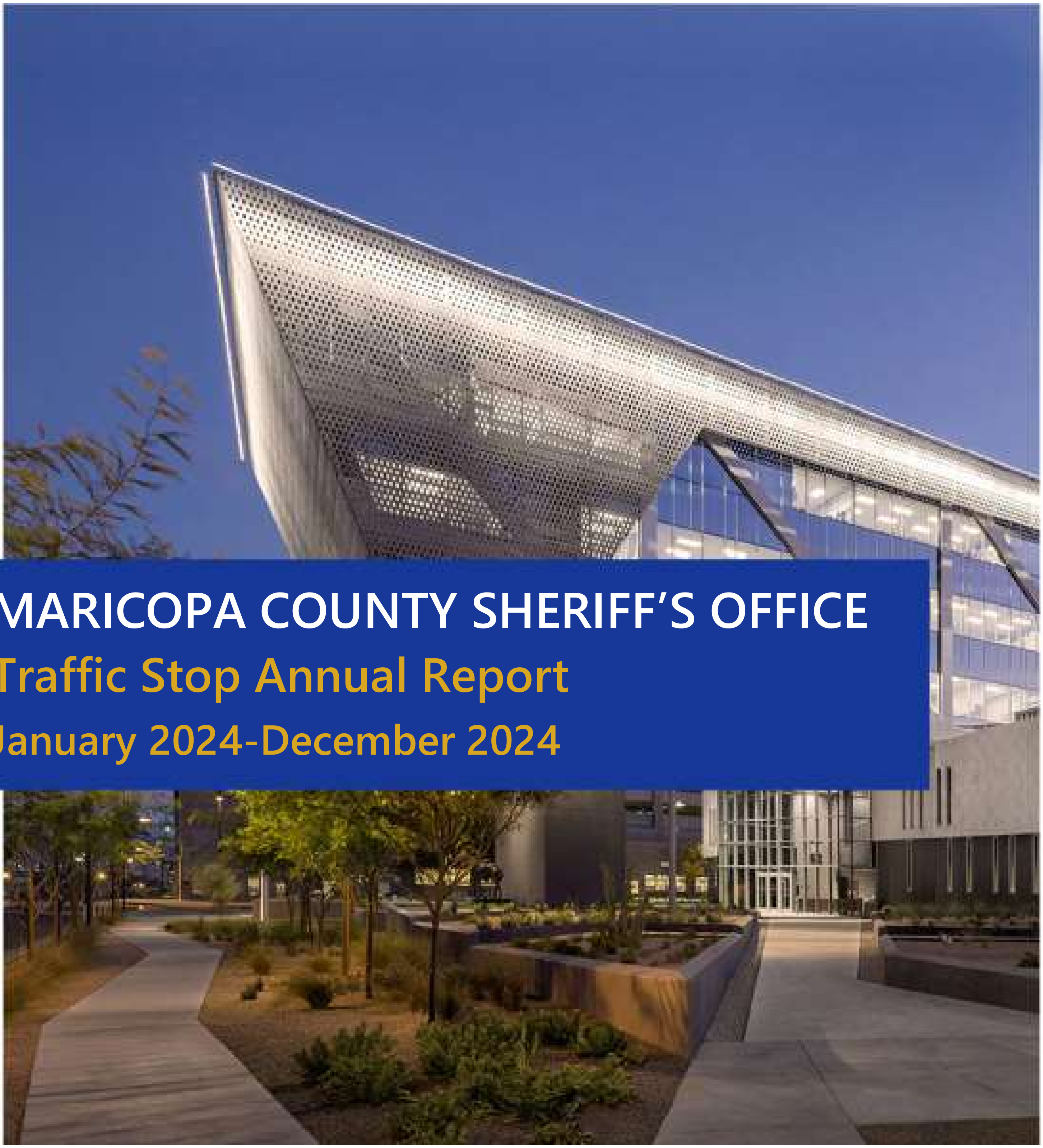


MARICOPA COUNTY SHERIFF'S OFFICE

Traffic Stop Annual Report

January 2024-December 2024

CNA

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

CNA Analysis Team:

Jennifer Lafferty, Project Director (May 2025-present); Kenneth Novak, Co-Project Director; Bryce Peterson, Analyst; L. Cait Kanewske, Analyst; Zoë Thorkildsen, Project Director (January 2025-April 2025)

Suggested citation:

Lafferty, J., Peterson, B., Kanewske, L. C., Novak, K., & Thorkildsen, Z. (2024). Maricopa County Sheriff's Office Traffic Stop Annual Report: January 2024–December 2024. Phoenix, AZ: Maricopa County Sheriff's Office.

Distribution

Distribution unlimited.

June 2025

TABLE OF CONTENTS

FIGURES

TABLES

This page intentionally blank.

Executive Summary

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.[1,2] Since 2014, the MCSO has worked toward achieving compliance with four federal court orders 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 four related court orders, respectively titled the First Order, Second Order, Third Order, and Fourth Order. As a feature of the First Order, the MCSO must conduct organizational- and individual-level analyses of patrol activity to determine whether racial/ethnic disparities exist in MCSO traffic stop outcomes. This report provides 13 baseline analyses that looked for disparities in the outcomes of traffic stops. It compares the outcomes of Hispanic, Black, and all racial/ethnic minority drivers as a group to those of White drivers across five benchmarks: stop length, citation rate, search rate, arrest rate, and seizures following a search. In this report, we identify five disparate outcomes when compared to White drivers: a 31-second longer stop length for Black drivers, a 0.45 percentage point difference in search rate for all minority drivers, and differences in arrest rates for all non-White groups (Hispanic, Black, and minority drivers). Findings of statistical significance for certain benchmarks may guide which actions and efforts the MCSO prioritizes to address racial/ethnic inequality in traffic stops. Per the requirements of Paragraph 70 of the First Order, the MCSO continues to monitor and, when necessary, intervene to address any potential racial/ethnic bias across all the benchmarks.

Paul Penzone was elected Maricopa County Sheriff in November 2016 and took office in January 2017. In 2018, the MCSO contracted with the CNA Center for Justice Research and Innovation to analyze annual patrol activity and support the development of quarterly reports on special topics related to traffic stops. In January 2024, Penzone resigned his position as Sheriff, and Chief Deputy Russell Skinner was appointed Sheriff by the Maricopa County Board of Supervisors for the remainder of the term. The current Sheriff, Jerry Sheridan, was elected in November 2024 and assumed office on January 1, 2025.

This report examines patterns of patrol activity within the MCSO; it does not analyze or identify individual deputies.[3] This report analyzes all traffic stops made by MCSO deputies from January 2024 through December 2024. The MCSO uses all traffic stop reports to understand patrol activity in the office and as a foundation to inform potential interventions, initiatives, and new or revised policies. This work is being carried out in conjunction with the court-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 hundreds of variables available through TraCS, we used a subset to analyze racial or ethnic disparities in stop outcomes. To accurately estimate the differential outcomes from traffic stops based on the race/ethnicity of the driver, we used two statistical approaches across the five relevant benchmarks: stop length, search rates, citation rates, arrest rates, and seizure rates. We used propensity score matching to analyze stop length, search rates, citation

[1] Much of the material in this section is identical to the executive summary from the *Maricopa County Sheriff's Office Traffic Stop Annual Report: January 2023–December 2023*.
[2] MCSO investigated traffic patrol activity in contracted jurisdictions and county communities in TSQR 16 (using 2023 traffic stop data) to identify whether certain jurisdictions evidenced disparity in traffic stop outcomes. Results of those analyses are available at: https://www.mcsobio.org/_files/ugd/b6f92b_7bec34a3c24c4212bb86c9cb9fd4e35c.pdf
[3] The MCSO analyzes individual deputies' traffic stops through the *Traffic Stop Monthly Report* and review process.

rates, and arrest rates. 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 racial/ethnic minority drivers[4] compared to White drivers—using a propensity score. To analyze seizure rates during searches, we used chi-square testing, which examines whether the racial/ethnic distribution of drivers subject to searches that result in seizures is different from the racial/ethnic distribution of drivers subject to searches that do not result in seizures.

MCSO deputies performed 20,265 traffic stops from January 2024 to December 2024. The rate of traffic stops per month was relatively steady throughout the year. The total number of stops in 2024 was approximately 8.76 percent more than in 2023. Within the 20,265 traffic stops, deputies perceived approximately 65 percent of drivers as White, approximately 23 percent as Hispanic, and approximately 8 percent as Black. The remaining 4 percent of stops involved individuals from other historically marginalized groups, including Asian and Native American individuals. In the dataset, more than 69 percent of the stops lasted fewer than 15 minutes. Approximately 61 percent of stops ended with a citation, 38 percent ended with a warning, and 6 percent ended with an arrest (including custodial and noncustodial arrests). Less than 1 percent of stops resulted in a non-incidental search of a driver or vehicle, meaning that the deputy decided to perform the search at their discretion.

The MCSO and the CNA analysis team conclude that there is evidence of disparity in one traffic stop outcome between White drivers and the plaintiff class (Hispanic drivers), two traffic stop outcomes between White drivers and Black drivers, and two traffic stop outcomes between White drivers and racial/ethnic minority drivers as a whole. This finding is a deviation from findings from previous court-ordered traffic stop analyses of the MCSO, especially last year's findings. Stops involving Hispanic drivers were more likely to result in an arrest than stops involving White drivers. Stops involving Black drivers were more likely to be longer than stops involving White drivers and more likely to result in an arrest. Stops involving all racial/ethnic minority drivers were more likely to result in a search and an arrest than stops involving White drivers. These disparities represent potential indicia of bias as described in the First Court Order, but do not directly represent discriminatory policing. The MCSO has been striving to reduce or eliminate disparities in traffic stop outcomes using the *Traffic Stop Monthly Reports* (TSMRs), which identify individual deputies with the most disparate outcomes and allows the MCSO to intervene when an indication of bias exists, and the *Traffic Stop Quarterly Reports* (TSQRs), which focus on specific areas with identified disparities to develop actionable responses, since 2020. In addition, the MCSO evaluates policies and procedures continually and conducts 21 different inspections to ensure compliance with them. The MCSO also provides ongoing training designed to combat bias. Since 2023, the MCSO has been documenting internal reviews and considerations from the results of Traffic Stop Annual Report (TSAR) and TSQR findings to give full transparency to all actions the MCSO considers and takes after each report. For additional discussion of findings from this and previous traffic stop reports, please refer to the conclusion section.

The MCSO and the CNA analysis team collaborated to collate the data for this analysis, address missing values and other data irregularities, analyze traffic stop outcomes, and develop this report. The MCSO was responsible primarily for collecting data, adjudicating missing values and data irregularities, and reviewing the annual report. The CNA analysis team was primarily responsible for developing and executing the analytical plan and writing the annual report. The MCSO then drew conclusions from the analytical results and developed a response plan to address the findings. The analytical plan was developed collaboratively by CNA, the MCSO, the Monitoring Team, and the Parties.

[4] The "all racial and ethnic minority drivers" (referred to as "all minority drivers") analysis includes Hispanic, Black, Asian, and Native American drivers, compared with White drivers.

The TSAR does not represent all of MCSO efforts to identify and mitigate disparities. Through our quarterly reports, ongoing TSMR and reviews of deputy activity, MCSO is situated to identify disparities across these benchmarks that are the result of factors other than bias or discriminatory policing. Upon review of statistically significant TSMR findings from ongoing research, the Court Monitor has concurred with MCSO that the inequalities in traffic stop outcomes were not associated with bias. Understanding this is necessary for meaningful interpretation of the results from this report. 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.

This page intentionally blank.

INTRODUCTION

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.[5] 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 4 million residents of Maricopa County. The MCSO operates the Fourth Avenue, Durango, Estrella, Lower Buckeye, and Towers jails; the Intake, Transfer, and Release facility; and smaller temporary holding facilities in district substations. In January 2024, Penzone resigned his position as Sheriff, and Chief Deputy Russell Skinner was appointed Sheriff by the Maricopa County Board of Supervisors for the remainder of the term. The current Sheriff, Jerry Sheridan, was elected in November 2024 and assumed office on January 1, 2025. The MCSO provides patrol and investigative operations for the county's six patrol districts, which include an array of businesses, residents, recreational areas, and communities. In addition, the MCSO operates specialized units and teams, such as specialized investigations, canine teams, and tactical operations.

Since 2014, the MCSO has worked toward 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. Sheridan*), a federal judge found that the MCSO violated the rights of Latinos and Latinas 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. Murray Snow of US District Court, Arizona, issued the First Supplemental Court Order (First Order) to the MCSO 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 data collection and analysis requirements, training, and policies. The Court appointed an independent monitor for oversight.[6] As a feature of the First Order, the MCSO must conduct organizational- and individual-level analyses of patrol activity to investigate racial or ethnic disparities in traffic stop outcomes. In 2018, the MCSO contracted the CNA Center for Justice Research and Innovation to analyze patrol activity annually and produce quarterly reports on special topics related to traffic stops.

