subset to analyze racial disparities in stop outcomes and construct data using variables from TraCS and appending data from other MCSO systems. To accurately estimate the differential outcomes from traffic stops based on the race of the driver, we used two statistical approaches across the five relevant outcome variables (stop length, search rates, citation rates, arrest rates, and seizure rates). To analyze the stop length, searches, citations versus warnings, and arrests, we used propensity score matching. Propensity score matching is a quasi-experimental method of statistical comparison that identifies the most similar events in a condition of interest (in this case, Hispanic, Black, or all minority drivers[2]) and White drivers using a propensity score. To analyze seizure rates during searches, we used chi-square testing, which examines whether the racial distribution of searches that result in seizures is different from the racial distribution of searches that do not result in seizures.

Over the 12-month period from January 2019 to December 2019, MCSO deputies performed 23,630 traffic stops. Rates of traffic stops have exhibited an overall upward trend since February of 2019, with a slight decrease in the winter months of 2019. Within the 23,630 traffic stops, deputies perceived 68 percent of drivers as White, 21 percent as Hispanic, and 7 percent as Black. The remaining 4 percent of stops involved other minorities (Asian and Native American). The drivers stopped were 59 percent male and 41 percent female. In the dataset, 86 percent of the stops that deputies made ranged from 5 to 19 minutes. Approximately 52 percent of stops ended with a citation; 47 percent ended with a warning. Just over 3 percent of stops ended with

1 Much of the material from this section is identical to the executive summary from the *Maricopa County Sheriff's Office Traffic Stops Analysis Report: July 2017–December 2018*.

2 The 'all minority drivers' analysis includes Hispanic, Black, Asian, and Native American drivers, compared with White drivers.

an arrest. Less than 0.2 percent of stops resulted in a driver search that was a discretionary decision by the deputy. The seizure rate during non-incidental searches of drivers was 27 percent, and the seizure rate during non-incidental vehicle searches was 12 percent.

The MCSO and CNA's analysis team conclude that there is evidence of disparate outcomes by driver race in traffic stops on most stop outcomes. This is consistent with past studies of traffic stop outcomes in other agencies (as noted in this report's introduction), as well as previous court-ordered traffic stop analyses within the MCSO. Stops involving Hispanic or Black drivers were more likely to be longer and more likely to involve a search than stops involving White drivers. Stops involving Hispanic drivers were more likely to result in citations or arrests compared with stops of White drivers. However, stops involving Black drivers were no more or less likely to end in a citation or arrest than stops involving White drivers, and searches involving minorities were no more or less likely to result in a seizure than searches involving White drivers. These disparities represent potential indicia of bias as described in the Court Order; as a result of these analyses, MSCO will take reasonable steps to investigate and monitor this situation and, where necessary, shall implement interventions. These results are consistent with those from the most recent annual report, Maricopa County Sheriff's Office Traffic Stops Analysis Report: July 2017–December 2018, other than the result for arrests during stops of Black drivers.

The MCSO and the CNA analysis team worked collaboratively to collate the data for this analysis, address missing values and other data irregularities, analyze traffic stops outcomes, and develop the Maricopa County Sheriff's Office Traffic Stops Analysis Report: January 2019–December 2019. The MCSO had primary responsibility for collating data and adjudicating missing values and data irregularities, as well as reviewing the annual report. The CNA analysis team had primary responsibility for developing and executing the analytical plan and authoring the annual report. MCSO had primary responsibility for drawing conclusions from the analytical results.

The MCSO will use this report to better understand its traffic stop activity and better serve the residents of Maricopa County. The MCSO and CNA will continue to work closely to analyze traffic stop activity by MCSO deputies, including developing additional annual analysis reports, monthly analysis reports focused on individual deputies, and quarterly reports on special topics as determined by the MCSO, CNA, and the Monitoring Team in consultation with the Parties.

INTRODUCTION[3]

Background

The Maricopa County Sheriff's Office (MCSO), established in 1871, serves and protects the unincorporated areas of Maricopa County and several cities to which the office provides law enforcement services on a contractual basis. In 2016, the residents of Maricopa County elected Sheriff Paul Penzone to lead the office, which includes more than 3,000 employees and provides enforcement and detention services to the more than four million residents of Maricopa County. The MCSO operates the Fourth Avenue, Durango, Estrella, Lower Buckeye, and Towers jails, as well as smaller temporary holding facilities in district substations. The MCSO provides patrol and investigative operations for the seven districts of the county, which include an array of businesses, residents, and communities. Additionally, the MCSO operates specialized units and teams, such as narcotics investigations, the animal crimes unit, canine teams, and tactical operations.

Since 2014, the MCSO has worked towards achieving compliance with a federal court order entered in 2013, requiring the MCSO to stop its immigration enforcement and refrain from using Hispanic ancestry as a factor in making law enforcement decisions. In *Manuel de Jesus Ortega Melendres v. Arpaio* (now *Manuel de Jesus Ortega Melendres v. Penzone*), a federal judge found that the MCSO violated the rights of Latinos in Maricopa County through racial profiling and a policy of unconstitutionally stopping persons without reasonable suspicion of criminal activity, in violation of their Fourth and Fourteenth Amendment rights. In 2013, Judge G. Snow of US District Court, Arizona, issued the First Supplemental Court Order (First Order) to the Maricopa County Sheriff's Office to address the pattern of disparate treatment of Hispanic community members in Maricopa County. The First Order established actions required for the MCSO to attain compliance, including introducing new analysis, training, and policies and appointing an independent monitor.[4] As a feature of the First Order, the MCSO must conduct organizational- and individual-level analyses of patrol activity to investigate racial disparities in traffic stop outcomes. In 2018, the MCSO contracted the CNA Institute for Public Research to analyze patrol activity on an annual and monthly basis and produce quarterly reports on special topics related to traffic stops.

This report directly responds to the First Order requirement to analyze the MCSO traffic stop activity to determine whether disparate outcomes exist by race of driver. This report continues with the new approach to conducting a traffic stop outcome analysis for the MCSO, established in the most recent annual report. This approach relies on propensity score matching (PSM) to compare stops that had similar characteristics other than the race of the driver. This report examines patterns of patrol activity within the MCSO; it does not analyze or identify individual deputies. The MCSO expects to use this report as a knowledge base of traffic stop activity in the organization and as a guide for potential interventions, initiatives, and new policies in conjunction with the Monitoring Team and Parties.

3 Much of the material from this section is identical to the introduction from the *Maricopa County Sheriff's Office Traffic Stops Analysis Report: July 2017–December 2018*.

4 In 2016, the Court issued the Second Supplemental Court Order (Second Order), establishing additional oversight and reforms for the MCSO. The Second Order does not include actions or requirements related to traffic stops.

Purpose of traffic stop analyses

Analyses of patrol activity are increasingly common across US law enforcement agencies. Law enforcement agencies face heavy scrutiny by the public and the media for concerns of bias and disparate outcomes by race in interactions between the police and community members. The interactions under scrutiny cover a wide variety of activities, including officer-involved shootings, use of force, searches, and traffic stops (see, for example, Correll et al. 2007; Fridell & Lim 2017; Fryer 2016; Ridgeway 2006; Ritter 2017). Although most law enforcement officers do not intentionally practice biased policing, they may exhibit behaviors that appear biased or that result from implicit bias (Marsh 2009; Nix et al. 2017; Spencer, Charbonneau, & Glaser 2016). Even though law enforcement strives for fair treatment, officers may unconsciously treat community members differently (Hall, Hall, & Perry 2016; Helfers 2016; Stroshine & Dunham 2008).

Implicit bias refers to the attitudes or stereotypes that unconsciously affect understanding, actions, and decisions (Staats, Capatoso, Wright, & Contractor 2015). In contrast to implicit bias, *explicit bias* refers to attitudes and beliefs about a person or group on a conscious level (James 2018), such as prejudice. Implicit bias occurs and affects individuals without their awareness or intentional control (Staats et al. 2015). An officer's implicit biases may affect his or her interactions with a driver when making a traffic stop and may affect stop outcomes on an individual level. This issue persists beyond the scope of law enforcement agencies—all people possess implicit biases, and implicit biases occur naturally on a subconscious level (Staats et al. 2015). Awareness of implicit bias gives law enforcement agencies the opportunity to work with organizations and researchers on methods and training to reduce implicit bias and its effects. Researchers can develop methods to identify officers who need implicit bias or other training through quantitative analysis of disparate outcomes.

Over time, methods for identifying evidence of disparate outcomes have evolved. Early research on bias in policing and disparate outcomes relied primarily on correlational and simple comparative methods (Gaines 2006; Novak 2004; Persico & Todd 2006; Rodriguez et al. 2015; Smith & Petrocelli 2001). Researchers now use methods like propensity score matching and weighting to analyze traffic stops and other law enforcement activity outcomes for evidence of racial disparity (Riley et al. 2005; Ridgeway 2006; Tillyer et al. 2010). Methods for assessing disparate outcomes have also evolved to incorporate measures beyond stop rates, focusing on stop outcomes such as citations and searches (Fridell 2004; Fridell 2005; Tillyer et al. 2010). Researchers also use more sophisticated benchmarks, moving away from population as an external benchmark for assessing disparate outcomes (Grogger & Ridgeway 2006; Lange, Johnson, & Voas 2007).

