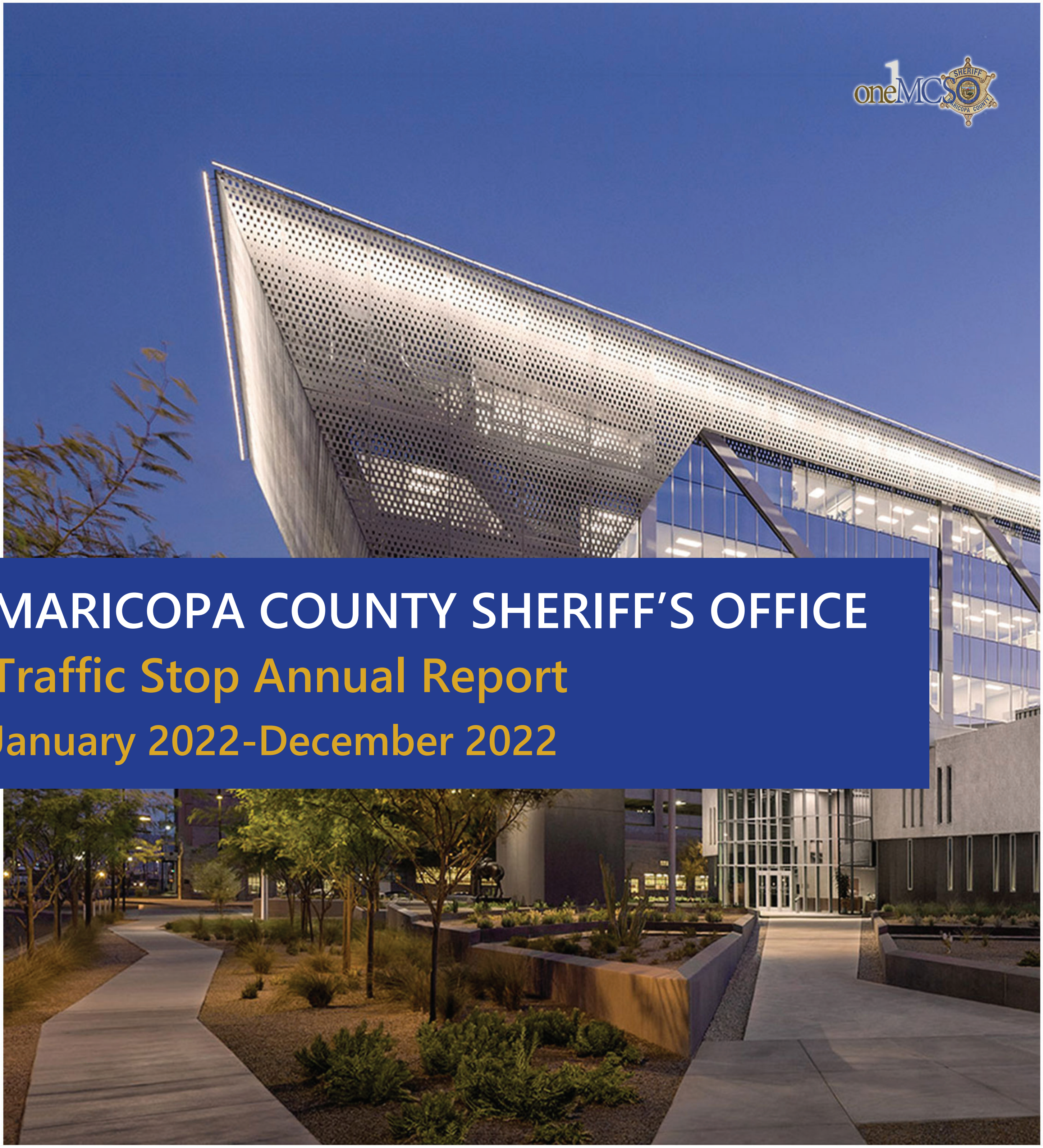


MARICOPA COUNTY SHERIFF'S OFFICE

Traffic Stop Annual Report

January 2022-December 2022

CNA

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

CNA Analysis Team:

Zoë Thorkildsen, Project Director; Jennifer Lafferty, Analyst; Bryce Peterson, Analyst; L. Cait Kanewske, Analyst

Suggested citation:

Thorkildsen, Z., Lafferty, J., Peterson, B. & Kanewske, L. C. (2023). Maricopa County Sheriff's Office Traffic Stop Annual Report: January 2022–December 2022. Phoenix, AZ: Maricopa County Sheriff's Office.

Distribution

June 2023

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] Since 2014, the MCSO has worked toward achieving compliance with three 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 three related court orders, respectively titled the First Order, Second Order, and Third Order. As a feature of the First Order, the MCSO must conduct organizational- and individual-level analyses of patrol activity to determine whether racial or ethnic disparities exist in MCSO traffic stop outcomes.