This report responds directly to the First Order requirement to analyze MCSO traffic stop activity to determine whether disparate outcomes exist between drivers of different races or ethnicities. This approach relies on propensity score matching to compare stops with similar characteristics other than the perceived race/ethnicity of the driver. This report examines patterns of patrol activity within the MCSO; it does not analyze or identify individual deputies. The MCSO analyzes individual deputies' traffic stop activity each month as part of its *Traffic Stop Monthly Report* (TSMR) process. This process has been ongoing since 2021. The MCSO expects to use this report as a knowledge

---

[5] Much of the material in this introduction is identical to the introduction in the *Maricopa County Sheriff's Office Traffic Stop Annual Report: January 2023–December 2023*.

[6] 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. In 2022, the Court issued a Third Order establishing requirements to address a backlog of misconduct investigations; and in 2023, a fourth order clarifying timelines of those investigations. The Third Order and Fourth Order do not establish requirements related to traffic stops.

base of traffic stop activity in the organization and as a guide for future research and potential interventions, initiatives, and policies. The MCSO works collaboratively with the Monitoring Team and Parties to develop policies and activities to address racial/ethnic inequality related to MCSO's mission.

Purpose of traffic stop analyses

Analyses of patrol activities are increasingly common across US law enforcement agencies. Law enforcement agencies face heavy scrutiny from the public and the media, who have concerns about bias and disparate outcomes by race/ethnicity from interactions between the police and community members. The interactions under scrutiny include officer-involved shootings, uses of force, searches, and traffic stops (see, for example, Baumgartner et al., 2021; Correll et al., 2007; Fridell & Lim, 2016; Fryer, 2016; Ridgeway, 2006; Ritter, 2017; Shoub, 2021). Although most law enforcement officers do not practice biased policing intentionally, they may exhibit behaviors that appear biased or that result from implicit bias (Bolas, 2022; Ekstrom et al., 2022; Marsh, 2009; Nix et al., 2017; Spencer et al., 2016; Stelter et al., 2022). Even though law enforcement strives for fair treatment, officers may unconsciously treat community members differently (Hall et al., 2016; Helfers, 2016; Roach et al., 2022; Stroshine et al., 2008). Of course, incidents involving explicit bias, such as racial profiling, have occurred in law enforcement practice, including the pattern of directed racial profiling that resulted in MCSO's court-ordered monitoring. Under Sheriff Sheridan's leadership, all allegations of explicit bias are taken seriously and investigated thoroughly, and discipline (including termination) is meted out as quickly as possible.

Implicit bias refers to attitudes or stereotypes that affect understanding, actions, and decisions unconsciously (Staats et al., 2015). Officers' implicit biases may affect their interactions with a driver when making a traffic stop and may affect individual stop outcomes. This issue persists beyond the scope of law enforcement agencies—all people possess implicit biases, and implicit biases occur naturally on a subconscious level throughout society (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 have developed methods to identify officers who need implicit bias or other training through quantitative analysis of disparate outcomes. Implicit bias occurs and affects all individuals without their awareness or intentional control (Staats et al., 2015). In contrast to implicit bias, *explicit bias* refers to conscious attitudes and beliefs about a person or group (James, 2018), such as prejudice.

Over time, methods for identifying evidence of disparate outcomes have evolved. Early research on bias in policing and disparate stop rates or 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 such as propensity score matching and weighting to analyze traffic stops and other law enforcement activity outcomes for evidence of racial or ethnic disparity (Knode et al., 2024; Ridgeway, 2006; Riley et al., 2005; Tillyer et al., 2010). Methods for assessing disparities have evolved to incorporate measures beyond stop rates, focusing instead on stop outcomes, such as citations and searches (Christiani et al., 2022; Fridell, 2004; Fridell, 2005; Onookome-Okome et al., 2022; 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 et al., 2005).

Understanding the expectations and limitations of quantitative analysis for investigating racial and ethnic disparities is important for interpreting the findings in this report. Research on traffic stops includes both pre-stop and post-stop analyses. Pre-stop analyses investigate whether the race/ethnicity of the driver affects stop rates; post-stop analyses investigate whether the race/ethnicity 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 because such data include non-drivers and may not accurately reflect the driving population or the racial/ethnic distribution of drivers who violate traffic laws (McMahon et al., 2002; Tregle et al., 2019). Other methods for estimating the racial/ethnic distribution of the driving population include observing and recording the race/ethnicity of drivers in a given jurisdiction over time or using driver license race/ethnicity data. However, these methods can be cost prohibitive or infeasible because of data unavailability (Fridell, 2004; Tillyer et al., 2010), while most states, including Arizona, do not capture race/ethnicity in their driver's license or registration documentation.

Post-stop analyses mitigate some of these issues because the population they capture is contained within the traffic stop data and does not need to be estimated (Ridgeway & MacDonald, 2010; Withrow et al., 2008). Post-stop analyses are not without their own limitations, however, such as unobserved differences between groups that can confound estimates of disparity (Ridgeway, 2006).

Despite improvements in analytical methods, analysts still need accurate and in-depth traffic stop data from agencies to measure and identify disparate outcomes; the absence of adequate data can limit the scope of analysis and make identifying policy responses to address disparities difficult. Some agencies collect data from their traffic stops meticulously, whereas other agencies track only limited information, such as when a stop occurred, the driver's race/ethnicity, and limited stop outcomes, or store data about traffic stops across data systems that cannot be linked readily.

Practitioners and consumers of bias research should understand that disparate outcomes do not definitively indicate the biases that exist across society (Fridell, 2004; Simoiu et al., 2017). Quantitative analysis cannot capture all the possible reasons that could explain the disparate outcomes. Even with these limitations, the results from statistical analyses can provide insight into policing practices in an agency, helping the agency identify disparate outcomes to address. Such analyses provide agencies with a tool to review officer traffic stop conduct and determine the necessary actions, if any, for officers and agencies.

Many law enforcement agencies now analyze their traffic stop data internally or in partnership with external researchers and analysts. Most analyses conducted to date have found racial or ethnic disparity in traffic stop outcomes (e.g., searches, citations, arrests). Tillyer et al. (2010) state, "Analyses of these data demonstrate a relatively consistent trend of racial/ethnic disparities in vehicle stops and vehicle outcomes." Most existing studies have found evidence of racially disparate rates of stops or outcomes of patrol activity in law enforcement agencies (Ariel & Tankebe, 2018; Baumgartner et al., 2018; Engel & Calnon, 2004; Gaines, 2006; Gelman et al., 2007; Hannon et al., 2020; Iwama & Mcevitt, 2025; Norris et al., 1992; Novak, 2004; Pierson et al., 2020; Roach et al., 2022; Rodriguez et al., 2019; Rojek et al., 2004; Rosenfeld et al., 2012; Shoub, 2025; Smith & Petrocelli, 2001; Stelter et al., 2022; Tillyer & Engel, 2013; Vito et al., 2020; Webb et al., 2021; Weiss & Rosenbaum, 2006). A few studies have documented findings of no racial or ethnic disparity in traffic stop outcomes (Grogger & Ridgeway, 2006; Higgins et al., 2012; McCabe et al., 2020; Taniguchi et al., 2016; Zhang & Zhang, 2021). The balance of the evidence suggests that disparate outcomes during traffic stop activity are common in law enforcement agencies in the United States; however, acknowledging the prevalence of the problem does not imply that agencies should not address disparate outcomes pragmatically and proactively by promoting anti-bias policy, training, and practices.

The MCSO currently combats biased policing through ongoing enhanced trainings, continual policy reviews and revisions, a series of inspections that include a statistical review of every deputy's traffic stops each month (approximately 40,000 stops over the course of a year), and an early intervention system to help supervisors identify and intervene into potentially problematic behaviors. In addition, the MCSO's Community Outreach Division is involved in community policing and recruitment efforts in the communities that the MCSO serves. In response to the findings from last year's *Traffic Stop Annual Report (TSAR) 9*, the MCSO conducted the following activities: communicated results to employees and the public through town halls and its website, provided monitor-approved enhanced training to personnel to address the findings in TSAR 9, developed a real-time information dashboard to help the Motors Unit patrol for traffic safety, developed a traffic stop review dashboard to enable supervisors to review subordinates' stops, validated the use of extended stop indicators, analyzed disparities both within and between districts, and explored data continuously to understand disparities and potential ways to mitigate them.

The use of statistical analysis for identifying racial/ethnic disparities in traffic stops has become increasingly crucial, and previous analyses indicate that disparities exist across the nation. The MCSO has been and continues to be committed to ongoing research to identify responses to address those disparities. The MCSO is at the forefront of analyzing racial/ethnic disparities in traffic stop outcomes in the United States and no other policing agency in the country conducts as extensive and in-depth analyses of racial/ethnic disparities in policing.

Organization of this report

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

In addition, 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. The appendices also include deprecated analyses related to previous TSAR reports. We provide the analyses to allow additional comparisons identifying potential influences on traffic stop outcomes. We reference the deprecated statistical models throughout the report if they produced results that were different from the Monitor's approved baseline models. 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 that we used in the traffic stop 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 that takes place before analysis. We then discuss the propensity score matching approach that we used to assess racial/ethnic disparity in the length of stops, search rates, citation rates, and arrest rates, as well as the chi-square analysis that we used to assess racial/ethnic disparity in seizure rates. We discuss the alternate specifications that 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

TraCS is a data collection, records management, and reporting software for public safety professionals. The MCSO uses TraCS to capture data about traffic stops. 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 GPS location of each traffic stop.[7] The system also requires the deputy to enter variables such as the perceived race/ethnicity of the driver,[8] the contact conclusion, whether an arrest took place, and search and seizure information. TraCS also includes data fields for capturing information about any delays during the stop, such as training, driving under the influence (DUI) investigations, tows, technical issues, language barriers, driving documentation issues, or other issues, and it includes a comment field for deputies to input additional information.[9] After the deputy fills out information about events in TraCS, the system forwards entries for supervisory review and, if necessary, revision. We used a subset of the several hundred variables available through TraCS to analyze racial or ethnic disparities in stop outcomes and we constructed and appended data using variables present in TraCS and other MCSO systems. Here, we briefly describe the variables that we 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 the perceived race/ethnicity 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 traffic stop 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 the stop time in hours and minutes. The stop classification summarizes the violation, per the Arizona Revised Statutes (ARS), classified into five categories: civil, criminal, civil traffic, criminal traffic, and petty. We also included the specific violation category, coded as speed, non-speed moving, equipment, license/insurance/registration issues, and other violations. 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, calling for a tow, or additional time spent verifying documentation (e.g., a driver without their insurance card verifying their

---

[7] 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.

[8] Arizona does not collect information about race/ethnicity as part of its driver license system; thus, all race/ethnicity categories within TraCS data are based on the perception of the deputy who made the stop.

[9] 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/about-us/general-info/mcso-policies

insurance by phone). We also included 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 (or motorcycle), supervisor, and other. For analyses presented in Supplemental Appendix 5, we noted whether the stop took place as part of a special assignment, such as DUI Task Force or Aggressive Driving Task Force. We included a variable that indicates whether a particular stop was for a speeding offense; if so, we included a categorical variable indicating the number of miles per hour over the posted speed limit documented for the stop.

Data about stop outcomes. Stop conclusion data describe the outcome of the stop as a citation, warning, or incidental contact. TraCS indicates whether a stop included a search of the driver or vehicle (we omitted passenger searches from this analysis because our focus is on drivers) and whether that search was incident to arrest or towing.[10] 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 were non-incidental (i.e., discretionary) in nature. For example, policy dictates that all individuals be searched before custodial arrest and courtesy rides, and that all vehicles be inventoried before tow; searches that occur incident to arrests for courtesy rides, or inventory searches for vehicle tows are not discretionary and were thus excluded from our analysis of search outcomes. Deputies also indicate in TraCS whether a search resulted in the seizure of contraband.

Data about the driver. We used the post-stop perceived race/ethnicity of the driver, as entered by the deputy, to classify the driver's race/ethnicity 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 drivers of unknown sex collapsed as the comparison category). We also included the reported license plate state of the vehicle that the driver was operating, classifying it as either an in-state or out-of-state plate.

The CNA analysis team appended data not housed in TraCS into our analysis, including information about special assignments. The MCSO manually compiles data about special assignments by deputy and by date. During the timeframe of this analysis, the special assignments included DUI Task Force, Click-It-or-Ticket Task Force, and the Aggressive Driver Task Force. The analysis team also constructed a deputy traffic stop count variable equal to the number of stops that the deputy made over the 12-month period, for descriptive purposes.

Data verification and missing data

The analysis team reviewed the 2024 TraCS data for data quality (e.g., missing data, out-of-range values) and verification.

The analysis team identified additional missing data that the MCSO could not adjudicate. For example, one stop lacked data on perceived driver race/ethnicity because the driver fled, and the traffic stop was never completed. These missing data represent less than 1 percent of the overall data, which is under any standard threshold that would trigger concerns about missing data biasing the analysis or findings. Supplemental Appendix 1 describes the missing data by variable.