Understanding the expectations and limitations of quantitative analysis for investigating implicit bias is important. Research on traffic stops includes both pre-stop and post-stop analysis. Pre-stop analysis studies whether the race of the driver affects stop rates; post-stop analysis studies whether the race of the driver affects the outcome of a stop. The different limitations of these two analyses illustrate the difficulties of traffic stop analysis. A pre-stop analysis requires estimating the local driving population, which is a complex problem. Using census data is imprecise, since it includes non-drivers and may not accurately reflect the driving population or the racial distribution of drivers who violate traffic laws (McMahon, Garner, Davis, & Kraus 2002; Tregle, Nix, & Alpert 2019). Other methods for estimating the racial distribution of the driving population, such as observing and recording the race of drivers in a given jurisdiction over time or using driver license race data, can be cost-prohibitive or infeasible due to data unavailability (Fridell 2004; Tillyer et al. 2010).

Conducting post-stop analysis mitigates some of these issues because the population under study is contained within the traffic stop data and does not need to be estimated (Withrow, Dailey, & Jackson 2008; Ridgeway & MacDonald 2010). Despite improvements in analytical methods, correct and in-depth traffic stop data from agencies is still necessary to accurately measure disparate outcomes; the absence of this data can limit the scope and effectiveness of the results. Some agencies track data for their traffic stops meticulously, while other agencies may track only limited information, such as when a stop occurred, the driver's race, and limited stop outcomes, or they may store data about traffic stops across data systems that cannot be readily linked.

Based in part on the limitations of traffic stop analysis, the presence of disparate outcomes does not necessarily indicate the presence of bias. Practitioners and consumers of bias research should understand that disparate outcomes do not definitively indicate bias (Fridell 2004; Simoiu, Corbett-Davies, & Goel 2017). Quantitative analysis cannot capture all the possible reasons that could explain the disparate outcomes. Even with these limitations, the results from statistical analysis can provide better insight into policing practices in an agency and serve as a useful system for identifying disparate outcomes for action by the agency. Such a system provides agencies with a tool to review officer traffic stop conduct and determine the necessary actions, if any, for officers and agencies as a whole.

A greater number of law enforcement agencies now analyze their traffic stop data internally or in partnership with researchers and analysts. The majority of analyses conducted to date find racial disparity in traffic stop outcomes. Tillyer et al. (2010) states, "Analyses of these data demonstrate a relatively consistent trend of racial/ethnic disparities in vehicle stops and vehicle outcomes." The majority of existing studies have shown evidence of racially disparate rates of stops or outcomes of patrol activity in law enforcement agencies (Norris, Fielding, Kemp, & Fielding 1992; Smith & Petrocelli 2001; Engel & Calnon 2004; Novak 2004; Rojek, Rosenfeld, & Decker 2004; Gaines 2006; Weiss & Rosenbaum 2006; Gelman, Fagan, & Kiss 2012; Rosenfeld, Rojek, & Decker 2012; Tillyer & Engel 2013; Baumgartner, Epp, & Shoub 2018; Ariel & Tankebe 2018; Rodriguez, Richardson, Thorkildsen et al. 2019). A few studies have documented findings of no racial disparity in traffic stops (Groggery & Ridgeway 2006; Higgins, Vito, Grossi & Vito 2012; Taniguchi et al. 2016). The balance of the evidence suggests that disparate outcomes during traffic stop activity is common in law enforcement agencies in the United States; however, the prevalence of the problem does not imply that agencies should not pragmatically and proactively address disparate outcomes by promoting anti-bias policy, training, and practices.

Researchers have analyzed patrol activity in many ways. Recently, the Stanford Computational Policy Lab (Pierson et al. 2019) compiled a dataset of 100 million traffic stops from municipal and state agencies. To date, this project is the largest traffic stop study to investigate racial disparities in outcomes. The study used a "veil of darkness" method to compare stop rates by race in situations in which officers presumably could see the race of the driver versus those (in conditions of darkness) in which officers presumably could not. The study found evidence of disparate outcomes in traffic stops across the compiled agencies when controlling for time of day. Several law enforcement agencies have taken on the task of analyzing their patrol activity data and developing a plan to reduce racial disparities, including the Cincinnati Police Department (Ridgeway 2009), Durham Police Department (Taniguchi et al. 2016), Minneapolis Police Department (Ritter & Bael 2005), and New York Police Department (Ridgeway 2007). Overall, the use of statistical analysis for identifying racial disparities in traffic stops is increasingly crucial, and previous analyses indicate that disparities exist across the nation.

Organization of this report

This report is organized into four sections: introduction, approach, findings, and conclusion. The approach section explains the MCSO and CNA's methods for analyzing traffic stop outcomes and developing this report. The findings section details results of the traffic stop analysis on the selected outcomes. Finally, the conclusion section reviews the significance of the analytical findings and discusses future analyses that the MCSO and CNA will conduct in response to the Order. The appendices provide a reference list and list of abbreviations.

Additionally, we provide supplemental appendices to this report in a separate companion document, including supporting data tables, alternate propensity score matching models, and analytical support and robustness checks. Law enforcement researchers and analytical practitioners looking to implement similar studies in other agencies will likely find these appendices of interest.

APPROACH

In this section, we discuss the data, variables, and methodology we used in the traffic stops analysis. We begin by describing the MCSO Traffic and Criminal Software (TraCS) data system, defining the variables used in the analysis, and describing the data cleaning process prior to analysis. We then discuss the propensity score matching approach used for assessing racial disparity in the length of stops, searches, citations, and arrests, as well as the chi-square analysis we used to assess racial disparity in seizure rates. We discuss the alternate specifications we used for the propensity score matching analyses. We close by noting specific considerations for interpreting the findings from this analysis, as well as limitations of the approach.

Overview of data and variables

The MCSO uses TraCS to capture data from the field about traffic stops. TraCS is a data collection, records management, and reporting software for public safety professionals. Deputies use TraCS to document aspects of traffic stops, including driver and vehicle characteristics and activities that occur during the stop. TraCS captures the start time, end time, and geolocation for the traffic stop.[5] The system also requires the deputy to enter variables such as the perceived race of the driver,[6] the contact conclusion, and search and seizure information. TraCS also includes data fields capturing information about technical issues or language barriers, and it includes a comment field for deputies to input additional information.[7] After the deputy fills out the event on TraCS, the system forwards entries for supervisory review and, if necessary, revision.

Of the 209 variables available through TraCS (including deprecated legacy variables), we used a subset to analyze racial disparities in stop outcomes, as well as construct and append data using variables present in TraCS and other MCSO systems. Here, we briefly describe the variables used in the analysis and those constructed by the analysis team. For all categorical variables coded into a single variable (such as stop classification or perceived race of the driver), we constructed indicator variables for each category.

Data about the stop. We used the stop date, stop start time, and stop end time variables to develop descriptive information about stops conducted by the MCSO. We also used the start time and end time to construct the stop length variable, which codes how long a stop lasted in minutes from reported start to finish. We also used stop time to construct an indicator variable capturing stops occurring between 8:00 p.m. and 8:00 a.m. as a proxy for time of day used as a matching variable.[8] Stop classification summarizes the reason for the stop, per the Arizona Revised Statutes (ARS), classified into four categories: criminal, civil traffic, criminal traffic, and petty. Deputies can also indicate whether circumstances beyond their control extended the length of a stop, including technical issues (e.g., a printer failure), a language barrier, a DUI stop, training, or calling for a tow. We also include a variable capturing information about the deputy's assignment (based on call sign), broken out as normal patrol, Lake Patrol, off-duty assignment, designated traffic stop car, supervisor, and other.

5 In some patrol areas, particularly within Lake Patrol's jurisdiction, GPS coverage can be inconsistent. In these cases, TraCS may not automatically capture the GPS coordinates of the stop. We discuss this issue further in the section on missing data.

6 Note that Arizona does not collect information about race as part of its driver's license system; thus, all race categories within the TraCS data are based on the perception of the deputy who made the stop.

7 A detailed description of the TraCS data collection system and included variables is available in MCSO policy #EB-2, "Traffic Stop Data Collection," available publicly on the MCSO website: https://www.mcso.org/Policy/policies.

8 The use of time of day as a matching variable is complicated by the cyclical nature of time variables, in which 23:59 is closer to 00:01 than it is to 23:00, which cannot be readily captured using any continuous variable construction.