In November 2016, Paul Penzone was elected as Maricopa County sheriff and took office in January 2017. In 2018, the MCSO contracted with the CNA Center for Justice Research and Innovation to analyze patrol activity on an annual and monthly basis and support the development of quarterly reports on special topics related to traffic stops.

This report examines patterns of patrol activity within the MCSO; it does not analyze or identify individual deputies.[2] The analysis in this report includes all traffic stops made by MCSO deputies from the start of January 2022 through the end of December 2022. The MCSO expects to use this report 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 or 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. To analyze the stop length, searches, citations versus warnings, and arrests, we used propensity score matching. *Propensity score matching* is a quasi-experimental method of statistical comparison that identifies the most similar events in a condition of interest—in this case, Hispanic, Black, or all racial and ethnic minority drivers[3] compared to White drivers—using a propensity score. To analyze seizure rates during searches, we used chi-square testing, which examines whether the racial or ethnic distribution of searches that result in seizures is different from the racial or ethnic distribution of searches that do not result in seizures.

Over the 12-month period from January 2022 to December 2022, MCSO deputies performed 19,787 traffic stops. The rate of traffic stops per month were relatively steady throughout the year. The total number of stops increased about 17 percent during 2022 compared to 2021. Within the 19,787 traffic stops, deputies perceived 65 percent of drivers as White, 24 percent as Hispanic, and 7 percent as Black. The remaining 4 percent of stops involved individuals from other historically marginalized groups, including Asian and Native American individuals.

---

[1] Much of the material in this section is identical to the executive summary from the *Maricopa County Sheriff's Office Traffic Stops Analysis Report: January 2021–December 2021*.

[2] The MCSO analyzes individual deputies' traffic stops through the Traffic Stop Monthly Report (TSMR) and Review.

[3] 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.

In the dataset, approximately 87 percent of the stops that deputies made ranged from 5 to 19 minutes. Approximately 52 percent of stops ended with a citation, 48 percent ended with a warning, and 5 percent ended with an arrest (arrests can occur in the same stop as a citation or warning), with 4 percent being cite-and-release arrests. Less than 1 percent of stops resulted in a non-incidental search of a driver or vehicle, meaning the search was a discretionary decision by the deputy.

The MCSO and the CNA analysis team conclude that there is evidence of disparate outcomes by driver race or ethnicity in traffic stops on many stop outcomes. This finding is consistent with past studies of traffic stop outcomes in other agencies (as noted in this report's introduction), as well as previous court-ordered traffic stop analyses within the MCSO. Stops involving Hispanic drivers were more likely to be longer and to result in an arrest or search than stops involving White drivers. This year, the raw citation rate for Hispanic drivers was lower than it was for White drivers; however, when matched on the characteristics of the stop, the calculated rate of citations was higher for Hispanic drivers. Stops involving Black or all racial and ethnic minority drivers were more likely to be longer, but not more or less likely to end in a citation, search, or arrest than stops involving White drivers. These disparities represent potential indicia of bias as described in the First Court Order. The MCSO has been striving to eliminate disparities in traffic stop outcomes through the use of the *Traffic Stop Monthly Reports*, which identify individual deputies with the most disparate outcomes and intervene when a possible indication of bias exists, as well as the *Traffic Stop Quarterly Reports*, which focus on specific areas with identified disparities to develop an actionable response. Additionally, the MCSO is continually evaluating policies and procedures and conducts 19 different inspections to ensure compliance. Finally, the MCSO is constantly providing ongoing training designed to combat bias. 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 worked collaboratively to collate the data for this analysis, address missing values and other data irregularities, analyze traffic stops outcomes, and develop the *Maricopa County Sheriff's Office Traffic Stop Annual Report: January 2022–December 2022*. The MCSO had primary responsibility for collating data and adjudicating missing values and data irregularities, as well as reviewing the annual report. The CNA analysis team had primary responsibility for developing and executing the analytical plan and authoring the annual report. The MCSO then had primary responsibility for drawing conclusions from the analytical results and developing an appropriate response plan to address findings. The analytical plan was developed collaboratively among CNA, MCSO, the Monitoring Team, and Parties.

In addition to other research on traffic stop activity (e.g., the *Traffic Stop Monthly Reports* and the *Traffic Stop Quarterly Reports*), the MCSO will use this report to better understand its traffic stop activity and better serve the residents of Maricopa County. The MCSO and CNA will continue to work closely to analyze traffic stop activity by MCSO deputies, including developing additional annual analysis reports, monthly analysis reports focused on individual deputies, and quarterly reports on special topics as determined by the MCSO, CNA, and the Monitoring Team in consultation with the Parties.