In addition, TraCS creates duplicate lines to capture data for multiple contacted passengers; because this analysis focuses solely on drivers, these lines represent duplicate data. We identified these duplicate entries based on the event number, deputy's badge number, driver's first and last name, and removed them.

---

[10] A detailed analysis of search activity during traffic stops is available in *Traffic Stops Quarterly Report 10: Searches*, https://www.mcsobio.org/_files/ugd/b6f92b_8fd0a6175a6f4d6483a8d97fa75f4d42.pdf.

Methodology

To accurately estimate differential outcomes from traffic stops based on the perceived race/ethnicity of the driver, we used two statistical approaches across the five benchmarks under consideration. To analyze the length of stops, search rates, citation rates, and arrest rates, 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 in this section.

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) is not feasible or practical; these techniques leverage specific data structure and statistical techniques to approximate experimental conditions (Shadish et al., 2002). In this case, propensity score matching matches individual events (i.e., traffic stops) with similar events based on their characteristics (listed in the next 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 racial and ethnic minority drivers[11]) using a propensity score (Apel & Sweeten, 2010; Rosenbaum & Rubin, 1983).

For this traffic stop analysis, we used the first stage of propensity score matching to determine the probability that a stop involved a driver of a particular race/ethnicity (i.e., Hispanic, Black, and all minorities). For all analyses, stops involving White drivers were the comparison conditions. We performed matching analyses using the observed characteristics of the stop—specifically the driver's sex, the stop longitude and latitude, the stop time, the stop classification (operationalized as civil traffic stops versus all others for all analyses, excluding arrests[12]), the violation type, whether the vehicle had out-of-state plates, the speed over the speed limit for stops with speeding offenses, whether the deputy indicated that the stop was extended for one of the reasons discussed in the previous section, and the call sign category under which the deputy was operating.[13,14] As an alternate specification, we also considered each model with the inclusion of special assignments as a matching variable. In addition, for the length of stop analysis only, we included whether the stop involved an arrest or a search; both these circumstances necessarily result in longer stops. For length of stop analysis, we excluded stops with extended stop indicators because the MCSO's *Traffic Stop Quarterly Report: Extended Traffic Stop Indicator Use* verified that deputies are using extended stop indicators appropriately.[15] For citation outcomes, we also included a variable indicating whether the stop was pursuant to ARS 28-3151A.[16]

To obtain the propensity scores, we used a logistic regression model with elastic net regularization. Regularization is a common tool used to prevent a model from "overfitting" to the data, which it does by preventing model coefficients from growing too large. It is especially useful for models with many features or for models with features that are strongly correlated, which is the case here. Logistic regression models use two common types of regularization—LASSO, which tends to set certain coefficients to 0, thus eliminating features from the model entirely, and Ridge, which tends to make certain coefficients very small but rarely exactly 0. Elastic net is a combination of Ridge and LASSO and

---

[11] The "all minority drivers" analysis includes Hispanic, Black, Asian, and Native American drivers, compared with White drivers.

[12] MCSO altered the analyses of arrests for this annual report. In previous reports, analyses of arrests included matching on stops that were classified as civil and criminal. Because all stops classified as criminal are also arrests, this variable was removed from the matching process.

[13] Both stop time and stop location are fitted using splines to allow a more flexible functional form.

[14] The logistic regression also included interaction terms for stop location, and between the speeding violation variable and the categorical variables capturing the speed recorded over the speed limit.

[15] Report available here: https://www.mcsobio.org/_files/ugd/c866a6_f37279fd33394818bb370ab6af46820e.pdf

[16] ARS 28-3151A indicates that individuals should not drive without a valid driver license.

requires the analyst to provide two "hyperparameters." One controls how much regularization to apply, and the other controls whether the model behaves more like LASSO or more like Ridge regression. To obtain these hyperparameters, we performed a grid search across possible pairs of candidate values for these two hyperparameters and evaluated each pair using five-fold cross-validation. For each model, we chose the pair resulting in the lowest validation error across the five folds. We then used this pair to fit a regularized logistic regression model on the entire dataset, and we used the fitted values for that model as the propensity scores.

Cross-validation is a common method for evaluating model settings. The dataset is split into five "folds," with each data point appearing in a single fold. The model is fit five times for each candidate setting. In each case, one fold is withheld from the dataset, the model is trained on the other four folds, and the model is then evaluated on the withheld fold. The objective function that we used to evaluate each model on each fold is cross-entropy error, also known as deviance. The setting that resulted in the best average performance across all folds was then used to fit the model on the entire dataset.

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 presents 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 5 presents detailed results from the robustness check analyses.

For all analyses, we present findings in terms of the average treatment effect on the treated (ATT)—that is, the average difference in outcomes between those who actually receive treatment. In this report, the ATT reflects the difference between outcomes in stops involving Hispanic, Black, or all racial and ethnic minority drivers versus those involving White drivers. We report the ATT as a measure of difference between Hispanic, Black, and minority drivers when compared to White drivers. To avoid confusion, *treatment* and *treated* in this context are terms derived from experimental methods identifying the treatment and control groups. In the context of the analyses presented in this report, *treatment* and *treated* refer to the racial or ethnic group analyzed and do not refer to deputies' interpersonal interaction with drivers. We conducted standard checks of balance and common support for all propensity score analyses. 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, perceived race or ethnicity). This test uncovers whether rates of seizures vary significantly across racial or ethnic categories. As noted in the literature, different rates of seizures may indicate racial or ethnic bias because differences suggest that deputies may use different decision criteria or thresholds before searches of drivers of different races or ethnicities (Persico & Todd, 2006; Ridgeway & MacDonald, 2009; Simoiu et al., 2017; Walker, 2003). For this analysis, we considered only discretionary searches. We used a standard chi-square analysis with Pearson's and likelihood ratio tests (Pearson, 1900). We also ran Fisher's exact test (because of 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 to match multiple other observations) for nearest neighbor and radius matching.[17] The supplemental appendices to this report present the results from the alternate specifications. In the supplemental appendices, we provide additional alternate specifications for stop length, citations, and arrests, which we analyzed using variations on the variables used in the analyses. All parties agreed upon the baseline analyses; to ease the interpretation of results, we present only baseline analyses in the main body of this report and provide all deprecated and alternate specifications in the supplemental appendices. We note in the main body of the report when the results from the baseline analyses and deprecated model specifications are different.

Considerations and limitations

Propensity score matching represents a substantial improvement over past methods of estimating racial/ethnic disparity in law enforcement activities because it does not rely on the development of imperfect or cost-prohibitive external benchmark data and it more precisely estimates the differences in outcomes when accounting for differences in circumstance among interactions (e.g., traffic stops). However, the methodology is not without limitations. First, propensity score matching does not directly measure bias or racial discrimination. Instead, the method identifies indicia of potential bias by measuring unequal outcomes. Second, as noted above, the matching step relies on the estimation of a logistic regression, which requires estimates to converge over iterative analysis steps. This requirement can limit the inclusion of variables and observations if convergence is impossible for a given model specification. The models do not account for the unobserved characteristics of the driver or the stop, for example, whether the driver committed multiple violations or when additional violations (e.g., license, insurance, or registration violations) are identified during the stop. This limitation can result in inaccurate estimates of treatment effects. The model also cannot account for any variable that perfectly predicts the condition of interest, such as DUI violations, or control for confounding characteristics of traffic stops not included in the models (Ridgeway, 2006; Rosenbaum and Rubin, 1983).

In addition, another limitation is the potential for model dependence, as findings may vary based on modeling choices (for example, matching variable selection). Although MCSO and CNA made efforts to select reasonable and theoretically grounded base models, alternate specifications have the potential to produce difference results (see Appendices 3 through 5). Furthermore, this TSAR introduced additional matching variables which will lead to a reduced effective sample size because not all observations in the original dataset will have a suitable match. As discussed with the Monitors and the Parties at the February 2025 site visit, introducing the ARS 28-3151A variable as a matching variable reduced the effective sample size of White drivers for the citations outcomes.

Finally, as with all statistical techniques used 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 or ethnicity—they cannot provide insight into the underlying causes of these disparities on their own.

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

This page intentionally 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 the race/ethnicity of the 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/ ethnicity. Most existing or proposed external benchmarks (e.g., census populations) provide inaccurate estimates of the driving population or are impractical and cost prohibitive (e.g., collecting data on driver race/ ethnicity using observations at intersections). We considered several emerging practices (e.g., comparing daytime versus nighttime stop rates, using accident data, comparing criminal versus civil traffic stop rates), but we could not implement them using the currently available data from the MCSO or other sources, such as Arizona driver's license information. Therefore, for stop rates, we present descriptive statistics only.

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

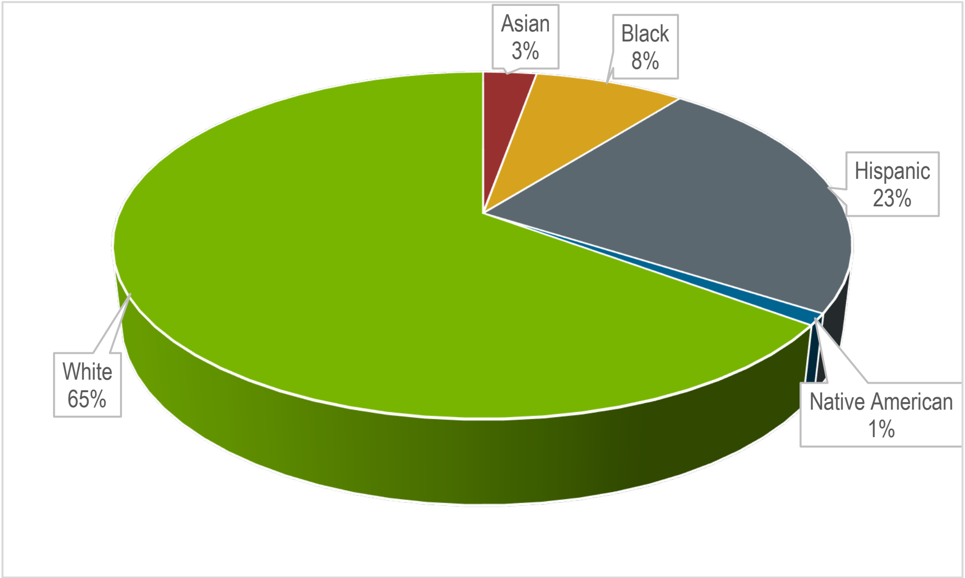
Descriptive statistics

In this section, we describe the data included in this analysis of traffic stops conducted by the MCSO between January 2024 and December 2024 (a 12-month period). We present the stops' characteristics, the characteristics of the stop outcomes, and the traffic stop count of the deputies making the stops. Supplemental Appendix 1 provides a full table of descriptive statistics for each variable.

Driver characteristics

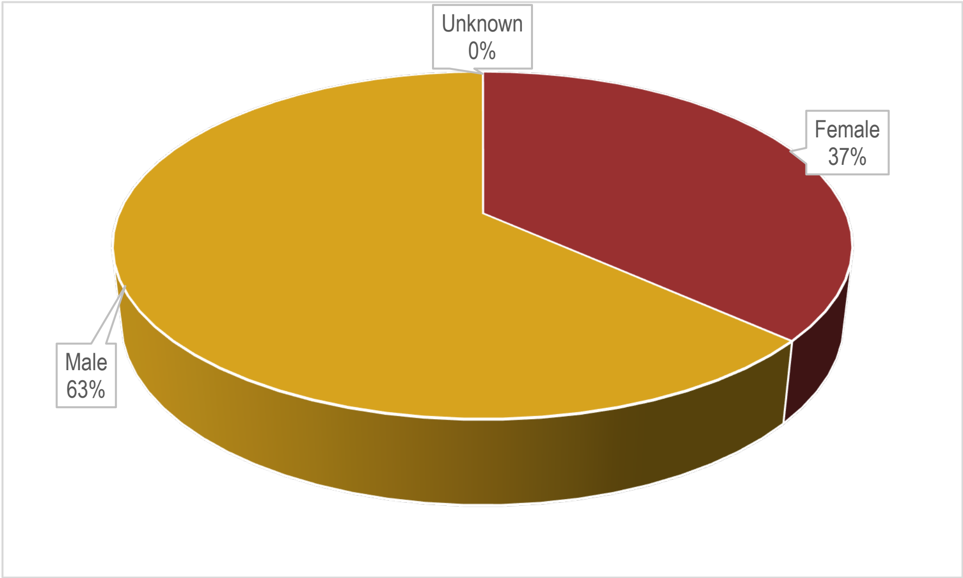
When deputies make a traffic stop, they document their observation of the perceived race/ethnicity of the driver both pre- and post-stop in TraCS. Deputies do not ask drivers questions about race/ethnicity during stops, and that information is not provided on Arizona driver's licenses. We omitted analysis of the pre-stop perception of driver race/ethnicity because this variable takes the value "unknown" in approximately 97 percent of stops. Post-stop, deputies perceived 65 percent of drivers as White, 23 percent as Hispanic, and 8 percent as Black. The remaining 4 percent of stops were of Native American and Asian drivers (Figure 1).