Data about stop outcomes. Stop conclusion data describe the outcome of the stop as a citation, warning, long-form, or incidental contact. Long-form is used in cases of a non-booked arrest.[9] TraCS indicates whether a stop included a search of the driver or vehicle (passenger searches are omitted from this analysis as our focus is on drivers) and whether that search was incident to arrest or towing. We constructed a variable for analyzing searches that indicates whether a search of the driver or vehicle took place. For this analysis, we restricted our interest in searches to those that are non-incidental (i.e., discretionary) in nature. For example, policy dictates that all individuals be searched prior to arrest detentions and all vehicles be inventoried prior to tow; searches that occur incident to arrests or tows are not discretionary and thus were excluded from our analysis of outcomes. Deputies also indicate in TraCS whether or not a search resulted in the seizure of contraband.

Data about the driver. We used the post-stop perceived race of the driver to classify driver race as Asian, Black, Hispanic, Native American, or White. We also used the post-stop perceived sex of the driver to create an indicator variable for male drivers (with female drivers and unknown sex drivers collapsed as the comparison category). We also include the reported license plate of the vehicle the driver was operating, classifying it as in-state or out-of-state.

The CNA analysis team appended data not housed with TraCS into our analysis, including information about special assignments. The MCSO manually compiled data about special assignments by deputy and by date. Special assignments during the timeframe of this analysis included DUI task forces, aggressive driver enforcement, and bicycle race support. The analysis team also constructed a deputy productivity variable equal to the number of stops the deputy made over the 12-month period, for descriptive purposes.

9 The state of Arizona allows law enforcement personnel to perform both custodial bookings as well as citations in lieu of detention (e.g., non-booked arrests).

Data verification and missing data

The analysis team reviewed the 2019 TraCS data for data quality (e.g., missing data or out-of-range values) and verification. We identified missing data in several fields. As noted previously, geolocation data should automatically be added to each TraCS entry, but it can be missing if the stop was made in an area without sufficient GPS coverage. The analysis team identified 236 stop data entries with missing latitude and longitude coordinates. The MCSO used the coded location for these stops to impute the latitude and longitude for all entries.

The analysis team identified additional missing data that the MCSO could not adjudicate or impute. Three stops were missing data for the vehicle license plate; we omitted these from all analyses, since in-state plate status is used as a propensity score matching variable. Two stops were missing either start or end times, and therefore we could not calculate stop lengths; these stops are omitted only from the stop length analysis. Taken together, the missing data represent less than 0.1 percent of the overall data, well below any standard thresholds for concerns about missing data biasing analysis or findings. Supplemental Appendix 1 describes missing data by variable.

To prepare the final dataset for analysis, in addition to constructing variables as noted above, the analysis team removed non-traffic stop data and dropped duplicate stop entries (TraCS creates duplicate lines to capture data for multiple contacted passengers; since this analysis focuses solely on drivers, these lines represent duplicate data). We removed all stops indicated as field interviews (FIs) in the contact conclusion variable since this meant the stop did not end in an arrest, citation, or warning and is not relevant for this analysis. We then identified duplicate entries based on the event number, deputy's badge number, and driver's first name and last name, and we removed all entries identified as duplicates based on these criteria. In addition, in the review of missing data entries, the MCSO identified 5 lines of data mistakenly classified as traffic stops (e.g., assisting a disabled motorist) and 44 lines of data where no enforcement actions took place and the driver was free to go (e.g., stops pre-empted by priority calls for service).

Methodology

To most accurately estimate differential outcomes from traffic stops based on the race of the driver, we used two statistical approaches across the five outcome variables under consideration. To analyze length of stops, searches, citations versus warnings, and arrests, we used propensity score matching. To analyze seizure rates during searches, we used chi-square testing. We discuss each of these approaches in more detail below.

Propensity score matching is a quasi-experimental method of statistical comparison. Researchers use quasi-experimental methods in circumstances in which random assignment (i.e., experimental approaches) are not feasible or practical; these techniques leverage specific data structure and statistical techniques to approximate experimental conditions (Shadish, Cook, & Campbell 2002). In this case, propensity score matching matches individual events (in this case, traffic stops) with similar events based on their characteristics (listed at the end of this paragraph). Specifically, propensity score matching identifies the most similar events in or not in a condition of interest (in this case, Hispanic, Black, or all minority drivers[10]) using a propensity score (Rosenbaum & Rubin 1983; Apel & Sweeten 2010). For this traffic stops analysis, we used a logistic regression in the first stage of propensity score matching to determine the probability that a stop involved a driver of a particular race (Hispanic, Black, and all minorities). For all analyses, stops involving White drivers are the comparison conditions. We performed matching analyses using observed characteristics of the stop—namely whether the stop was conducted as part of a special assignment, the driver's sex, the stop longitude and latitude, whether the stop took place between 8:00 a.m. and 8:00 p.m., the stop classification (operationalized as civil traffic stops versus all others), whether the vehicle had out-of-state plates, whether the deputy indicated the stop was extended for one of the five reasons discussed above, and the call sign category the deputy was operating under. The call sign classification variable is a new matching variable as of this annual report. In addition, for the length of stop analysis only, we include whether the stop involved an arrest or a search; both these circumstances necessarily result in longer stops. The addition of these variables for the stop length analysis is also new as of this annual report.

After this matching step, we conducted comparisons using the propensity scores to match observations. For the baseline analysis presented in the main body of this report, we used nearest neighbor matching (in which stops in the condition of interest are compared by propensity score with the nearest one stop that is not in the condition of interest). We chose nearest neighbor matching as the baseline case because it is the least susceptible to problems with achieving common support (Caliendo & Kopeinig 2005), a necessary condition for validating propensity score matching results. Supplemental Appendix 6 describes common support and results from common support testing in more detail. To check the robustness of our results, we ran each analysis using radius matching (in which stops in the condition of interest are compared with all stops within a certain propensity score range that are not in the condition of interest) using multiple radii values. Finally, we also used nearest N-neighbor matching (in which stops in the condition of interest are compared with the nearest N stops by propensity score that are not in the condition of interest). We also considered matching with and without replacement as a sensitivity check. Supplemental Appendix 6 presents detailed results from the robustness check analyses.

For all analyses, we present findings in terms of the average treatment effect—that is, the average difference of outcomes between stops in and not in the condition of interest (Rosenfeld, Rojek, &

10 The 'all minority drivers' analysis includes Hispanic, Black, Asian, and Native American drivers, compared with White drivers.

Decker 2012). We report the average treatment effect, reflecting the difference between outcomes in stops involving Hispanic, Black, or all minority community members versus White community members. We report the average treatment effect in lieu of average treatment on the treated, since average treatment on the treated is appropriate when individuals can choose their assignment into the condition of interest, which is not the case for minority status. For all propensity score analyses, we conducted standard checks of balance and common support. We summarize these results in the body of the report and present them in detail in Supplemental Appendix 4.

We analyzed the rate of seizures during searches using a standard chi-square test of homogeneity across mutually exclusive categories (in this case, race). This test uncovers whether rates of seizures vary significantly across racial categories. As noted in the literature, different rates of seizures may indicate racial bias, since it suggests deputies may use different decision criteria or thresholds prior to searches of minority and non-minority drivers (Persico & Todd 2006; Ridgeway & MacDonald 2009; Walker 2003; Simoiu et al. 2017). For this analysis, we considered only searches that were not incident to arrest or towing. We used a standard chi-square analysis with Pearson's and likelihood ratio tests (Pearson 1900). We also ran Fisher's exact test (due to the small number of stops of Asian and Native American drivers) for comparison purposes.

Alternate specifications

As noted above, we varied the propensity score approach for the propensity score matching analyses to encompass two matching methods: radius and neighbor. We also varied the parameters used for the radius caliper size and the number of neighbors matched. Finally, we considered the effect of allowing replacement (i.e., whether an observation can be used as a match for multiple other observations) for nearest neighbor and radius matching.[11] The Supplemental Appendices to this report present the results from the alternate specifications.

For the length of stop analysis, we also considered an alternate specification in which we added controls for extended stop indicators to estimate the average treatment on the treated, as well as a model in which we used the extended stop indicators as matching variables. We also considered models limited to those stops with and without the extended stop indicators for comparison purposes and because deputies self-select the extended stop indicators in the TraCS form. We anticipate the MCSO will further explore the extended stop indicators in future analyses with support from the analysis team. Including control variables in the second stage of the propensity score matching analysis is feasible only when nearest neighbor matching is used; therefore, we present only findings from that specification for these alternate specifications.

Finally, to allow for direct comparison of the findings in this report to the previous annual report, we also include analyses using all other stops as the comparison condition (e.g., all Hispanic stops are compared with all non-Hispanic stops), instead of stops of White drivers. The results from this analysis are presented in Supplemental Appendix 2.

Considerations and limitations

Propensity score matching represents a substantial improvement over past methods of estimating racial disparity in law enforcement activities, since it does not rely on the development of imperfect or cost-prohibitive external benchmark data and it more precisely estimates the true differences in outcomes when accounting for differences in circumstance between interactions (e.g., traffic stops). However, the methodology is not without limitations. First, as noted above, the matching step relies on the estimation of a logistic regression, which requires estimates to converge over iterative analysis steps. This can limit the inclusion of variables and observations if convergence is impossible for a given model specification. The

11 Matching without replacement cannot be feasibly conducted on N-to-1 neighbor matching analyses.

model also cannot account for any variable that perfectly predicts the condition of interest, though this did not occur in any of the estimated models in this analysis.