INTRODUCTION

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.[4] In 2016, the residents of Maricopa County elected Sheriff Paul Penzone to lead the office, which includes more than 3,000 employees and provides enforcement and detention services to the more than four million residents of Maricopa County. The MCSO operates the Fourth Avenue, Durango, Estrella, Lower Buckeye, and Towers jails; the Intake, Transfer, and Release facility; and smaller temporary holding facilities in district substations. The MCSO provides patrol and investigative operations for the six patrol districts of the county, which include an array of businesses, residents, recreational areas, and communities. Additionally, 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. Penzone*), 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 Maricopa County Sheriff's Office to address the pattern of disparate treatment of Hispanic community members in Maricopa County. The First Order established actions required for the MCSO to attain compliance, including introducing new data collection and analysis requirements, training, and policies. The Court appointed an independent monitor for oversight.[5] 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 on an annual basis and produce quarterly reports on special topics related to traffic stops.

This report directly responds to the First Order requirement to analyze the MCSO traffic stop activity to determine whether disparate outcomes exist by the race or ethnicity of drivers. This approach relies on propensity score matching to compare stops that had similar characteristics other than the perceived race or ethnicity of the driver. This report examines patterns of patrol activity within the MCSO; it does not analyze or identify individual deputies. The MCSO expects to use this report as a knowledge base of traffic stop activity in the organization and as a guide for 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 or ethnic inequality related to MCSO's mission.

---

[4] Much of the material in this introduction is identical to the introduction in the *Maricopa County Sheriff's Office Traffic Stops Analysis Report: January 2021–December 2021*

[5] 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. The Third Order does not establish requirements related to traffic stops.

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 or ethnicity in interactions between the police and community members. The interactions under scrutiny include officer-involved shootings, use of force, searches, and traffic stops (see, for example, Baumgartner et al., 2021; Correll et al., 2007; Fridell & Lim, 2017; Fryer, 2016; Ridgeway, 2006; Ritter, 2017; Shoub, 2021). Although most law enforcement officers do not intentionally practice biased policing, they may exhibit behaviors that appear biased or that result from implicit bias (Bolas, 2022; Ekstrom et al., 2021; 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 & Dunham, 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 Penzone leadership, all allegations of explicit bias are taken seriously, investigated thoroughly, and discipline (including termination) is meted out as quickly as possible.

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

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 (Riley et al., 2005; Ridgeway, 2006; Tillyer et al., 2010). Methods for assessing disparities have evolved to incorporate measures beyond stop rates, focusing on stop outcomes such as citations and searches (Christiniani 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., 2007).

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 analysis studies whether the race or ethnicity of the driver affects stop rates; post-stop analysis studies whether the race or 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, since it includes non-drivers and may not accurately reflect the driving population or the racial or ethnic distribution of drivers who violate traffic laws (McMahon et al., 2002; Tregle et al., 2019). Other methods for estimating the racial or ethnic distribution of the driving population, such as observing and recording the race or ethnicity of drivers in a given jurisdiction over time or using driver license race or ethnicity data, can be cost-prohibitive or infeasible because of data unavailability (Fridell, 2004; Tillyer et al., 2010), which is the case in Arizona because race or ethnicity is not captured in Arizona driver license or registration documentation.

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

Practitioners and consumers of bias research should understand that disparate outcomes do not definitively indicate bias (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 as a whole.

A greater number of law enforcement agencies now analyze their traffic stop data internally or in partnership with external researchers and analysts. The majority of analyses conducted to date have found racial or ethnic disparity in traffic stop outcomes (e.g., searches, citations, and arrest). Tillyer et al. (2010) states, "Analyses of these data demonstrate a relatively consistent trend of racial/ethnic disparities in vehicle stops and vehicle outcomes." The majority of existing studies have 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., 2012; Hannon et al., 2020; Norris et al., 1992; Novak, 2004; Pierson et al., 2020; Rodriguez et al., 2019; Rojek et al., 2004; Roach et al., 2022; Rosenfeld et al., 2012; 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 (Groggery & Ridgeway, 2006; Higgins et al., 2012; Taniguchi et al., 2016; McCabe et al., 2020; 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 pragmatically and proactively address disparate outcomes by promoting anti-bias policy, training, and practices. MCSO currently combats bias-based policing through ongoing enhanced trainings, continual policy reviews and revisions, a series of inspections that includes a statistical review of every deputy's traffic stops each month, and the use of an early intervention system to assist supervisors in identifying and intervening on potentially problematic behaviors. Additionally, MCSO has a Community Outreach Division that is actively involved in community policing and recruitment efforts in the communities MCSO serves.

Overall, the use of statistical analysis for identifying racial and ethnic disparities in traffic stops is 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 actionable findings that will address those disparities.