Figure 1. Stops by post-stop perceived driver race or ethnicity



The deputies also enter post-stop perceived sex in TraCS. Deputies perceived 63 percent of drivers as male and 37 percent as female. In 3 stops (less than 1 percent), the deputy could not determine the sex of the driver (Figure 2).

Figure 2. Stops by post-stop perceived driver sex

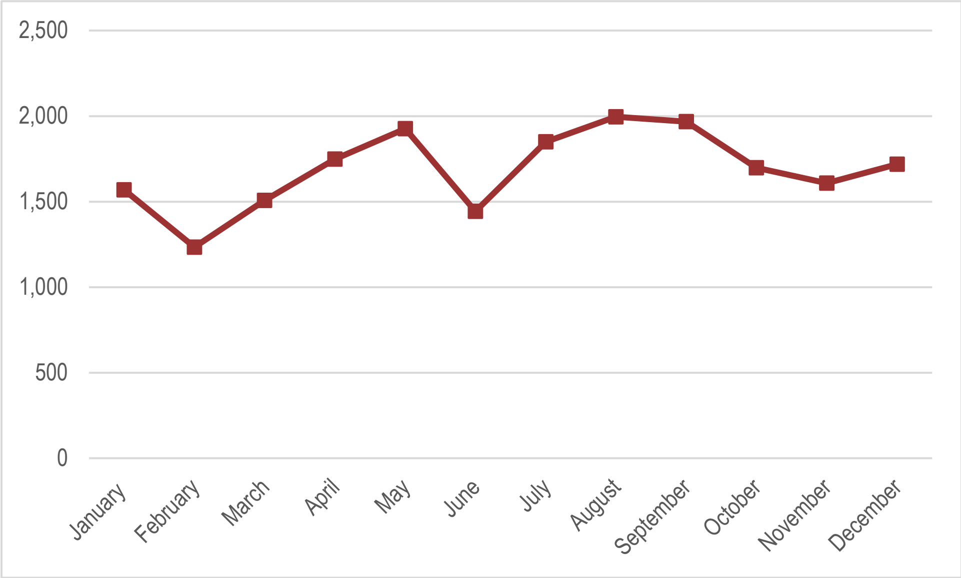


Stop characteristics

MCSO deputies performed 20,265 traffic stops during the 12-month period observed for this analysis. Monthly traffic stops increased between February and May 2024 and again between June and August 2024. Traffic stops decreased between January and February 2024 and again between May and June 2024 (Figure 3). This overall trend is different from what we observed in 2023, when there was a downward trend between January and May 2023 and again between August and October 2023. Traffic stops increased between May and June 2023 and again between

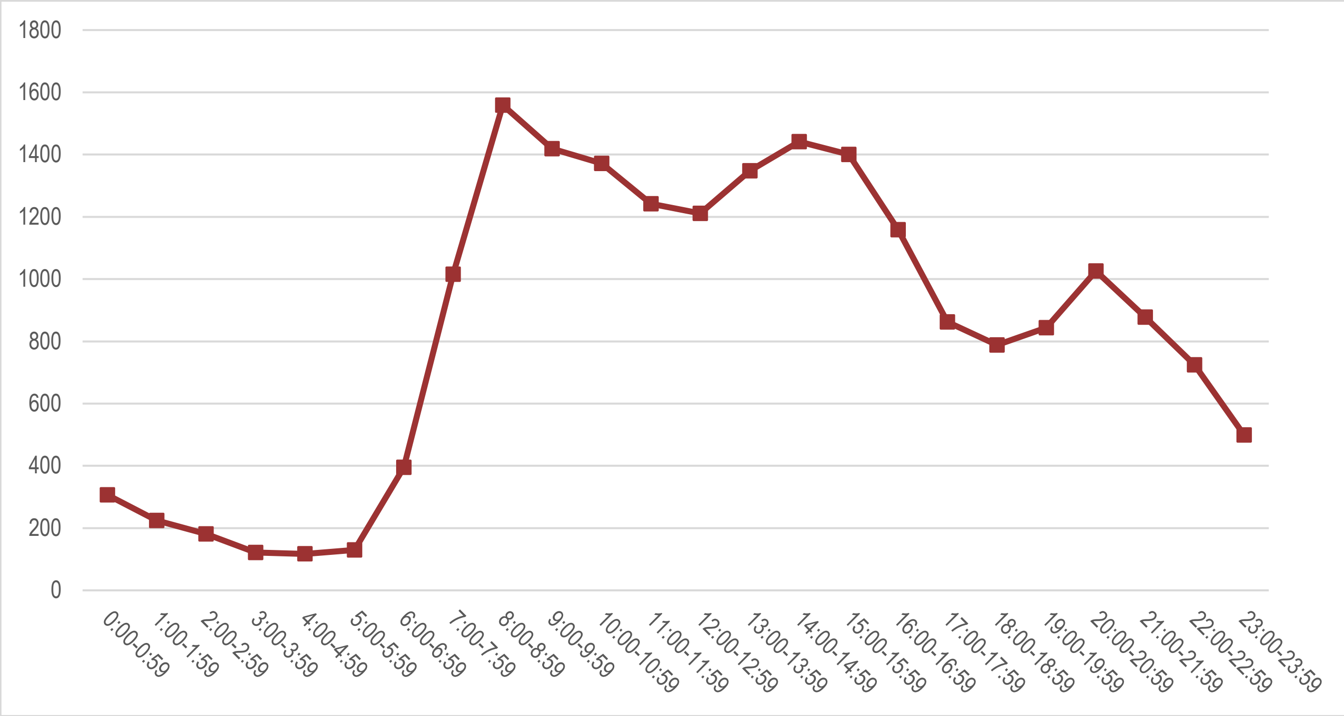
June and August 2023. There was an overall increase in stops of approximately 9 percent in 2024, compared to 2023.

Figure 3. Stops by month, January 2024 to December 2024



We also considered the time of day that a stop took place. Most stops occurred between 7:00 a.m. and 8:00 p.m., which is similar to trends in previous years. After 8:00 PM, there was a steady decrease until 6:00 AM. (Figure 4).

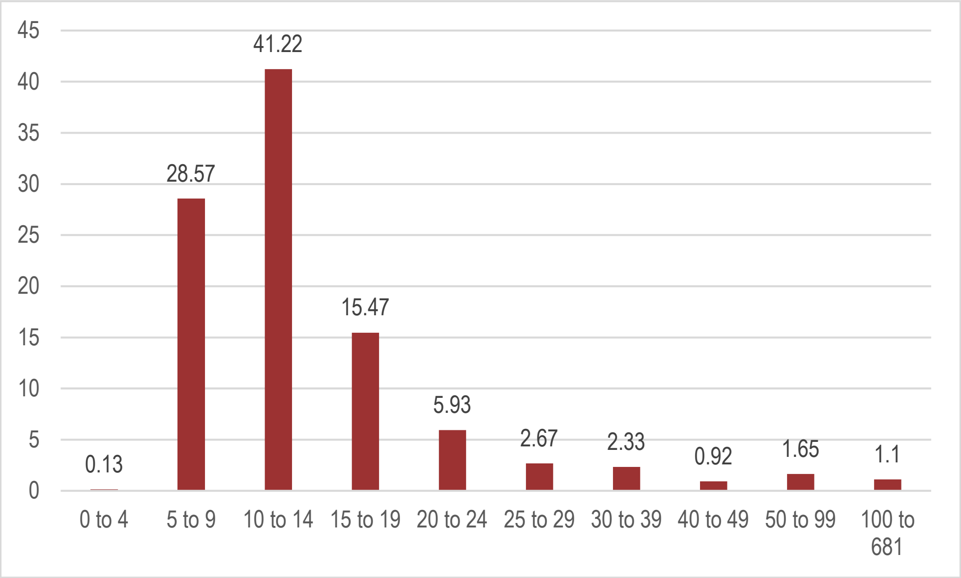
Figure 4. Stops by time of day



Stop length is particularly important to this analysis because it is a core aspect of the Court Order. Stops lasted an average of 15.99 minutes (with a standard deviation of 22.19 minutes), a 4.65 percent decrease from the annual report on the 2023 data (in which the average stop length was 16.77 minutes with a standard deviation of 22.24

minutes). Most stops lasted between 5 and 14 minutes (Figure 5).

Figure 5. Stop length, in minutes



Deputies document in TraCS whether a stop was extended for reasons that would reasonably extend a stop. The extended stops field contains seven options: DUI stop, language barrier, technical issues, training stops, vehicle towed, driving-related documentation issues, and other issues. Deputies selected extended stop indicators for 10,194 stops, representing 50.30 percent of total stops, which is an increase from the previous annual report, in which extended stops represented 37.30 percent of total traffic stops.[18] The most frequent reason for an extended stop was documentation issues (36.80 percent), and the next most frequent reason was other issues (12.50 percent). This finding differs from that in the previous year's report, in which documentation and technical issues were the most common reasons for extended stops (Table 1).

Table 1. Extended stop reasons

Reason indicated	Percentage of stops
DUI Stop	1.83%
Language Barrier	2.70%
Technical Issue	10.44%
Training Stop	5.93%
Vehicle Towed	1.78%
Documentation Issue	36.80%
Other Issue	12.50%

We also considered which stops occurred while the deputy was on a special patrol assignment. Of the 20,265 stops

[18] The MCSO has investigated extended stop indicator use with 2020, 2023, and 2024 stop data in *Traffic Stops Quarterly Reports* 3, 13, and 17 which can be found here: https://www.mcsobio.org/traffic-stop-data

in the dataset, 938 stops occurred while deputies were on DUI Task Force assignment, 1,205 occurred while deputies were on the Aggressive Driver Task Force, and 97 stops occurred while deputies were on the Click-it-or Ticket Task Force (Table 2).

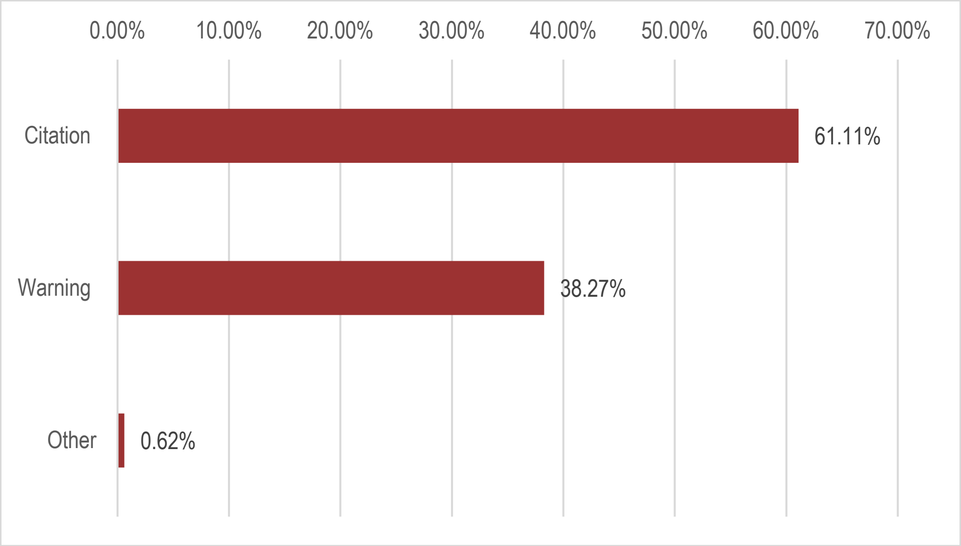
Table 2. Stops conducted during special assignments

Special assignment	Count
DUI Task Force	938
Aggressive Driver	1,205
Click-It-or-Ticket	97

Stop outcomes

Contact conclusion documents the outcomes from each stop. Of all traffic stops, 61.11 percent concluded with a citation, 38.27 percent ended with a warning, and less than 1 percent ended with non-enforcement outcomes, such as when a deputy is preempted for a priority call and ends the traffic stop without issuing a citation or documented warning (Figure 6).[19] Unlike the previous year (2023), White drivers were cited at higher rates than both Black and minority drivers and at slightly lower rates than Hispanic drivers (Table 3).

Figure 6. Traffic stop contact conclusions



[19] This citation rate has increased over the TSARs in recent years. For example, TSAR 8 reported a citation rate of 51.6 percent, and TSAR 9 reported a citation rate of 52.27 percent. With the increase in citations, MCSO has reported decreased disparities in citation rates.