Finally, as with all statistical techniques to assess outcomes and behavior from law enforcement personnel, the results from these analyses can uncover only evidence of disparities in outcomes based on race—they cannot provide insight into the underlying causes of these disparities on their own.

THIS PAGE INTENTIONALLY LEFT BLANK.

FINDINGS

In this section, we begin by describing the included variables. As part of the descriptive statistics, we present the rates of traffic stops by race of driver. The analysis team worked closely with the MCSO to assess various options for external benchmarks to use as a comparison condition for stop rates by race. Most existing or proposed external benchmarks provide inaccurate estimates of the driving population (census population) or are cost-prohibitive (collection of data on driver race using observations at intersections). We considered several emerging practices (comparison of daytime versus nighttime stop rates, use of accident data, comparison of criminal versus civil traffic stop rates), but we could not implement them using the currently available data from the MCSO. Therefore, for stop rates, we present descriptive statistics only.

Below, we present the findings from the comparative propensity score matching and chi-square test of homogeneity. For each stop outcome analyzed using propensity score matching, we include results from comparing Hispanic drivers to White drivers, comparing Black drivers to White drivers, and comparing all minority drivers to White drivers. We did not specifically analyze Asian or Native American drivers because of the relative sparsity of stops involving drivers of these races. The chi-square analysis includes drivers of all races.

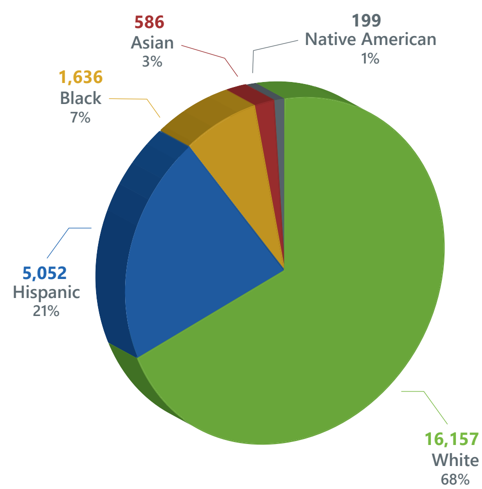
Descriptive statistics

In this section, we describe the data included in this analysis of traffic stops conducted by the MCSO between January 2019 and December 2019 (a 12-month period). We present the characteristics of the stops themselves, characteristics of stop outcomes, and the productivity of the deputies making the stops. Supplemental Appendix 1 provides a full table of descriptive statistics for each variable.

Driver characteristics

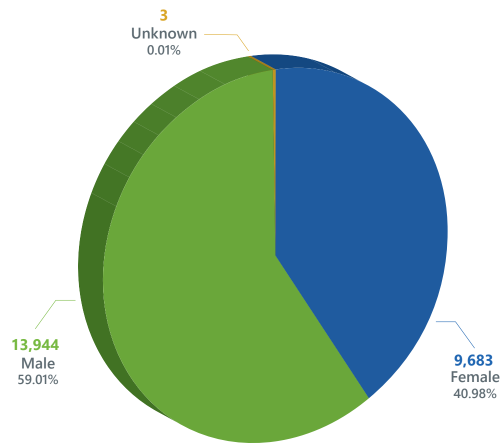
When deputies make a traffic stop, they document their observation of the perceived race of the driver both pre- and post-stop in TraCS. We omitted analysis of the pre-stop perception of driver race, since this variable takes the value “unknown” in almost 96 percent of stops. Post-stop, deputies perceived 68 percent of drivers as White, 21 percent as Hispanic, and 7 percent as Black. The remaining 4 percent of stops were of Native American and Asian drivers.

Figure 1. Stops by post-stop perceived driver race



The deputies also enter post-stop perceived sex in TraCS. The drivers stopped were 59 percent male and 41 percent female, plus three stops (less than one percent) in which the deputy could not determine the sex of the driver.

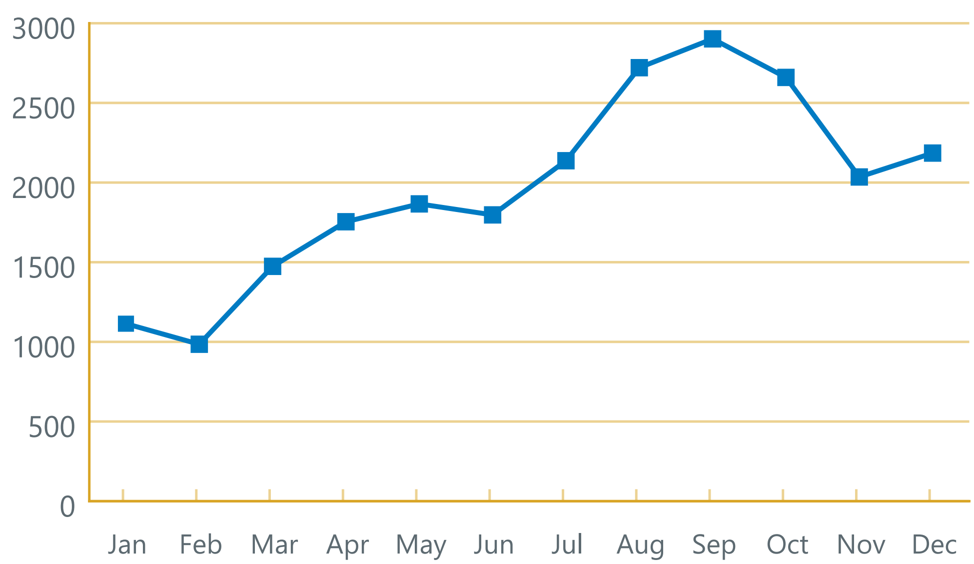
Figure 2. Stops by post-stop perceived driver sex



Stop characteristics

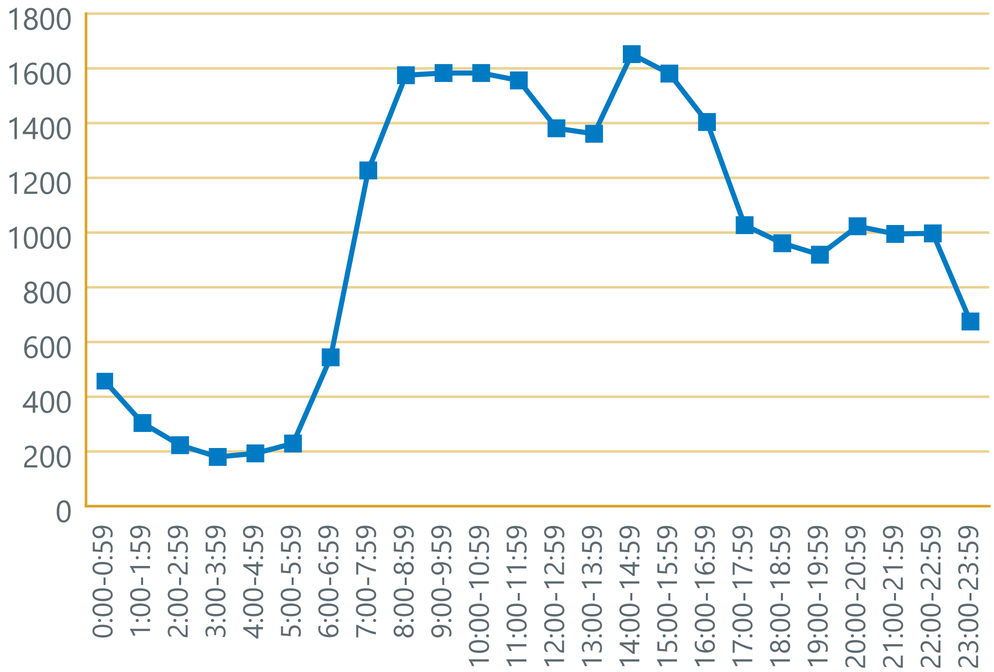
Over the 12-month period for this analysis, the MCSO deputies performed 23,630 traffic stops. Traffic stops over this period exhibit an upward trend, with temporary decreases in February 2019 and in June 2019, and finally again from October 2019 through November 2019. This overall upward trend continues a pattern observed in late 2018.