Organization of this report

This report is organized into four sections: introduction, approach, findings, and conclusion. The approach section explains the MCSO and CNA's methods for analyzing traffic stop outcomes and developing this report. The findings section details the results of the traffic stop analysis on the selected outcomes. Finally, the conclusion section reviews the analytical findings, discusses ongoing and future activities 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 list of abbreviations.

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

Approach

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

Overview of data and variables

The MCSO uses TraCS to capture data from the field about traffic stops. TraCS is a data collection, records management, and reporting software for public safety professionals. Deputies use TraCS to document aspects of traffic stops, including driver and vehicle characteristics and activities that occur during the stop. TraCS captures the start time, end time, and geolocation for each traffic stop.[6] The system also requires the deputy to enter variables such as the perceived race or ethnicity of the driver,[7] the contact conclusion, whether an arrest took place, and search and seizure information. TraCS also includes data fields 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.[8] After the deputy fills out information about events in TraCS, the system forwards entries for supervisory review and, if necessary, revision. Of the several hundred variables available through TraCS, we used a subset to analyze racial or ethnic disparities in stop outcomes, as well as construct and append data using variables present in TraCS and other MCSO systems. Here, we briefly describe the variables 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 or 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 reason for the stop, 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

[6] 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.
[7] Note that Arizona does not collect information about race or ethnicity as part of its driver license system; thus, all race or ethnicity categories within the TraCS data are based on the perception of the deputy who made the stop.
[8] 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.

without their insurance card verifying their 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, supervisor, and other. 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.[9] 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 prior to custodial arrest and all vehicles be inventoried prior to tow; searches that occur incident to arrests or tows are not discretionary and thus were excluded from our analysis of outcomes. Deputies also indicate in TraCS whether a search resulted in the seizure of contraband.

Data about the driver. We used the post-stop perceived race or ethnicity of the driver, as entered by the deputy, to classify the driver's race or 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 unknown sex drivers collapsed as the comparison category). We also included the reported license plate state of the vehicle 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 time frame of this analysis, the special assignments included DUI Task Force, Click-It-or-Ticket, and Aggressive Driver Task Force. The analysis team also constructed a deputy traffic stop count variable equal to the number of stops the deputy made over the 12-month period, for descriptive purposes.

Data verification and missing data

The analysis team reviewed the 2022 TraCS data for data quality (e.g., missing data or out-of-range values) and verification.

The analysis team identified additional missing data that the MCSO could not adjudicate or impute. For example, 12 stops lacked data for the vehicle license plate; we omitted these from all comparative analyses, since we used in-state plate status as a propensity score matching variable. These missing data represent less than 1 percent of the overall data, below 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; since 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, and driver's first and last name, and we removed them.

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

Methodology

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

Propensity score matching is a quasi-experimental method of statistical comparison. Researchers use quasi-experimental methods in circumstances in which random assignment (i.e., experimental approaches) 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 at the end of this paragraph). Specifically, propensity score matching identifies the most similar events in or not in a condition of interest (in this case, Hispanic, Black, or all racial and ethnic minority drivers[10]) using a propensity score (Rosenbaum & Rubin, 1983; Apel & Sweeten, 2010).

For this traffic stops analysis, we used the first stage of propensity score matching to determine the probability that a stop involved a driver of a particular race or 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), 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 the stop was extended for one of the reasons discussed above, and the call sign category the deputy was operating under.[11,12] 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. 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.[13]

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 a large number of features, or for models with a number of features that are strongly correlated, which is the case here. For logistic regression models, there are 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 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

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

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

[12] 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.

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

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. For each candidate setting, the model is fit five times. In each case, one fold is withheld from the dataset, the model is trained on the other 4 folds, and then evaluated on the withheld fold. The objective function 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 describes common support and results from common support testing in more detail. To check the robustness of our results, we ran each analysis using radius matching (in which stops in the condition of interest are compared with all stops within a certain propensity score range that are not in the condition of interest) using multiple radii values. Finally, we also used nearest N-neighbor matching (in which stops in the condition of interest are compared with the nearest N stops by propensity score that are not in the condition of interest). We also considered matching with and without replacement as a sensitivity check. Supplemental Appendix 6 presents detailed results from the robustness check analyses.

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

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

Alternate specifications

As noted above, we varied the propensity score approach for the propensity score matching analyses to encompass two matching methods: radius and neighbor. We also varied the parameters used for the radius caliper size and the number of neighbors matched. Finally, we considered the effect of allowing replacement (i.e., whether an observation can be used as a match for multiple other observations) for nearest neighbor and radius matching.[14] The Supplemental Appendices to this report present the results from the alternate specifications.