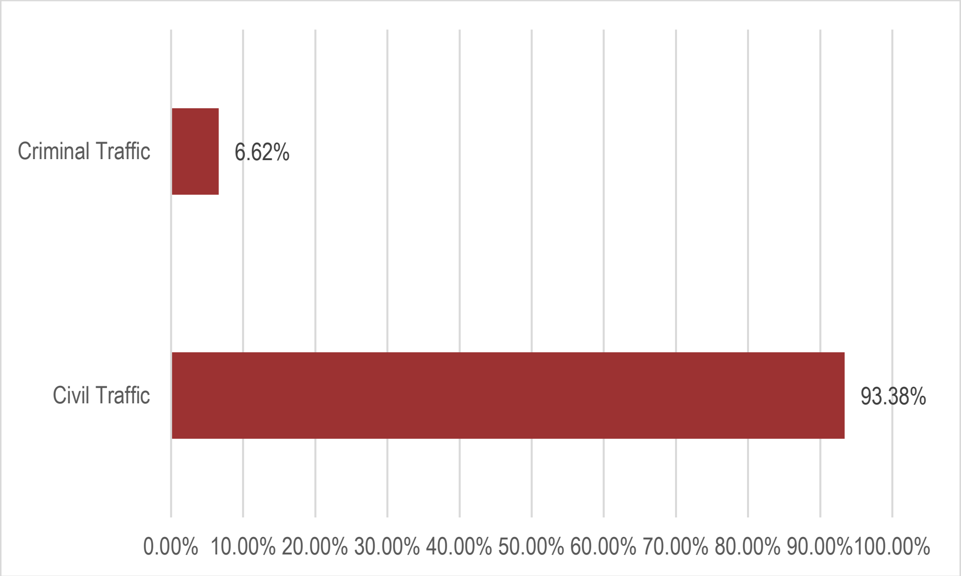
Table 3 presents the stop outcomes when considering the race/ethnicity of the driver.

Table 3. Traffic stop contact conclusions by race and ethnicity

Race and ethnicity	Citation	Warning	Other
White drivers	61.59%	37.89%	0.52%
Hispanic drivers	62.12%	37.10%	0.78%
Black drivers	55.18%	43.82%	1.00%
Minority drivers	60.23%	38.97%	0.80%

The MCSO organizes stops into five categories based on ARS code: civil traffic, criminal traffic, petty, civil, and criminal. In 2024, all stops were either civil traffic or criminal traffic stops.[20] Most stops were civil traffic stops. Civil traffic violations are violations in which the driver does not face jail time and instead may be subject to a fine. Examples include speeding, equipment violations, and seatbelt violations. Criminal traffic violations are traffic violations that result in a fine and involve possible jail time. Examples include criminal speeding, reckless driving, DUI, and driving with a revoked or canceled license. All criminal or criminal traffic violations are considered arrests whether the driver was detained or not. Petty violations are criminal violations with less severe penalties that do not include the possibility of jail time. Examples include boating violations, park violations, and curfew violations. Criminal violations, which are not in this year's dataset, 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 unrelated to the stop. Of the 2024 traffic stops, 93.38 percent resulted in a civil traffic classification and 6.62 percent resulted in a criminal traffic classification (Figure 7).

Figure 7. Traffic stop classifications



The MCSO categorized stops into five violation categories: speed, non-speed moving, equipment, license/insurance/registration issues, and other violations. Violation categories were derived from ARS sections and subsections that were entered into citation or warning forms issued to drivers during a stop. Speeding violations were violations associated with exceeding the speed limit (e.g., civil speeding, criminal speeding, speeding in a school zone). Non-speed moving violations were safety violations for which the vehicle was moving, such as failing

[20] MCSO had approximately 20 stops that were criminal, but these were classified as criminal traffic for this report.

to signal when changing lanes, failing to stop, tailgating, or driving too slowly. DUI violations were included in the non-speed moving category. Equipment violations were violations in which a driver's automobile lacked proper equipment, had non-functioning equipment, or had equipment deemed unsafe (e.g., broken taillights or headlights, cracked windshields, illegal modifications, opacity on window tint). License/insurance/registration issue violations were violations associated with licensing (vehicle or driver), insurance, and registration. Examples include driving with a suspended or revoked license or expired registration, failing to possess insurance, and driving without license plates. Stops with ARS 28-3151A violations were categorized separately. Finally, other violations included all violations that could not be identified as one of the above categories. The other violation category included a diverse collection of offenses, such as drug violations, seat belt and cell phone violations, parking violations, noise violations, and littering violations. Like in 2023, speed violations were the most common violations, followed by license/insurance/registration issue violations (Figure 8).

Figure 8. Violation categories[21]

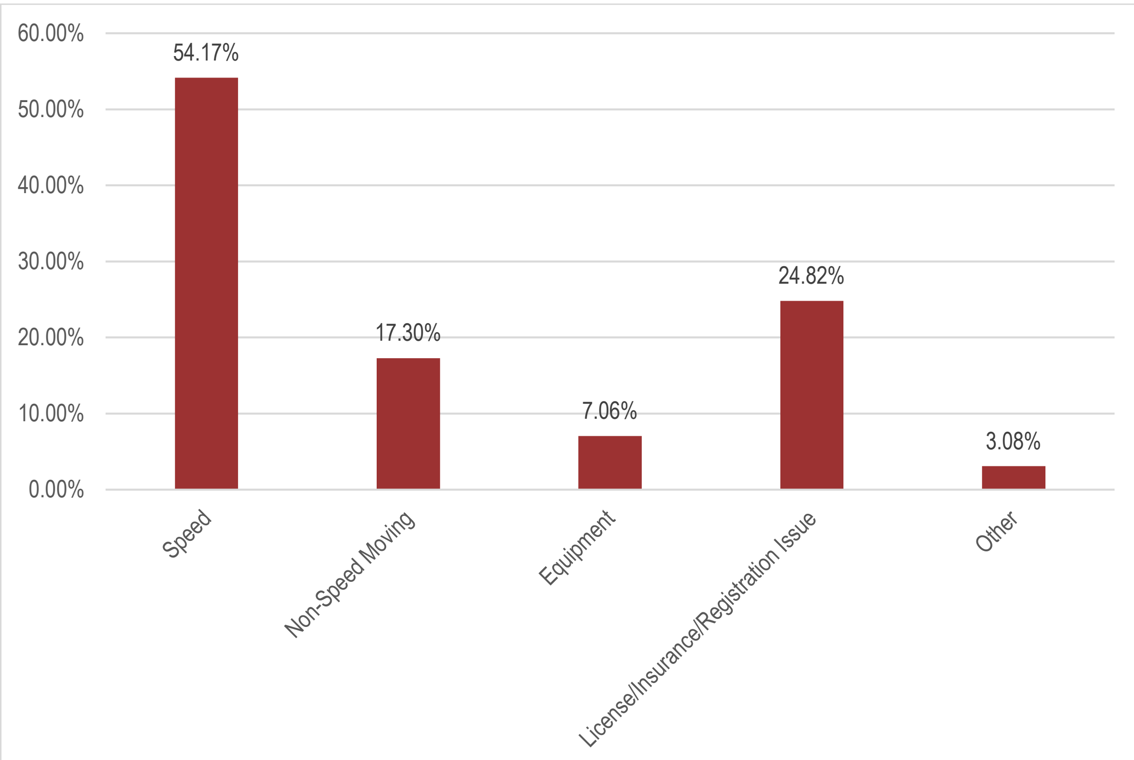
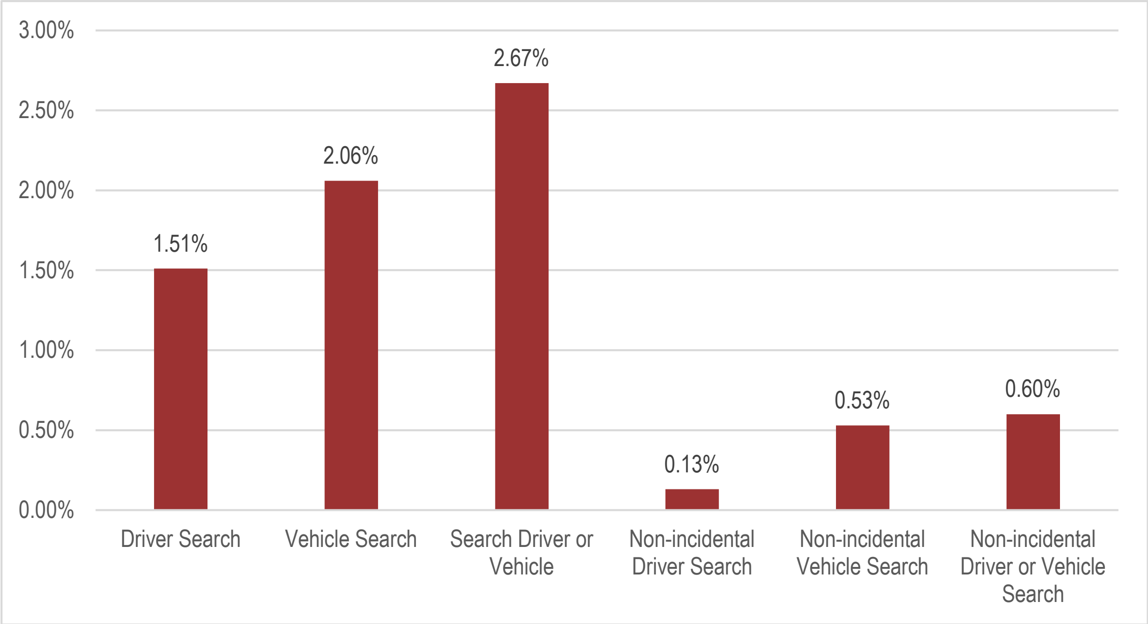


Figure 9 presents information about searches as a percentage of total stops. MCSO policy dictates that deputies search all arrested drivers, drivers given courtesy rides, and all towed vehicles; these searches are not discretionary on the deputy's part. Non-incidental searches are not connected to arrests, courtesy rides, or tows, meaning they are discretionary searches conducted by deputies. Most searches of drivers occurred incident to arrest. However, these represent 0.06 percent of all traffic stops. For the comparative analysis, we considered non-incidental searches of drivers or vehicles as a search outcome; MCSO deputies conducted slightly more searches of vehicles than of drivers.

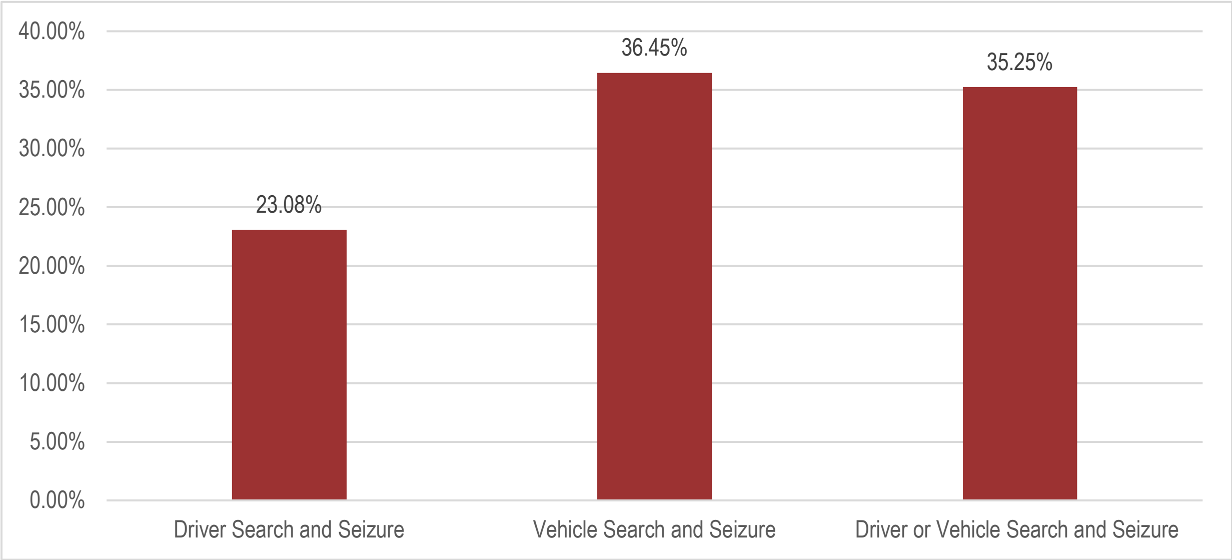
[21] Note that combined percentages exceed 100 percent because drivers may be cited or warned for more than one offense.

Figure 9. Searches



For all stops involving a search, deputies record whether the search turned up contraband (i.e., the incidence of seizures predicated on searches, or "hit rate"). Overall, 35.25 percent of non-incidental searches resulted in seizures (Figure 10).