Figure 3. Stops by month, January 2019–December 2019



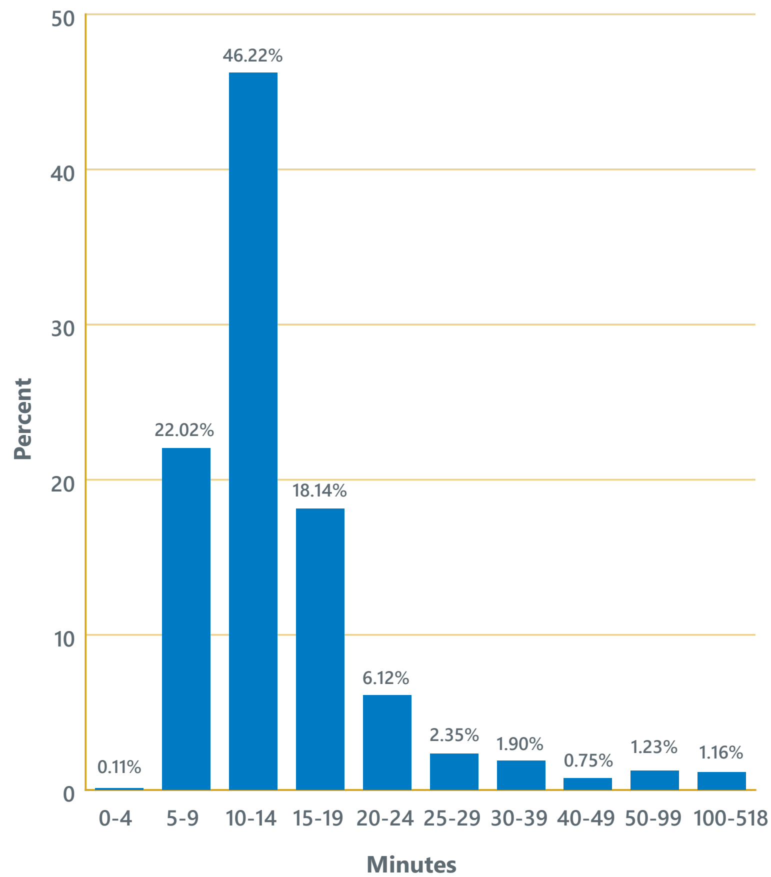
We also consider the time of day that a stop took place. A majority of the stops occurred between 7:00 a.m. and 5:00 p.m.

Figure 4. Stops by time of day



Stop length is of particular importance to this analysis, since it is a core aspect of the Order. Stops lasted an average of 16.1 minutes, a decrease from the previous annual report (in which the average stop length was 19.4 minutes). The majority of stops lasted between 5 and 25 minutes.

Figure 5. Stop lengths in minutes



Deputies document in TraCS whether a stop is extended for reasons beyond their control. The extended stops field contains five options: DUI stop, language barrier, technical issues, training stop, and vehicle towed. Deputies selected extended stop indicators for 3,200 stops, representing 13.5 percent of total stops. This represents a decrease from the previous annual report, which seems to be driven primarily by a decrease in stops involving technical issues and tows. Technical issues occurred in 6.45 percent of stops, while training stops were close behind, at 4.47 percent. Based on observations during traffic stop ride-alongs and body-worn camera footage review while conducting a Traffic Stop Quarterly Report, technical issues often involved equipment failures in the deputy's vehicle, such as printer failures or automated license and registration barcode scanner failures.

Table 1. Extended stop reasons

Reasons indicated	Percentage of stops
DUI stop	1.95%
Language barrier	1.08%
Technical issues	6.45%
Training stop	4.47%
Vehicle towed	1.45%

We also considered which stops occurred while the deputy was on special patrol assignment. Of the 23,630 stops in the dataset, 523 stops occurred while deputies were on DUI Task Force assignments, the most common special assignment in the traffic stop data. Stops conducted while deputies were assigned to the aggressive driver enforcement were the next most common, followed by the Prickly Pedal Bike Race.

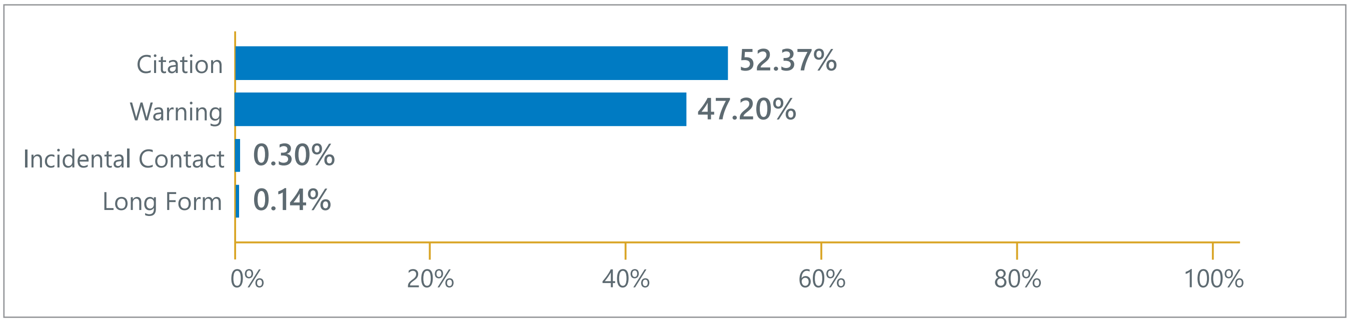
Table 2. Stops conducted during special assignments

Special assignment	Percentage of stops
DUI Task Force	2.21%
Aggressive driver	0.05%
Bike race	0.01%

Stop outcomes

Contact conclusion documents the outcomes from each stop. Of the stops, 52 percent concluded with a citation, 47 percent ended with a warning, and less than 1 percent ended with a long-form submission or incidental contact. The incidental contact refers to situations in which a deputy makes incidental contact with the driver or other occupant of the vehicle and that person does not receive a warning, citation, or long-form charges. A long form is used when the officer submits charges but the person is not booked into jail. They may be charged if the county attorney or judge pursue the charges submitted in the long-form complaint.

Figure 6. Traffic stop contact conclusions



The MCSO organizes stops into four categories, based on ARS code: civil traffic, criminal traffic, criminal, and petty. The majority of stops are civil traffic stops. Civil traffic violations include violations in which the driver pays a fine and does not face jail time. Examples of these include speeding, equipment violation, or seatbelt violations. Criminal traffic violations are traffic violations that result in a fine and involve possible jail time. These include criminal speeding, reckless driving, driving under the influence, or driving on a revoked or cancelled license. Petty violations are criminal violations with less severe penalties that do not include the possibility of jail time. These include boating violations, park violations, and curfew violations. Criminal violations are non-traffic violations that involve possible jail time and typically are incident to the traffic stop, such as stopping an individual with an active warrant for criminal activity or identifying criminal activity not related to the stop. Almost 98 percent of stops have a civil traffic classification; 2 percent have a criminal traffic classification. The dataset contains 40 criminal classifications and 3 petty classifications, accounting for less than 0.25 percent of the stop classification reasons.

Figure 7. Traffic stop classifications

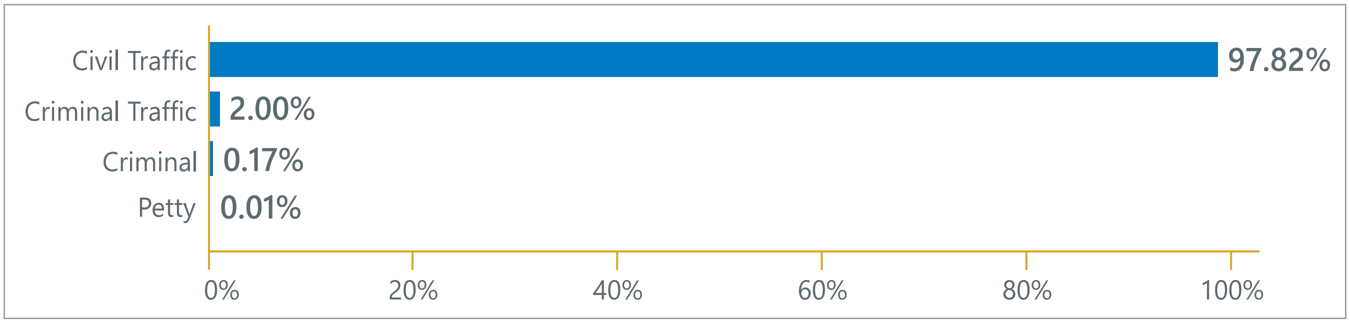
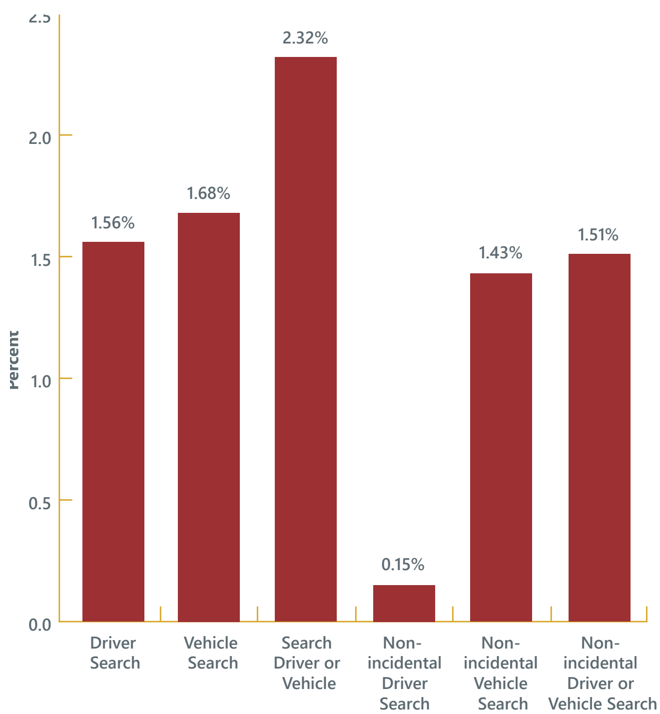