For the length of stop analysis, we also considered an alternate specification in which we added controls for extended stop indicators to estimate the average treatment on the treated, as well as a model in which we used the extended stop indicators as matching variables. Including control variables in the second stage of the propensity score matching analysis is feasible only when nearest neighbor matching is used; therefore, we present findings from that specification for only these alternate specifications.

For the arrest analysis, we also considered alternate specifications related to the type of arrest: warrant arrests, cite-and-release arrests, and booked arrests. These alternate specifications are presented in the findings section.

Considerations and limitations

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

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

[14] 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 or 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 or 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 or 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.

Below, we present the findings from the comparative propensity score matching and chi-square test of homogeneity. 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 and 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.

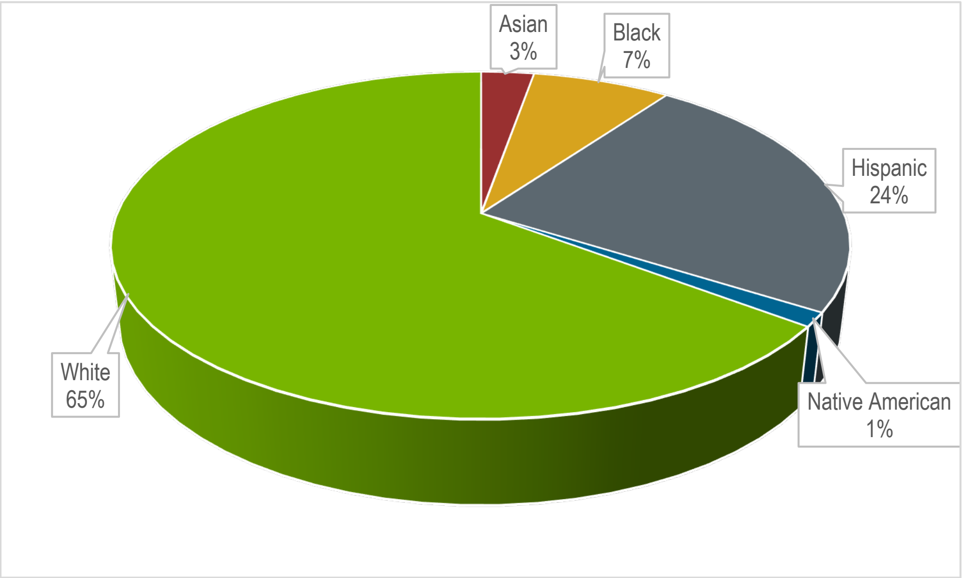
Descriptive statistics

In this section, we describe the data included in this analysis of traffic stops conducted by the MCSO between January 2022 and December 2022 (a 12-month period). We present the characteristics of the stops themselves, 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

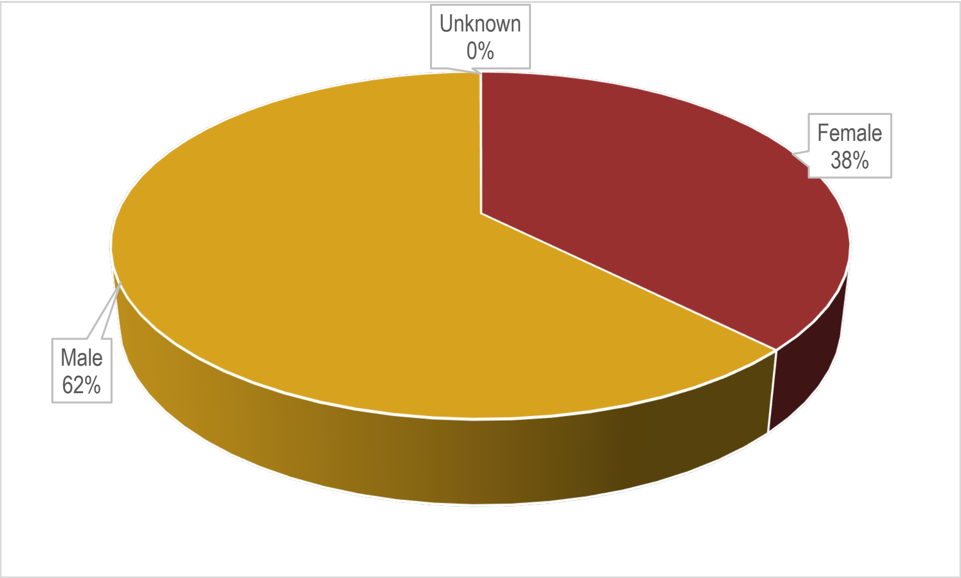
When deputies make a traffic stop, they document their observation of the perceived race or ethnicity of the driver both pre- and post-stop in TraCS. We omitted analysis of the pre-stop perception of driver race or ethnicity since this variable takes the value "unknown" in over 98 percent of stops. Post-stop, deputies perceived 65 percent of drivers as White, 24 percent as Hispanic, and 7 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. The drivers stopped were 62 percent male and 38 percent female, plus seven stops (less than 1 percent) for which the deputy could not determine the sex of the driver (Figure 2).