Figure 10. Seizures during non-incidental searches



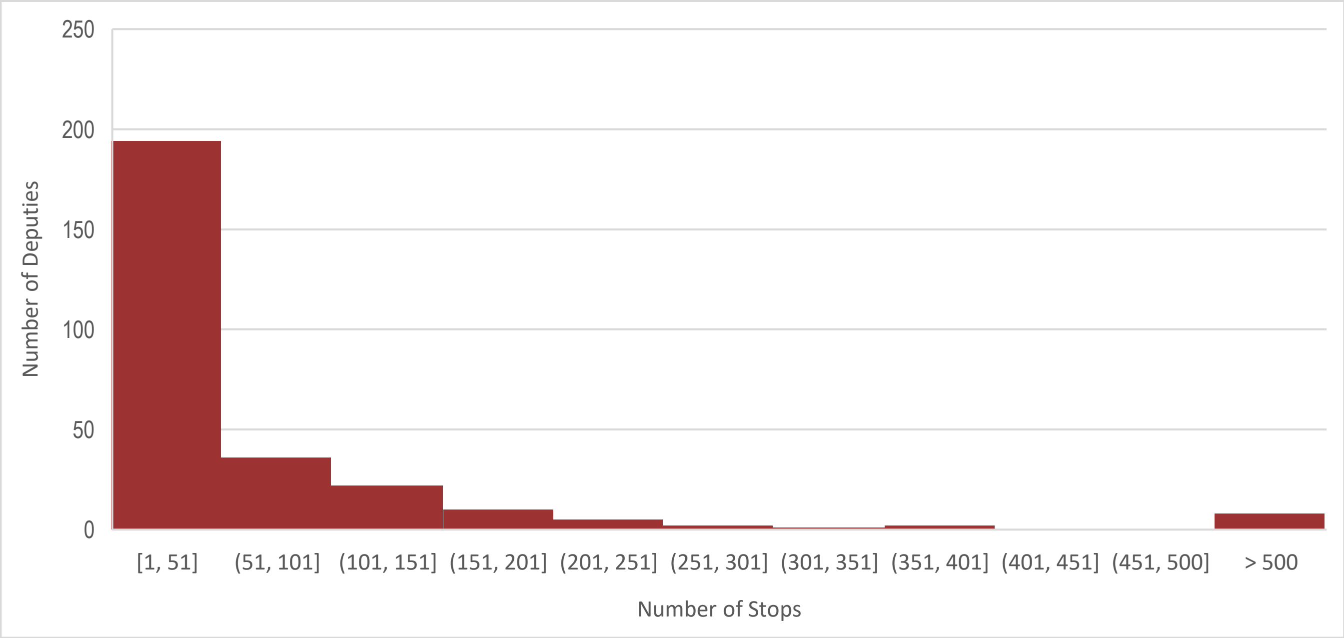
Deputies use the driver arrest variable field to document whether arrests are classified as cite and release (i.e., citation in lieu of detention) or bookings. Depending on the charges against the driver, deputies can use their discretion to choose between the two options. For example, a deputy arresting an individual for DUI may use their discretion regarding whether the individual is too impaired to be released on their own recognizance or to a sober individual, licensed driver, or should be booked for the night. Arrests of drivers are rare among traffic stops; cite-and-release arrests with no custodial element represent 92.57 percent of traffic stop arrests. Overall, cite-and-release arrests (non-booked) represent 4.01 percent of total traffic stops, with less than 1 percent of stops ending in a custodial arrest where the driver was booked (Figure 11).

Figure 11. Arrests during traffic stops

Deputy characteristics

The dataset includes traffic stops made by 280 deputies from the MCSO. We present data about deputy traffic stop activity measured as the total number of stops conducted by deputies over the 12-month period in this analysis. As Figure 12 shows, most deputies conducted between 1 and 51 stops during this period, but a few deputies made more than 500 stops in the same period. This finding is similar to those of previous annual reports.

Figure 12. Deputy traffic stop count (number of stops over the 12-month period)



Comparative analysis

In this section, we present the findings from analyzing each stop outcome, and we summarize the findings from the propensity score matching statistical analysis. This report also includes the following supplemental appendices:

- Supplemental Appendix 1 presents descriptive statistics for all variables.
- Supplemental Appendix 2 includes output from the regularization models creating the propensity scores.
- Supplemental Appendix 3 includes detailed tables of the propensity score matching results.
- Supplemental Appendix 4 provides results from the analyses of stop length that include extended stop indicators.
- Supplemental Appendix 5 provides results from all other alternate specifications.
- Supplemental Appendix 6 provides details on the results of the common support and balance tests for our main models.

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 statistical significance. Given that the sample size for all analyses was more than 100, the critical t-statistic was 1.96 (t-statistics above this value indicate statistical significance, and those below indicate a failure to reject the null hypothesis, or no statistically significant difference).

Common support and balance assumptions were met for all the baseline analyses (see Supplemental Appendix 6 for further details on these tests). In propensity score matching analysis, common support is necessary for valid estimation, meaning that all observations contain a positive probability of being in the condition of interest or not, based on the propensity score (Khandker et al., 2010). Balance evaluates the effectiveness of the matching procedure in reducing observable differences between observations in and out of the condition of interest (Khandker et al., 2010). After matching takes place, the differences between observations in the condition of interest and their matches on the observable characteristics used for matching should be minimal.

Stop length

The analysis team investigated differences in stop length between Hispanic and White drivers, Black and White drivers, and all racial/ethnic minority and White drivers. As noted in the previous section, deputies can indicate whether they experienced specific circumstances that extended the length of a stop beyond their control, including technical issues (e.g., a printer failure), a language barrier, a DUI stop, training, calling for a tow, documentation issues, and other issues. The stop length analyses removed all extended stops based on MCSO's *Traffic Stop Quarterly Reports* (TSQR 3, 13, and 17), which audited extended stop length indicators and found that deputies were using them appropriately.

To provide context and a comparison point, the average stop length for White drivers in 2024 was 14.00 minutes, with a standard deviation of 18.35 minutes (or 18 minutes and 21 seconds). Removing the extended stops, the average stop length for White drivers in 2024 was 10.17 minutes (or 10 minutes and 10 seconds), with a standard deviation of 3.72 minutes (or 3 minutes and 43 seconds). Table 4 summarizes the findings from this analysis. **Our analysis found statistically significant differences in stop lengths between Black drivers and White drivers. The average stop length for Black drivers was 31 seconds longer than the average stop length for White drivers.**[22] The deprecated specifications found in Appendices 3 and 4 resulted in several different findings. Three models found no statistically significant difference in stop length between Black and White drivers while one model identified a statistically significant difference in stop length between minority and White drivers.

Table 4. Propensity score matching results for stop length, extended stops removed

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.23	1.67	No
Black vs. White drivers	0.52	2.32	Yes
Minority vs. White drivers	0.10	0.92	No

Citations

The analysis team investigated differences in citation rates (i.e., the percentage of stops that resulted in citations rather than warnings or incidental contact receipts) between Hispanic and White drivers, Black and White drivers, and all racial/ethnic minority and White drivers. To provide context and a comparison point, 61.59 percent of stops involving White drivers ended in a citation. This year, the analysis included ARS 28-3151A as a matching variable. ARS 28-3151A citations were present in 6.91 per cent of stops of Hispanic drivers but only in 0.54 percent of White drivers. Table 5 summarizes the findings from this analysis. **For similarly situated stops, we found no statistically significant differences in citation rates**. The deprecated specification found in Appendices 3 and 4 identified statistically significant difference in citation rate between Hispanic and White drivers when including special assignments as a matching variable and excluding ARS 28-3151A violations as a matching variable.

[22] Note that the "all racial and ethnic minorities" analysis includes Hispanic, Black, Asian, and Native American drivers. Because of the inclusion of additional racial and ethnic groups, these results may differ from those of individual analysis of Hispanic or Black drivers.

Table 5. Propensity score matching results for citations

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	2.23	1.71	No
Black vs. White drivers	0.38	0.19	No
Minority vs. White drivers	1.50	1.30	No

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 all racial/ethnic minority and White drivers. To provide context and a comparison point, 0.37 percent of stops of White drivers involved a non-incidental search. Table summarizes the findings from this analysis. **We found that, compared to White driver stops, all racial/ethnic minority driver stops were 0.45 percentage points more likely to involve a search not incidental to an arrest or tow.** Deprecated analyses, found in Appendices 3 and 4, identified no statistically significant differences in search rates for any groups.

Table 6. Propensity score matching results for non-incidental searches

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.06	0.26	No
Black vs. White drivers	0.31	0.86	No
Minority vs. White drivers	0.45	2.27	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 all racial/ethnic minority and White drivers. To provide context and a comparison point, 4.78 percent of stops involving White drivers ended in an arrest. Table summarizes the findings from this analysis. **We found statistically significant differences in arrest rates across all three comparisons.** For deprecated specifications, found in Appendices 3 and 4, we identified eight different findings. We found no statistically significant differences, for any group, for booked arrests; no statistically significant differences for all arrests for Black or minority drivers when including special assignments as a matching variable; and no statistically significant differences for all groups in booked arrests, compared to cite and release arrests.

Table 7. Propensity score matching results for arrests

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	3.86	6.51	Yes
Black vs. White drivers	3.08	2.94	Yes
Minority vs. White drivers	3.05	5.30	Yes

Seizures

The analysis team investigated differences in seizure rates predicated on non-incidental searches by the race/ethnicity of the driver. Deputies made 122 stops involving non-incidental searches during the analysis period. Table presents the breakdown of searches with and without seizures by the race/ethnicity of the driver. The chi-square test of homogeneity returned χ^2=7.11, p=0.13, and the Fisher's exact test returned p=0.08, indicating **no**

statistically significant difference in the distributions of searches with and without seizures across driver race or ethnicity. These findings are consistent with those of the previous annual reports.

Table 8. Seizures during non-incidental searches by the race/ethnicity of the driver

Race/ethnicity of the driver	Number of searches	Percentage of searches without seizures	Percentage of searches with seizures
Asian	1	0.00%	100.00%
Black	18	77.78%	22.22%
Hispanic	53	54.72%	45.28%
Native American	2	50.00%	50.00%
White	48	72.92%	27.08%
Overall	122	64.75%	35.25%

CONCLUSION

The MCSO and the CNA analysis team **conclude that there is evidence of disparate stop lengths, search rates, and arrest rates by drivers' race/ethnicity in the 2024 traffic stops in the baseline analysis.** Stops involving Black drivers were more likely to be longer than stops involving White drivers. Although results this year still indicate statistically significant findings in stop length when comparing all Black drivers to White drivers, the MCSO is encouraged that the difference in stop length between the comparison groups is approximately 31 seconds. Stops involving Black, Hispanic, and all racial and ethnic minority drivers were not more or less likely to result in a citation. For search rates, we found only a statistically significant difference when comparing all racial/ethnic minority drivers to White drivers. However, in terms of arrest rates, we found statistically significant differences in all comparison groups.

We identified disparities in some stop outcomes, and these disparities are higher than those found in the 2023 analysis (as indicated in Table 9).[23] Notably, we did not identify statistically significant differences for any group in the baseline analysis for citation rates or seizures following a search. The calculated differences for each year cannot necessarily be assumed to represent statistically significant changes over time; this information is purely descriptive. In Table 9, red checkmarks represent statistically significant differences and green null symbols represent a lack of statistically significant differences between the identified group and White drivers.

[23] Notes on models used for comparisons:

- All models use White drivers as the comparison condition, reflecting the change made for the 2019 analysis.
- All models reflect a correction to the statistical syntax used to classify the time of day and define non-incidental searches of vehicles. The uncorrected syntax was present in the 2017–2018 and 2019 models.
- All models use the matching variables used in the original analysis and the 2017–2018 analysis includes fewer matching variables (see *Maricopa County Sheriff's Office Traffic Stops Analysis Report: January 2019–December 2019* for details on the added variables).
- All stop length models reflect the analysis with extended stops removed, reflecting the change in the baseline model made for the 2021 analysis. This model is found in Table 7 in previous TSAR reports.
- All models for the 2022 and 2023 analyses include the introduction of splining certain variables and regularization.
- Models for the citation outcomes included the use of ARS 28-3151A as a matching variable. Models for the arrest outcomes removed Stop Classification as a matching variable.