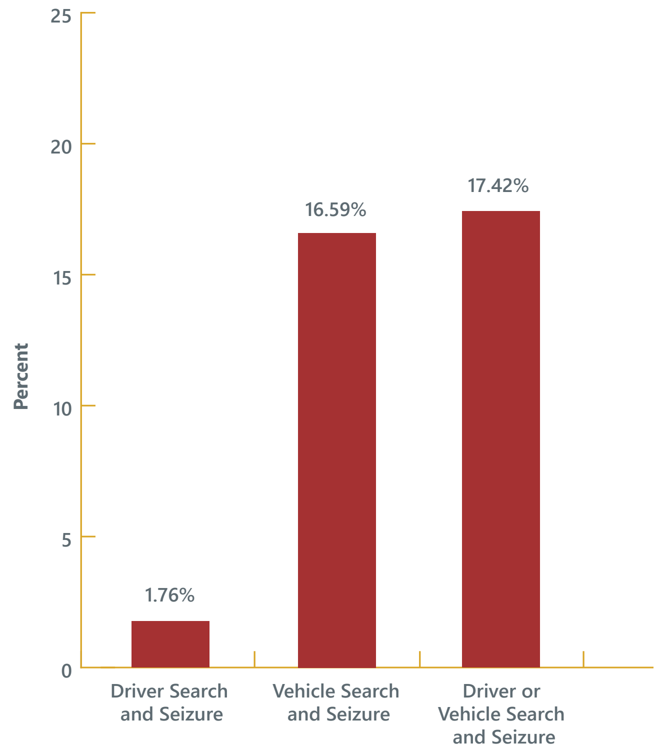
Figure 8 presents information about searches. MCSO policy dictates that deputies search all arrested drivers and search all towed vehicles; these searches are not discretionary on the part of the deputy. Non-incidental searches refer to searches that are not connected to arrests or tows; these are discretionary searches conducted by deputies. As Figure 8 shows, the majority of searches of drivers occurred incident to arrest. For this analysis, we considered searches of drivers or vehicles as a search outcome; there were more searches conducted by MCSO deputies of vehicles than searches of drivers.

Figure 8. Searches



For all stops involving a search, deputies record whether the search turned up contraband (i.e., the incidence of seizures predicated on searches). Overall, 28 percent of non-incidental searches result in seizures, and these seizures are proportionately distributed across the searches of drivers and vehicles.

Figure 9. Seizures during non-incidental searches



Deputies use the driver arrest variable field to document whether arrests are classified as cite and release (i.e., citation in lieu of detention) or bookings. Depending on the charges against the driver, deputies can use their discretion to choose between the two options. For example, a deputy arresting an individual for driving under the influence may use his or her discretion regarding whether the individual is too impaired to be released on his or her own recognizance and should be booked for the night. Arrests of drivers are rare among traffic stops, representing 3 percent of total traffic stops.

Figure 10. Arrests during traffic stops

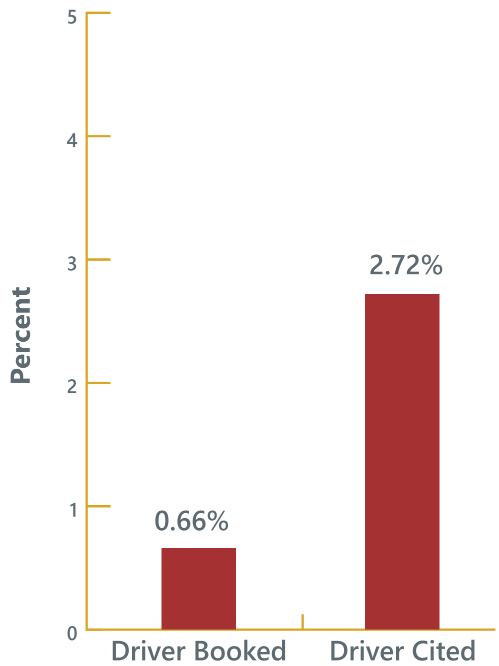
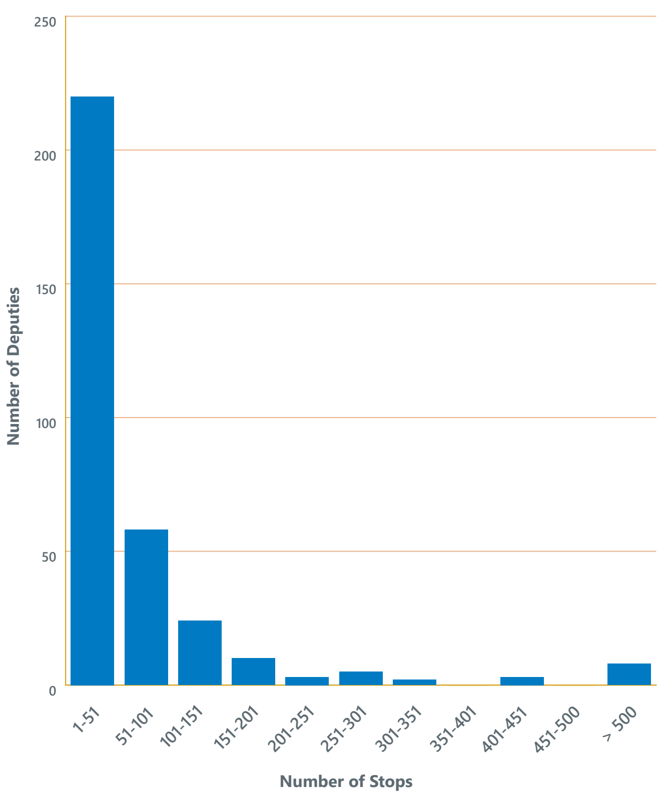


Figure 11. Deputy productivity (number of stops over the 12-month period)



Deputy characteristics

The dataset includes 333 deputies from the MCSO. We present data about deputy productivity, measured as the total number of stops conducted by deputies over the 12-month period in this analysis. As Figure 11 shows, most deputies conducted between 1 and 51 stops during this period, but a notable minority of deputies made over 500 stops in the same period.

Comparative analysis

In this section, we present the findings from analyzing each stop outcome and summarize the findings from the statistical analysis. Supplemental Appendix 2 includes results from the alternate specification using all other stops as the comparison condition. Supplemental Appendix 3 includes results from the logistic regressions for each of the conditions of interest. Supplemental Appendix 4 includes detailed tables of the propensity score matching results. Supplemental Appendix 5 provides results from the analyses of stop length that include extended stop indicators. Supplemental Appendix 6 provides results from the other alternate specifications, and Supplemental Appendix 7 provides details on the results of the common support and balance tests for each specification. Note that we present the full analysis of seizures predicated on searches in the main body of the report.

For the propensity score matching results, we used a p-value of 0.05 or less to indicate significance. Given that the sample size for all analyses was more than 100, this resulted in a critical t-statistic of 1.96 (t-statistics above this value indicate significance, and those below indicate a failure to reject the null hypothesis of no statistically significant difference).

Common support and balance assumptions were met for all the baseline analyses (see Supplemental Appendix 7 for further details on these tests). In propensity score matching analysis, common support is assumed for valid estimation, meaning that all observations contain a positive probability of being in the condition of interest or not, based on the probability score (p-score) (Khandker, Koolwal, & Samad 2010). Balance evaluates the effectiveness of the matching procedure in reducing observable differences between observations within and out of the condition of interest (Khandker, Koolwal, & Samad 2010). After matching takes places, the differences between observations in the condition of interest and their matches on the observable characteristics used for matching should be minimal.

Analyses presented in this section include all observations unless otherwise noted.

Stop length

The analysis team investigated differences in stop length between Hispanic and White drivers, Black and White drivers, and minority and White drivers. To provide context and a comparison point, the average stop length for stops of White drivers in 2019 was 14.87 minutes (or 14 minutes and 52 seconds). Table 3 summarizes the findings from this analysis. **Our analysis found statistically significant differences in stop lengths for all the comparisons (with all t-statistics exceeding 1.96), with all differences falling between 1 and 2 minutes.** These findings are consistent with those from the previous annual report in terms of statistical significance, though all observed differences are smaller. The findings were consistent across all alternate specifications of the main propensity score matching model with replacement.[12] Note that all stop length analyses omit the two stops for which we could not calculate stop length.