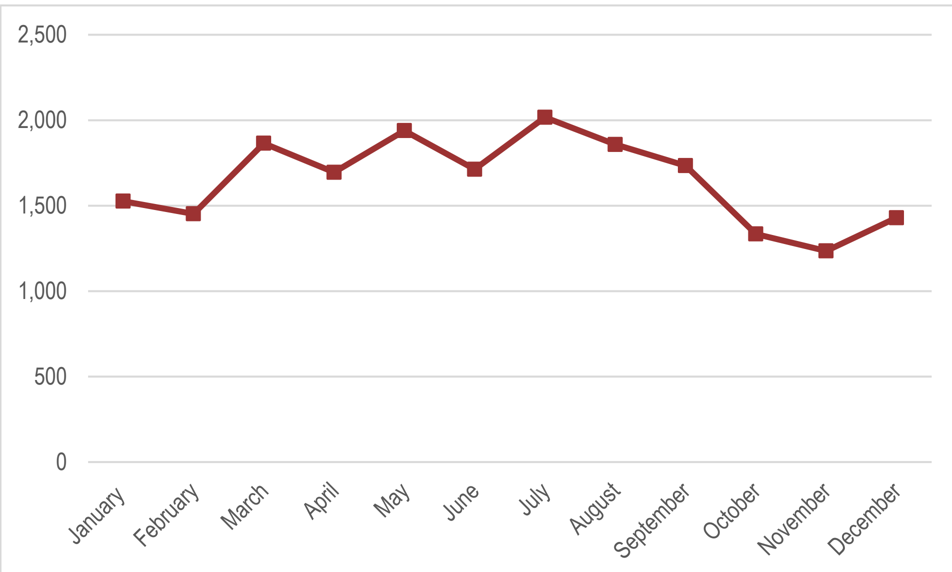
Figure 2. Stops by post-stop perceived driver sex



Stop characteristics

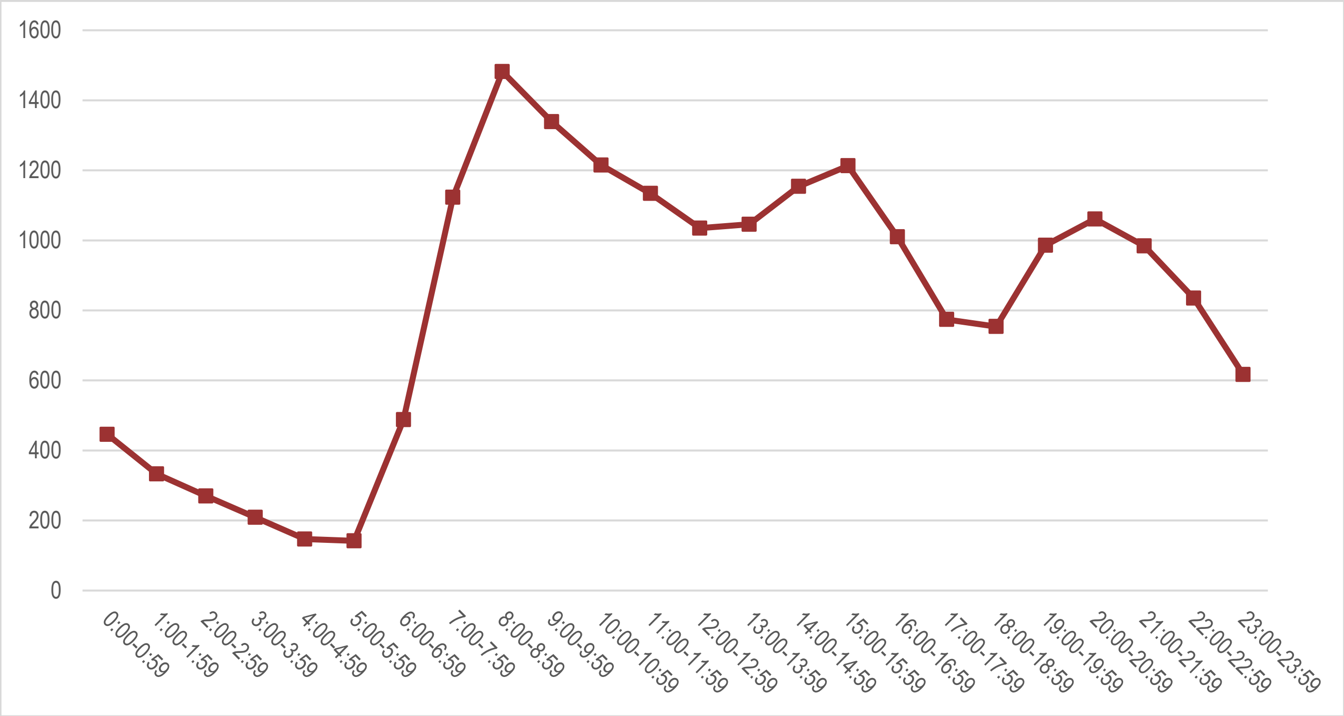
Over the 12-month period for this analysis, the MCSO deputies performed 19,797 traffic stops. Traffic stops over this period exhibited an upward trend between February and March 2022 and again in July 2022. Traffic stops decreased from July 2022 to November 2022 before increasing back to a number of stops that was similar to January 2022 in December 2022. This overall trend is different from what was observed in late 2018 and throughout 2019, when there was a steady upward trend of stops. Additionally, this trend is different from what was observed in 2020, when there was a downward trend at the beginning of the year with more stable rates for the remainder of the year. There was an increase of approximately 17 percent in stops during 2022 compared to 2021 (Figure 3).