Table 9. Comparison of statistical significance and differences across TSARs

Outcome	2017–2018 Finding		2019 Finding		2020 Finding		2021 Finding		2022 Finding		2023 Finding		2024 Finding	
	Stat. sig.	Diff.	Stat. sig.	Diff.	Stat. sig.	Diff.	Stat. sig.	Diff.	Stat. sig.	Diff.	Stat sig.	Diff.	Stat sig.	Diff.
Stop length	H: ✓ B: Ø M: ✓	0.49 min 0.35 min 0.64 min	H: ✓ B: ✓ M: ✓	0.91 min 1.28 min 0.94 min	H: ✓ B: ✓ M: ✓	1.15 min 1.52 min 0.97 min	H: ✓ B: Ø M: Ø	0.99 min N/A N/A	H: ✓ B: ✓ M: ✓	0.48 min 0.72 min 0.83 min	H: Ø B: Ø M: ✓	N/A N/A 0.28 min	H: Ø B: ✓ M: Ø	N/A 0.52 min N/A
Citations	H: ✓ B: Ø M: ✓	2.3% N/A 2.0%	H: ✓ B: Ø M: ✓	4.1% N/A 4.1%	H: ✓ B: Ø M: ✓	4.6% N/A 3.6%	H: ✓ B: ✓ M: Ø	2.6% -7.3% N/A	H: ✓ B: Ø M: Ø	3.7% N/A N/A	H: Ø B: Ø M: ✓	N/A N/A 2.5%	H: Ø B: Ø M: Ø	N/A N/A N/A
Searches	H: Ø B: ✓ M: ✓	N/A 1.7% 0.6%	H: Ø B: ✓ M: ✓	N/A 0.9% 0.4%	H: ✓ B: Ø M: ✓	0.9% N/A 1.0%	H: ✓ B: Ø M: ✓	0.7% N/A 0.5%	H: ✓ B: Ø M: Ø	0.7% N/A N/A	H: Ø B: Ø M: Ø	N/A N/A N/A	H: Ø B: Ø M: ✓	N/A N/A 0.45
Arrests	H: ✓ B: ✓ M: ✓	2.1% 3.7% 2.1%	H: ✓ B: Ø M: ✓	1.6% N/A 1.3%	H: ✓ B: Ø M: ✓	1.5% N/A 1.3%	H: ✓ B: Ø M: ✓	1.3% N/A 1.5%	H: Ø B: Ø M: Ø	N/A N/A N/A	H: Ø B: Ø M: Ø	N/A N/A N/A	H: ✓ B: ✓ M: ✓	3.86 3.08 3.05
Seizures	Ø		Ø		Ø		Ø		Ø		Ø		Ø	

Note: Red checkmarks represent statistical significance and green null symbols represent a lack of statistically significant differences between the identified group and White drivers.

Figure 13 depicts the trend in the difference in the average length of traffic stops across the last six TSARs and this TSAR. This year, we see a decline for both Hispanic drivers and all racial and ethnic minority drivers. The difference in average stop length for Hispanic drivers was approximately 14 seconds and was not statistically significant. For all racial and ethnic minorities, the difference was approximately 6 seconds and was also not statistically significant. The MCSO remains committed to investigating the cause of these disparate outcomes. MCSO has investigated extended stop indicator use for three data years (2020, 2023, and 2024) and published results of those studies that identified that deputies are using the extended stop indicators appropriately. Through *Traffic Stop Quarterly Report 3: Extended Traffic Stop Indicator Use* and *Traffic Stop Quarterly Report 4: Long Non-extended Traffic Stops*, the MCSO identified the need to collect additional data regarding stop length because previous TSARs did not account for unmeasured characteristics of stops. In May 2022, the MCSO began collecting information that identifies delays in traffic stops caused by drivers not having the required driving documentation (i.e., license, registration, and insurance) with them at the time of the stop or when drivers take extra time to find their required documentation. The MCSO also provided an "other" category in which deputies can identify a reason for an extended stop other than those captured by the existing extended stop indicators, which include technical issues (e.g., a printer failure), a language barrier, a DUI stop, training, or having a vehicle towed. The MCSO is committed to identifying and documenting the contributing factors to these differences in traffic stop length and is working to mitigate them when possible.

Figure 13. Difference in the average length of traffic stops by race/ethnicity (compared to White drivers)

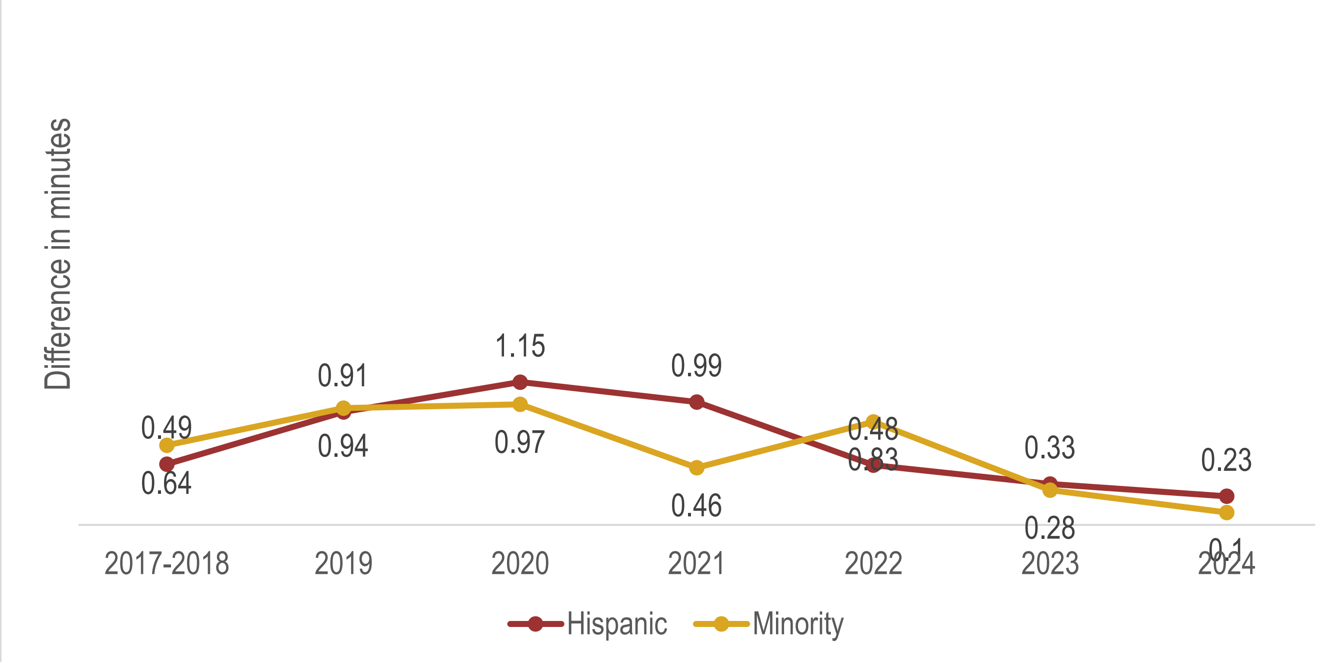


Figure 14 presents a downward trend in the difference in citation rates for Hispanic drivers compared with White drivers, which decreases to a point that is no longer statistically significant. The difference in citation rates increased for all racial and ethnic minorities in from 2022 to 2023, and those differences were statistically significant. The MCSO completed a quarterly report (*Traffic Stop Quarterly Report 6: Citations and Warnings*) analysis to further investigate the cause. Examining the types and number of violations that result in citations and warnings helped the MCSO gain insight into the cause of these disparities and therefore understand how to target efforts to combat them. In 2023, MCSO used the offense type of violations as a matching variable for the first time. Notably, the unmatched citation rates reveal that White drivers were more likely to be cited than any other group; however, once offense types were matched, the direction of the disparity changed, indicating that Hispanic drivers were more likely to receive a citation, though the difference from White drivers was not statistically significant. Based on TSQR 15 results, MCSO has included ARS 28-3151A as a matching variable in the analysis of citation outcomes. Disparity in citation rates were not identified as statistically significant for 2024.

Figure 14. Difference in citation rates by race/ethnicity (compared to White drivers)

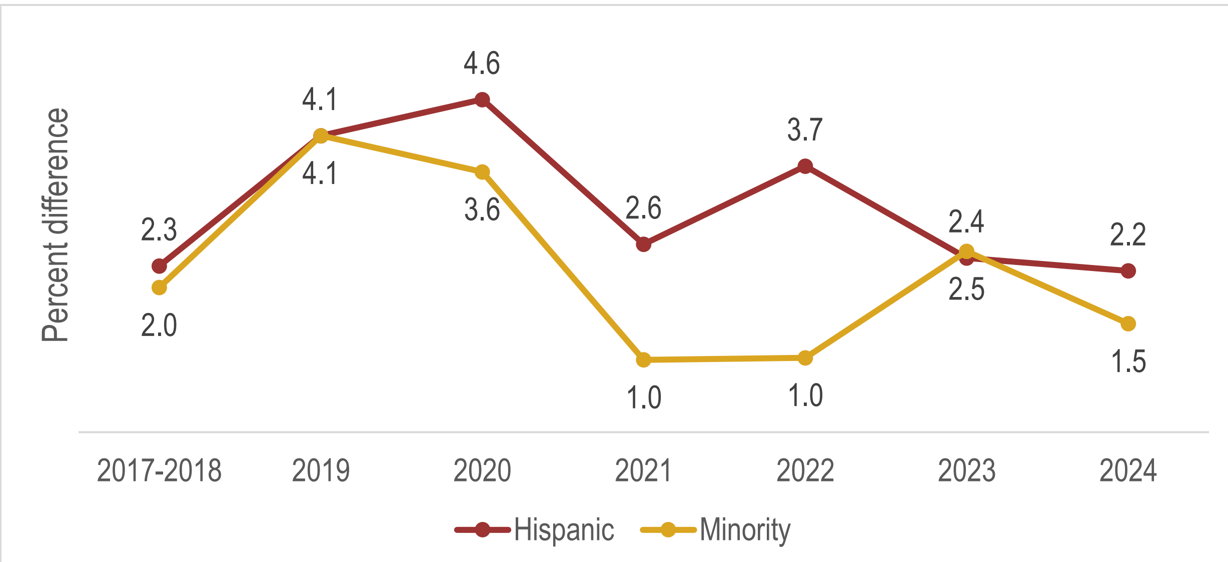


Figure 15 shows the difference between search rates for Hispanic and all racial and ethnic minority drivers compared with White drivers. There were statistically significant differences in search rates between all racial/ethnic minority groups and White drivers in 2024. The differences in non-incidental search rates indicate that racial/ethnic minority drivers were more likely to be searched than White drivers. In addition, statistically significant differences in seizure rates following a search (i.e., hit rates) have not been observed since the TSAR methodology was approved.

Figure 15. Difference in search rates by race/ethnicity (compared with White drivers)

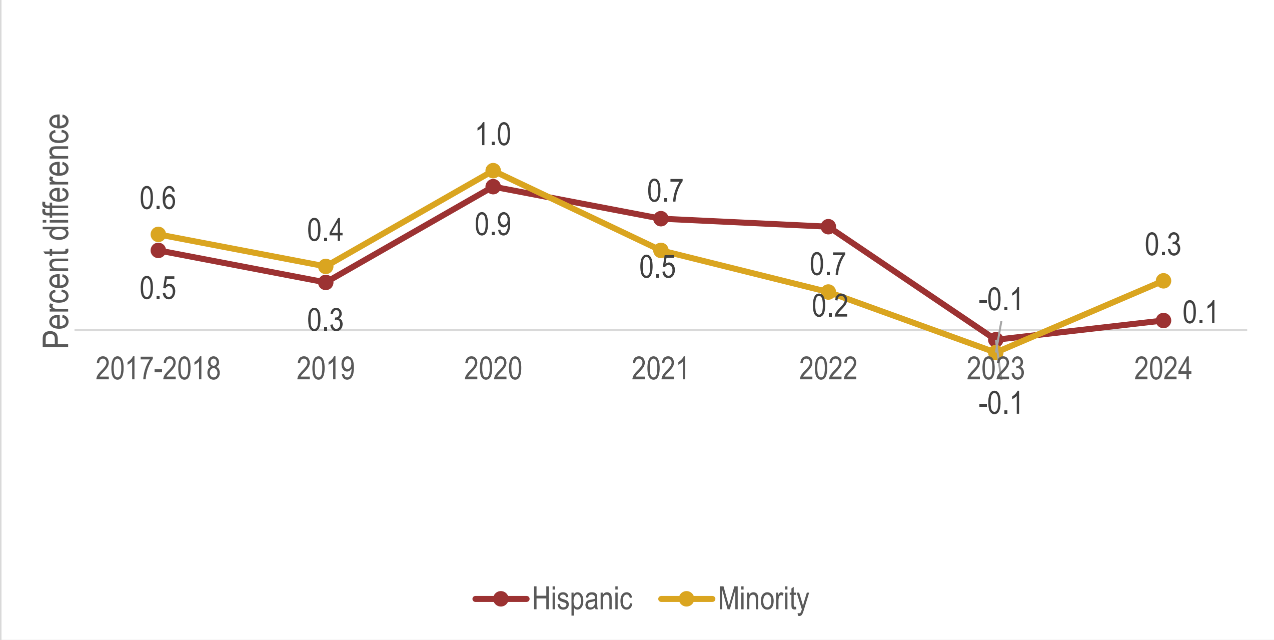
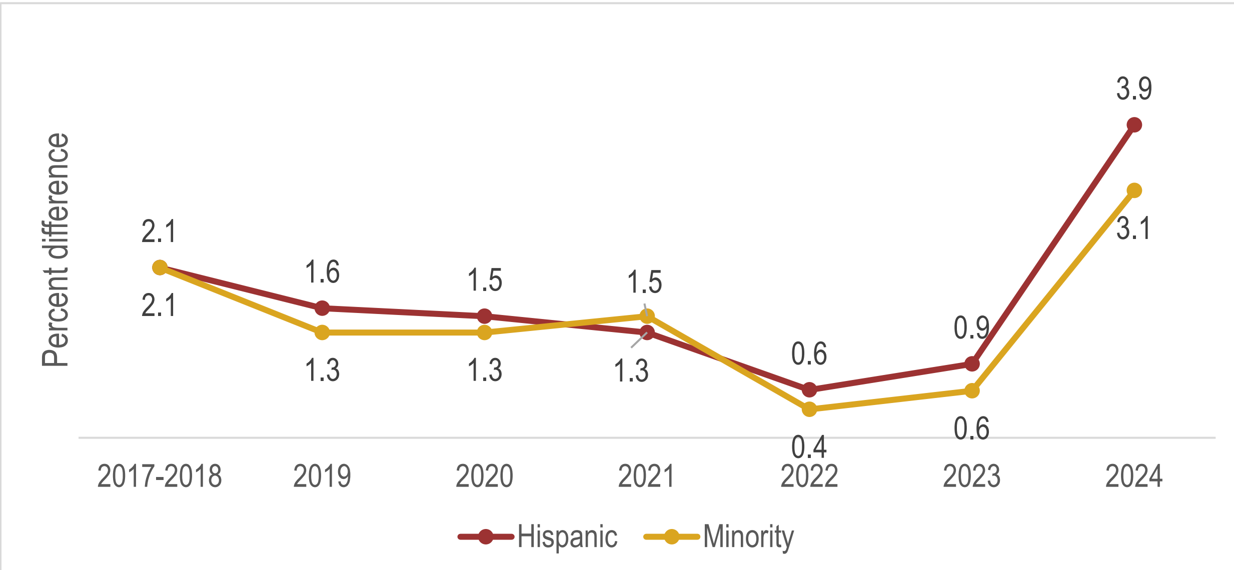