Table 3. Propensity score matching results for stop length

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic v. White drivers	1.91	5.24	Yes
Black v. White drivers	1.55	2.67	Yes
Minority v. White drivers	1.38	4.61	Yes

We considered alternate specifications for the analysis of stop length; we used the extended stop indicators to modify the propensity score matching model in three ways. As noted above, deputies can indicate whether they experienced specific circumstances that extended the length of a stop beyond their control, which include technical issues (e.g., a printer failure), a language barrier, a DUI stop, training, or calling for a tow. We considered three mechanisms of controlling for the stop length indicators. First, we introduced the stop length indicators in the second stage matching analysis as control variables. Since we had to conduct the alternate specification analysis manually after calculating the propensity scores, we could not compare the observed difference due to driver race directly with the baseline analyses. The observed difference in this estimation represents the treatment on the treated; in the baseline propensity score matching analysis, the average treatment on the treated and average treatment effect always fell within 2 minutes of each other. Since the results were not directly comparable, we focused instead on consistency or inconsistency in statistical significance. Table 4 presents the results from this analysis. In this specification, the differences observed are slightly smaller than without the extended stop indicators included, but remain statistically significant. As can be seen in the detailed tables in Supplemental Appendix 5, the extended stop indicators all had a large impact on stop length (both in estimated effect per the coefficient and in terms of being highly statistically significant).

12 Models without replacement are less stable because of the likelihood of matching less similar events; we therefore comment in the report on consistency among only the models with replacement. Details on results from the models without replacement can be found in Supplemental Appendix 6.

Table 4. Propensity score matching results for stop length, controlling for extended stop indicators

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic v. White drivers	1.09	5.12	Yes
Black v. White drivers	1.38	3.43	Yes
Minority v. White drivers	0.77	3.7	Yes

As a second test of the impact of the extended stop length indicator variables, we introduced those variables as matching variables. However, we omit language barrier, as it is not appropriate to include variables in the matching step that could be caused by the condition of interest (i.e., the race of the driver). Table 5 presents the results from this analysis; the differences here are average treatment effect and are thus directly comparable to those in the baseline analysis. In this model, all of the differences are statistically significant, and again the observed differences are similar to those in the original specification.

Table 5. Propensity score matching results for stop length, including extended stop indicators as matching variables

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic v. White drivers	1.59	4.21	Yes
Black v. White drivers	1.68	2.84	Yes
Minority v. White drivers	1.22	4.11	Yes

Lastly, we ran separate propensity score matching analyses for stops with an extended stop reason and stops without an extended stop reason. Table 6 and Table 7 present the results from these analyses, respectively. For stops with an extended stop reason, length did not differ significantly by race. Considering stops that were not noted as extended, the differences observed are all statistically significant, but all are smaller than in the baseline model.

Table 6. Propensity score matching results for stop length, including only stops noted as extended

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic v. White drivers	1.10	0.49	No
Black v. White drivers	-2.99	-1.20	No
Minority v. White drivers	1.89	0.92	No

Table 7. Propensity score matching results for stop length, including only stops not noted as extended

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic v. White drivers	0.91	4.76	Yes
Black v. White drivers	1.28	3.88	Yes
Minority v. White drivers	0.94	5.66	Yes

Taken together, the results for the stop length outcomes suggest that the extended stop indicators played a role in understanding the length of stops that MCSO deputies conducted. The MCSO and CNA expect to further investigate the use and effect of extended stop indicators in a future quarterly report.

Citations

The analysis team investigated differences in citation rates (i.e., the percentage of stops that involved citations, as opposed to warnings or long forms) between Hispanic and White drivers, Black and White drivers, and minority and White drivers. To provide context and a comparison point, 52 percent of stops involving White drivers end in a citation. Table 8 summarizes the findings from this analysis. **Hispanic drivers and minority drivers were more likely to receive citations, as opposed to warnings or other stop outcomes, than White drivers. Black drivers, however, did not experience statistically significant differences in citation rates compared with White drivers.** These findings are consistent with those from the previous annual report in terms of statistical significance, though the observed differences are larger in this analysis. The findings were consistent across all alternate specifications of the main propensity score matching model with replacement.

Table 8. Propensity score matching results for citations

Model	Difference (percentage)	t-statistic	Statistically significant?
Hispanic v. White drivers	4 percent	4.42	Yes
Black v. White drivers	3 percent	1.8	No
Minority v. White drivers	4 percent	4.82	Yes

Searches

The analysis team investigated differences in search rates (i.e., the percentage of stops that involved searches not incident to arrest or tow) between Hispanic and White drivers, Black and White drivers, and minority and White drivers. To provide context and a comparison point, 1 percent of stops involving White drivers involve a search. Table 9 summarizes the findings from this analysis. **Search rates had statistically significant differences for all the comparisons, ranging from a difference of 1 percent for Black and minority drivers compared with White drivers to 2 percent for Hispanic drivers compared with White drivers.** These findings are consistent with those from the previous annual report in terms of statistical significance, though the observed differences are slightly larger in this analysis. The findings are consistent across all alternate specifications of the main propensity score matching model with replacement.

Table 9. Propensity score matching results for searches

Model	Difference (percentage)	t-statistic	Statistically significant?
Hispanic v. White drivers	2 percent	5.51	Yes
Black v. White drivers	1 percent	2.31	Yes
Minority v. White drivers	1 percent	6.12	Yes

Arrests

The analysis team investigated differences in arrest rates (i.e., the percentage of stops that involved arrests) between Hispanic and White drivers, Black and White drivers, and minority and White drivers. To provide context and a comparison point, 3 percent of stops involving White drivers end in an arrest. (Across all drivers, booked arrests account for 19 percent of all arrests.) Table 10 summarizes the findings from this analysis. **We found statistically significant differences in arrest rates for Hispanics and minority drivers overall, but not between Black drivers and White drivers.** These findings are consistent with those from the previous annual report in terms of statistical significance, except for the result for Black drivers, and are consistent in the size of the observed differences. The findings were consistent across all alternate specifications of the main propensity score matching model with replacement.

Table 10. Propensity score matching results for arrests

Model	Difference (percentage)	t-statistic	Statistically significant?
Hispanic v. White drivers	1 percent	3.06	Yes
Black v. White drivers	1 percent	1.2	No
Minority v. White drivers	1 percent	4.14	Yes

Seizures

The analysis team investigated differences in seizure rates, predicated on non-incidental searches, by the race of the driver. Deputies made 356 stops involving non-incidental searches during the analysis period. Table 11 presents the breakdown of searches with and without seizures by the race of the driver.[13] The chi-squared test of homogeneity returned $\chi^2=4.07$, $p=0.396$, and the Fisher's exact test returned $p=0.391$, indicating **no statistically significant difference in the distributions of searches with and without seizures across driver race.** These findings are consistent with those of the previous annual report.

13 Note that only one search involving an Asian driver was conducted.

Table 11. Seizures during non-incidental searches by race of driver

Race of driver	Percentage of searches without seizures	Percentage of searches with seizures
Asian	100%	0%
Black	59%	41%
Hispanic	75%	25%
Native American	75%	25%
White	71%	29%
Overall	72%	28%

CONCLUSION

The MCSO and CNA's analysis team conclude that there is evidence of disparate outcomes by driver race in traffic stops. This finding is consistent with past studies of traffic stop outcomes in other agencies (as noted in this report's introduction), as well as with previous traffic stop analyses within the MCSO under the Court Order. In particular, stops involving Hispanic or Black drivers were more likely to be longer and were more likely to involve a search than stops involving White drivers. Stops involving Hispanic drivers were more likely to result in citations or arrests compared with stops of White drivers. However, stops involving Black drivers were no more or less likely to end in a citation or arrest than stops involving White drivers, and searches involving minorities were no more or less likely to result in a seizure than searches involving White drivers. Analysis also suggests that the indicators for extended stop reasons may explain some of the differences in stop lengths, a potential area for further inquiry by the MCSO and the analysis team. Taken together, we identified disparities in many, but not all, stop outcomes. These disparities may indicate a systemic problem. These disparities represent potential indicia of bias as described in the Court Order; as a result of these analyses, MSCO will take reasonable steps to investigate and monitor this situation and, where necessary, shall implement interventions.

The MCSO takes the findings of disparity presented in this report seriously, as disparities in traffic stop outcomes are a potential indicator of implicit bias. Accordingly, the MCSO will proceed with training, policy, and practice to reduce bias. The MCSO remains committed to this charge and will continue to implement necessary changes to address bias—implicit and otherwise—in the organization and in traffic stops in particular. As required by the Court Order, MCSO will take reasonable steps to investigate and closely monitor the situation based on these results. The MCSO will use the analyses in this report and other forthcoming analyses to better understand traffic stop behavior and better serve the residents of Maricopa County. The information in this report will be foundational to the MCSO's efforts to implement data-driven approaches to improving the effectiveness and fairness of patrol activity. Additionally, this analysis places the MCSO at the forefront of comprehensive, in-depth studies of traffic stop activity in US law enforcement.