Figure 3. Stops by month, January 2022 to December 2022



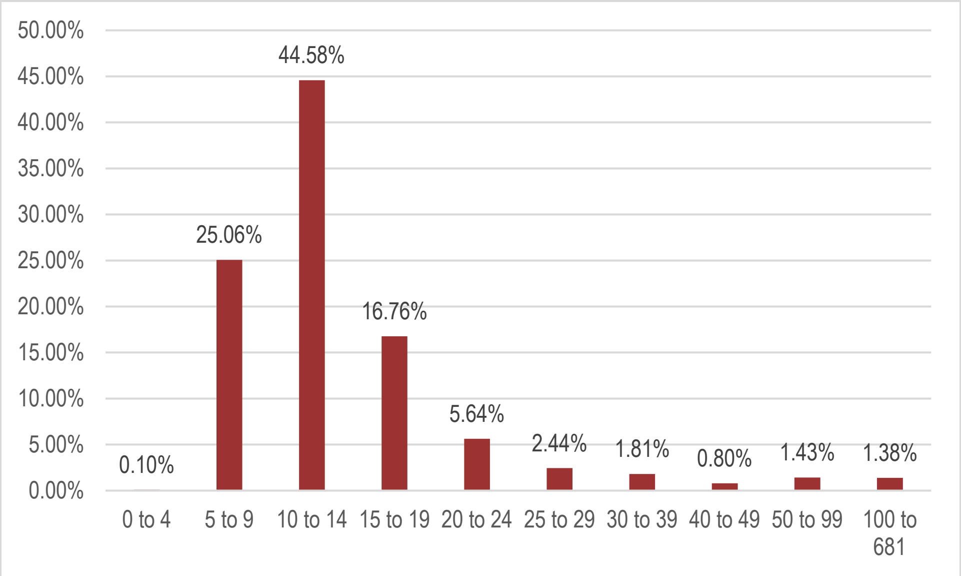
We also considered the time of day that a stop took place. The majority of the stops occurred between 7:00 a.m. and 5:00 p.m., which is similar to trends in previous years (Figure 4).

Figure 4. Stops by time of day



Stop length is of particular importance to this analysis because it is a core aspect of the Court Order. Stops lasted an average of 16.46 minutes (with a standard deviation of 24.90 minutes), a 7.16 percent decrease from the 2021 annual report (in which the average stop length was 17.73 minutes). The majority of stops lasted between 5 and 19 minutes (Figure 5).

Figure 5. Stops 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 stop, vehicle towed, driving-related documentation issues, and other issues. Deputies selected extended stop indicators for 5,366 stops, representing 27.1 percent of total stops. This represents an increase from the previous annual report, in which extended stops represented 20.4 percent of total traffic stops. Documentation issues occurred the most at 13.89 percent, while the next most frequently documented reason was technical issues at 6.91 percent. This finding is different from the previous year's report, in which training stops and technical issues had the highest percentages; however, this year's report includes two new extended stop reasons: documentation issues and other issue (Table 1).

Table 1. Extended stop reasons

Reason indicated	Percentage of stops
DUI Stop	1.94%
Language Barrier	1.96%
Technical Issue	6.91%
Training Stop	5.91%
Vehicle Towed	1.73%
Documentation Issue	13.89%
Other Issue	3.35%

We also considered which stops occurred while the deputy was on a special patrol assignment. Of the 19,797 stops in the dataset, 1,231 stops occurred while deputies were on DUI Task Force assignment, 459 occurred while deputies were on the Aggressive Driver Task Force, and 9 stops occurred while deputies were on the Click-it-or Ticket Task Force (Table 2).