Figure 16 shows the trends in arrest rates for Hispanic and all racial and ethnic minority drivers compared with White drivers. The figure documents the overall downward trend in the difference in arrest rates across the last six TSARs, followed by a spike in 2024. We see an increase in arrest rates from last year in this TSAR, and the differences for all comparison groups are statistically significant. The MCSO remains concerned about any disparate outcomes and recognizes that these changes stem largely from a change in the methodology.[24]

Figure 16. Difference in arrest rates by race/ethnicity (compared with White drivers)



The MCSO is firmly committed to eliminating bias across its operations. *Critical Policy-8: Preventing Racial and Other Bias-Based Profiling* expressly forbids explicit bias. Training on implicit bias has been incorporated into the MCSO's required ongoing training curriculum. MCSO has implemented a program to analyze traffic stop data monthly to look for warning signs or indicia of possible bias-based policing or racial profiling. The monthly analysis of traffic stop data is designed to identify disparities between all racial and ethnic minority drivers (i.e., Asian, Black, Hispanic, and Native American) and White drivers for the length of stop, citation rate, search rate, seizure rate, and arrest rate benchmarks using both comparative and descriptive analyses. The patrol activity of deputies identified by these analyses is reviewed extensively. This process allows the early identification and monitoring of deputies with identified disparities in outcomes across any of the five benchmarks and provides a mechanism for conducting deputy-level interventions when appropriate.

The MCSO remains vigilant and committed to addressing disparities because they indicate possible systemic racial bias and because of how they affect the community. For context, we discuss the statistically significant findings based on propensity score matching estimates here. Black drivers in 2024 experienced a 30-second difference in stop length compared with an average stop length of 10 minutes and 17 seconds for white drivers. All Minority drivers experienced 0.45 percent increase in search rates from the 0.37 percent search rate for White drivers. All three comparison groups (Black, Hispanic, and all Minority drivers) had arrest rates that were 3.08, 3.86, and 3.05 percentage points higher, respectively, than the 4.78 percent arrest rate for White drivers. Although these disparities

[24] In 2024, Criminally Classified stops were not used as matching variable for the arrest benchmark for the first time, as they were all arrests, and therefore the outcome variable as well.

indicate possible systemic bias, in MCSO's patrol function, they may have other causes that are not accounted for in this study, including societal systemic bias, which MCSO cannot control. The results of the TSQR 6 identify potential differential offending that is not accounted for in this or previous annual analyses.

The MCSO continues to be at the forefront of traffic stop analysis and reporting and has already implemented many nationally recognized and recommended strategies to combat disparities in traffic stops (Council on Policing Reforms & Race, 2023). These strategies include the elimination of performance incentives based on "quotas." Officers receive training in procedural justice, and the MCSO has and continues to deploy a survey for community members who have interacted with its deputies that is designed to capture how well deputies adhere to the principles of procedural justice. The MCSO has implemented and uses its early intervention system to track any disparities identified in deputies' traffic stops and conducts interventions when deemed necessary through the *Traffic Stop Monthly Report* and review process. The MCSO produces several reports each year, including the TSAR and the TSQRs, examining disparities in a continual effort to improve disparate outcomes for the community members that it serves. In addition, the MCSO continually evaluates and reviews its policies and procedures. It conducts 19 different inspections and publishes the results to ensure compliance. In addition, the MCSO has implemented training courses over several years designed to improve cultural competencies, reduce explicit and implicit bias, and ensure that bias-based policing does not occur.

Given the cyclical nature of the annual reports (particularly since 2019), many of the actions MCSO takes in response to the TSARs are meant to address each of the benchmarks. Findings of statistical significance may help the MCSO identify which actions and efforts to prioritize, but the MCSO also strives to continually improve results for each of the benchmarks. In the past year, the MCSO has developed a process for responding to each of the traffic analysis reports: First, it publishes a report on its website and then solicits feedback from internal and external stakeholders. The MCSO command then considers whether to implement this feedback, and the MCSO communicates the results via the traffic stop reports page of the MCSO website. The TSQR 12 was the first study to go through this process. MCSO is excited to continue this ongoing feedback process with both internal and external stakeholders as these efforts continue.

In response to the findings from last year's TSAR 9, the MCSO conducted the following activities: communicated results to employees and the public through town halls and its website, provided monitor-approved enhanced training to personnel to address the findings in TSAR 9, developed a real-time information dashboard to help the Motors Unit patrol for traffic safety, developed a traffic stop review dashboard to enable supervisors to review subordinates' stops, examined the impact of ARS 28-3151A on citation rate disparities, examined contracted jurisdictions individually, validated the use of extended stop indicators, analyzed disparities both within and between districts, and continuously explored data to understand disparities and potential ways to mitigate them.

The information in this report builds upon the MCSO's efforts to implement data-driven approaches to improve the effectiveness and fairness of its traffic patrol activity. The Traffic Stop Annual Report does not represent all of MCSO efforts to identify and mitigate disparities. Through our quarterly reports, ongoing Traffic Stop Monthly analyses and reviews of deputy activity, MCSO is able to identify disparities across these benchmarks that are the result of factors other than bias or discriminatory policing. Upon review of statistically significant TSMR findings from ongoing research the Court Monitor has concurred with MCSO that the inequalities in traffic stop outcomes were not associated with bias. Understanding the data and the causes of the traffic stop disparities identified in this report, is part of MCSO's ongoing efforts to identify potential bias, monitor deputy activity, and intervene when necessary.

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.

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.

Baumgartner, F. R., Bell, K., Beyer, L., Boldrin, T., Doyle, L., Govan, L., Halpert, J., Hicks, J., Kyriakoudes, K., Lee, C., Leger, M., McAdon, S., Michalak, S., Murphy, C., Neal, E., O'Malley, O., Payne, E., Sapirstein, A., Stanley, S., & Thacker, K. (2021). Intersectional encounters, representative bureaucracy, and the routine traffic stop. *Policy Studies Journal, 49*(3), 860–886.

Bolas, T. (2022*). Fewer stops, but no less bias: The effect of the 2020 Black Lives Matter protests on traffic enforcement in California.* Johns Hopkins School of Public Health. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4094498

Caliendo, M., & Kopeinig, S. (2005). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Christiani, L., Shoub, K., Baumgartner, F. R., Epp, D. A., & Roach, K. (2022). Better for everyone: Black descriptive representation and police traffic stops. *Politics, Groups, and Identities*, *10*(5), 807–816.

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.

Council on Policing Reforms & Race. (2023). *Recommendation area: Addressing racial disparities in traffic stops*. Retrieved June 13, 2023, from https://www.councilonpolicingreforms.org/addressing-racial-disparities-in-traffic-stops/#recommendations.

Ekstrom, P. D., Le Forestier, J., & Lai, C. K. (2022). Racial demographics explain the link between racial disparities in traffic stops and county-level racial attitudes. *Psychological Science*, *33*(4), 497–509.

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.* 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, R. G., Jr. (2016). An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, https://fryer.scholars.harvard.edu/publications/empirical-analysis-racial-differences-police-use-force

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.

Hannon, L., Neal, M., & Gustafson, A. R. (2020). Out-of-place and in-place policing: An examination of traffic stops in racially segregated Philadelphia. *Crime & Delinquency*, *67*(6).

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.

Iwama, J., & McDevitt, J. (2025). Racial Profiling in Traffic Stops: An Investigation of Race/Ethnicity and Minor Traffic Violations. *Race and Justice*, *0*(0). https://doi.org/10.1177/21533687251346502

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*. The World Bank. https://documents1.worldbank.org/curated/en/650951468335456749/pdf/520990PUB0EPI1101Official0Use0Only1.pdf

Knode, J. L., Carter, T. M., & Wolfe, S. E. (2024). Driving While Broke: The Role of Class Signals in Police Discretion. *Justice Quarterly*, 42(3), 1–30. https://doi.org/10.1080/07418825.2024.2398543

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.

McCabe, J. E., Kaminski, R. J., & Boehme, H. M. (2020). Racial profiling and CT motor vehicle stops: An observational study in three towns. *Police Practice and Research*, *22*(6) 1–18.

McMahon, J., Garner, J., Davis, R., & Kraus, A. (2002). *How to correctly collect and analyze racial profiling data: Your reputation depends on it!* Office of Community Oriented Policing. https://portal.cops.usdoj.gov/resourcecenter/content.ashx/cops-p044-pub.pdf

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*, *43*(2) 207–224.

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

Onookome-Okome, T., Gorondensky, J., Rose, E., Sauer, J., Lum, K., & Moodie, E. E. (2022). Characterizing patterns in police stops by race in Minneapolis from 2016 to 2021. *Journal of Ethnicity in Criminal Justice*, *20*(2), 142–164.

Pearson, K. (1900). 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. *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. *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.

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Ramachandran, V., Phillips, C., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745.

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. https://www.rand.org/pubs/technical_reports/TR534.html

Ridgeway, G. (2009). *Cincinnati Police Department traffic stops: Applying RAND's framework to analyze racial disparities*. RAND. https://www.rand.org/pubs/monographs/MG914.html

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. (2009). *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.

Roach, K., Baumgartner, F. R., Christiani, L., Epp, D. A., & Shoub, K. (2022). At the intersection: Race, gender, and discretion in police traffic stop outcomes. *Journal of Race, Ethnicity, and Politics*, *7*(2), 239–261.

Rodriguez, D., Kunard, L., Johnson, W., LaRochelle, J., & Thorkildsen, Z. (2015). *Assessment report on the Fayetteville (North Carolina) Police Department*. Washington, DC: Office of Community Oriented Policing Services.

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*. 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.

Shoub, K. (2025). Driving While Different: Examining Traffic Stop Outcomes Experienced by Historically Marginalized Communities. *The Journal of Politics*, 87(1)

Simoiu, C., Corbett-Davies, S., & Goel, S. (2017). The problem of infra-marginality in outcome tests for discrimination. *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, Ohio State University.

Stelter, M., Essien, I., Sander, C., & Degner, J. (2022). Racial bias in police traffic stops: White residents' county-level prejudice and stereotypes are related to disproportionate stopping of Black drivers. *Psychological Science*, *33*(4), 483–496.

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.

Vito, A. G., Griffin, V. W., Vito, G. F., & Higgins, G. E. (2020). "Does daylight matter"? An examination of racial bias in traffic stops by police. *Policing: An International Journal, 43*(4), 675–688.

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

Webb, P., Rodriguez, J. J., & Vasquez, A. (2021). Stigmatization and stops: Power threat theory and traffic stop statistics in North Carolina.

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.

Zhang, Y., & Zhang, L. (2021). Racial characteristics of areas and police decisions to arrest in traffic stops: Multilevel analysis of contextual racial effects. *Policing: An International Journal*.

Appendix B. Acronyms

Acronym	Definition
ARS	Arizona Revised Statutes
ATT	average treatment effect on the treated
DUI	driving under the influence
MCSO	Maricopa County Sheriff's Office
TraCS	Traffic and Criminal Software
TSAR	Traffic Stop Annual Report
TSQR	Traffic Stop Quarterly Report
TSMR	Traffic Stop Monthly Report

This page intentionally blank.

3003 Washington Boulevard, Arlington, VA 22201

www.cna.org | 703-824-2000

CNA is a not-for-profit research organization that serves the public interest by providing in-depth analysis and result-oriented solutions to help government leaders choose the best course of action in setting policy and managing operations.

CNA