The MCSO and CNA will continue to work closely to analyze traffic stop activity by MCSO deputies. This work will include developing additional annual analysis reports, monthly analysis reports analyzing individual deputies, and quarterly reports on special topics selected by the MCSO, CNA, and the Monitoring Team, in consultation with the Parties.

THIS PAGE INTENTIONALLY LEFT BLANK.

APPENDIX A. References

Apel, R. J., & Sweeten, G. (2010). Propensity score matching in criminology and criminal justice. *Handbook of Quantitative Criminology, 10*, 543-562.

Ariel, B., & Tankebe, J. (2018). Racial stratification and multiple outcomes in police stops and searches. *Policing and Society*, *28*(5), 507-525.

Baumgartner, F. R., Epp, D. A., & Shoub, K. (2018). *Suspect citizens: What 20 million traffic stops tell us about policing and race*. New York, NY: Cambridge University Press.

Caliendo, M., & Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. *Discussion Paper No. 1588*.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, *92*(6), 1006.

Engel, R. S., & Calnon, J. M. (2004). Examining the influence of drivers' characteristics during traffic stops with police: Results from a national survey. *Justice Quarterly, 21*(1), 49-90.

Fridell, L. A. (2004). By the numbers: A guide for analyzing race data from vehicle stops. Washington, DC: Police Executive Research Forum.

Fridell, L. (2005). Understanding race data from vehicle stops: A stakeholder's guide. Police Executive Research Forum.

Fridell, L., & Lim, H. (2016). Assessing the racial aspects of police force using the implicit- and counter-bias perspectives. *Journal of Criminal Justice, 44*, 36-48.

Fryer Jr, R. G. (2016). An empirical analysis of racial differences in police use of force (No. w22399). National Bureau of Economic Research.

Gaines, L. K. (2006). An analysis of traffic stop data in Riverside, California. *Police Quarterly, 9*(2), 210-233.

Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the New York City police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association, 102*(479), 813-823.

Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association, 101*(475), 878-887.

Hall, A. V., Hall, E. V., & Perry, J. L. (2016). Black and blue: Exploring racial bias and law enforcement in the killings of unarmed black male civilians. *American Psychologist, 71*(3), 175.

Helfers, R. C. (2016). Ethnic disparities in the issuance of multiple traffic citations to motorists in a southern suburban police agency. *Journal of Ethnicity in Criminal Justice, 14*(3), 213-229.

Higgins, G. E., Vito, G. F., Grossi, E. L., & Vito, A. G. (2012). Searches and traffic stops: Racial profiling and capriciousness. *Journal of Ethnicity in Criminal Justice, 10*(3), 163-179.

James, L. (2018). The stability of implicit racial bias in police officers. *Police Quarterly, 21*(1), 30-52.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2010). *Handbook on Impact Evaluation*. Washington, DC: The World Bank.

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

Marsh, S. (2009). The lens of implicit bias. *Juvenile and Family Justice Today*, *18*, 16-19.

McMahon, J., Garner, J., Davis, R., & Kraus, A. (2002). How to correctly collect and analyze racial profiling data: Your reputation depends on it! In *Final Project Report for Racial Profiling Data Collection and Analysis*.

Nix, J., Campbell, B. A., Byers, E. H., & Alpert, G. P. (2017). A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy*, *16*(1), 309-340.

Norris, C., Fielding, N., Kemp, C., & Fielding, J. (1992). Black and blue: An analysis of the influence of race on being stopped by the police. *British Journal of Sociology*, 207-224.

Novak, K. J. (2004). Disparity and racial profiling in traffic enforcement. *Police Quarterly, 7*(1), 65-96.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157-175.

Persico, N., & Todd, P. (2006). Generalising the hit rates test for racial bias in law enforcement, with an application to vehicle searches in Wichita. *The Economic Journal*, *116*(515), F351-F367.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Goel, S. (2019). A large-scale analysis of racial disparities in police stops across the United States. Stanford Computational Policy Lab.

Ridgeway, G. (2006). Assessing the effect of race bias in post-vehicle stop outcomes using propensity scores. *Journal of Quantitative Criminology, 22*, 1-29.

Ridgeway, G. (2007). Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices. Rand.

Ridgeway, G. (2009). Cincinnati Police Department traffic stops: Applying RAND's framework to analyze racial disparities. Rand.

Ridgeway, G., & MacDonald, J. M. (2009). Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association, 104*(486), 661-668.

Ridgeway, G., & MacDonald, J. (2010). Methods for assessing racially biased policing. In S. K. Rice & M. D. White (Eds.), *Race, Ethnicity, and Policing: New and Essential Readings* (pp. 180-204). New York, NY: New York University Press.

Riley, K. J., Turner, S., MacDonald, J., Ridgeway, G., Schell, T., Wilson, J., Dixon, T. L., Fain, T., & Barnes-Proby, D. (2005). Police-community relations in Cincinnati. Rand.

Ritter, J. A. (2017). How do police use race in traffic stops and searches? Tests based on observability of race. *Journal of Economic Behavior and Organization*, *135*, 82–98.

Ritter, J. A., & Bael, D. (2005). Detecting racial profiling in Minneapolis traffic stops: A new approach. *Economic Review, 95*, 94-98.

Rodriguez, D., Kunard, L., Johnson, W., LaRochelle, J., & Thorkildsen, Z. (2015). *Assessment Report on the Fayetteville (North Carolina) Police Department*. Washington, DC.

Rodriguez, D., Richardson, K., Thorkildsen, Z., Monroe, R., Medlock, H., Rickman, S. (2019). *Final Report: Racial Bias Audit of the Charleston, South Carolina, Police Department*. Arlington, VA: CNA.

Rojek, J., Rosenfeld, R., & Decker, S. (2004). The influence of driver's race on traffic stops in Missouri. *Police Quarterly, 7*(1), 126-147.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Rosenfeld, R., Rojek, J., & Decker, S. (2012). Age matters: Race differences in police searches of young and older male drivers. *Journal of Research in Crime and Delinquency, 49*(1), 31-55.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth Cengage Learning.

Simoiu, C., Corbett-Davies, S., & Goel, S. (2017). The problem of infra-marginality in outcome tests for discrimination. *The Annals of Applied Statistics*, *11*(3), 1193-1216.

Smith, M. R., & Petrocelli, M. (2001). Racial profiling? A multivariate analysis of police traffic stop data. *Police Quarterly, 4*(1), 4-27.

Spencer, K. B., Charbonneau, A. K., & Glaser, J. (2016). Implicit bias and policing. *Social and Personality Psychology Compass*, 10(1), 50-63.

Staats, C., Capatosto, K., Wright, R. A., & Contractor, D. (2015). State of the science: Implicit bias review 2015. Kirwan Institute for the Study of Race and Ethnicity, The Ohio State University.

Stroshine, M., Alpert, G., & Dunham, R. (2008). The influence of "working rules" on police suspicion and discretionary decision making. *Police Quarterly, 11*(3), 315-337.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016). Exploring racial disproportionality in traffic stops conducted by the Durham Police Department. Research Triangle Park, NC: RTI International.

Taniguchi, T., Hendrix, J., Aagaard, B., Strom, K., Levin-Rector, A., & Zimmer, S. (2016). A test of racial disproportionality in traffic stops conducted by the Fayetteville Police Department. Research Triangle Park, NC: RTI International.

Tillyer, R., & Engel, R. S. (2013). The impact of drivers' race, gender, and age during traffic stops: Assessing interaction terms and the social conditioning model. *Crime & Delinquency, 59*(3), 369-395.

Tillyer, R., Engel, R. S., & Cherkauskas, J. C. (2010). Best practices in vehicle stop data collection and analysis. *Policing: An International Journal of Police Strategies & Management, 33*(1): 69-92.

Tregle, B., Nix, J., & Alpert, G. P. (2019). Disparity does not mean bias: making sense of observed racial disparities in fatal officer-involved shootings with multiple benchmarks. *Journal of Crime and Justice, 42*(1), 18-31.

Walker, S. (2003). Internal benchmarking for traffic stop data: An early intervention system approach. Police Executive Research Forum.

Weiss, A., & Rosenbaum, D. P. (2006). Illinois traffic stops statistics study: 2005 annual report. Evanston, IL: Northwestern University Center for Public Safety.

Withrow, B. L., Dailey, J. D., & Jackson, H. (2008). The utility of an internal benchmarking strategy in racial profiling surveillance. *Justice Research and Policy*, *10*(2), 19-47.

THIS PAGE INTENTIONALLY LEFT BLANK.

APPENDIX B. Acronyms

Acronym	Definition
ARS	Arizona Revised Statutes
DRE	Drug recognition expert
FI	Field interview
MCSO	Maricopa County Sheriff's Office
PSM	Propensity score matching
TraCS	Traffic and Criminal Software

THIS PAGE INTENTIONALLY LEFT BLANK.

3003 Washington Boulevard, Arlington, VA 22201

www.cna.org | 703-824-2000

CNA

oneMCSO