Table 2. Stops conducted during special assignments

Special assignment	Counts
DUI Task Force	1,231
Aggressive Driver	459
Click-It-or-Ticket	9

Stop outcomes

Contact conclusion documents the outcomes from each stop. Of all traffic stops, almost 52 percent concluded with a citation, almost 48 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). This year, White drivers were cited at higher rates than any other group (Table 3).

Figure 6. Traffic stop contact conclusions

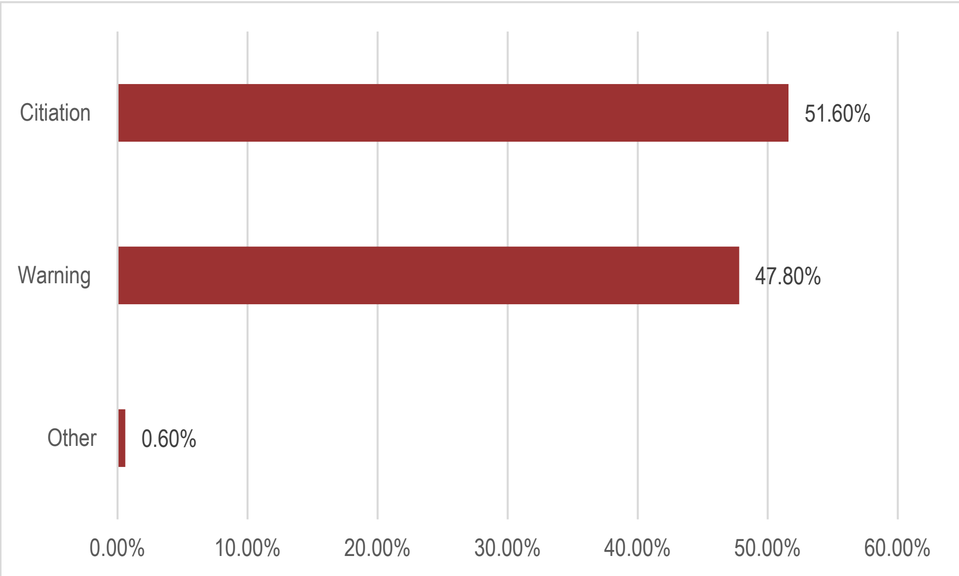


Table 3 presents the stop outcomes when considering the race or ethnicity of the driver.

Table 3. Traffic stop contact conclusions by race and ethnicity

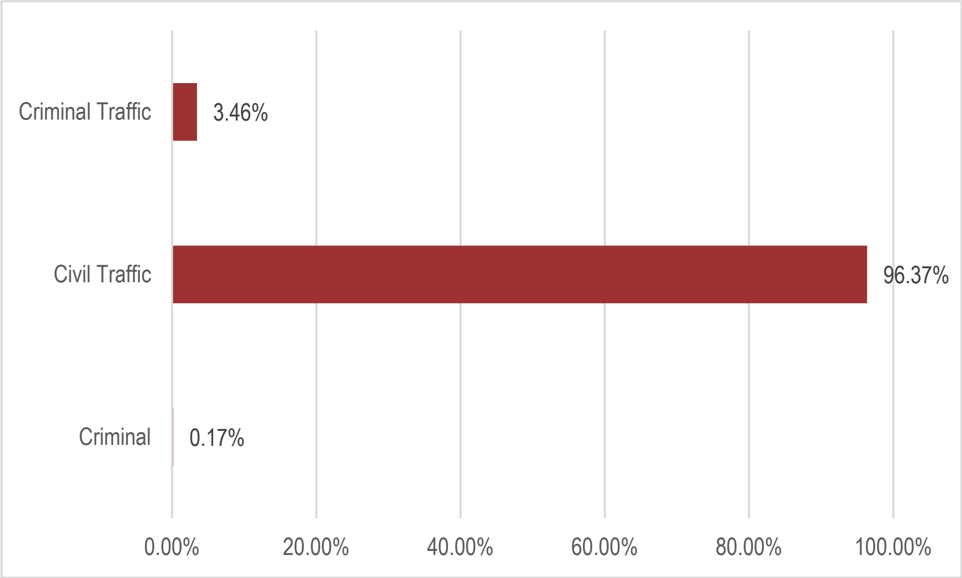
Race and ethnicity	Citation	Warning	Other
White drivers	52.93%	46.55%	0.52%
Hispanic drivers	50.24%	49.00%	0.76%
Black drivers	46.33%	53.04%	0.63%
Minority drivers	49.17%	50.07%	0.76%

The MCSO organizes stops into five categories based on ARS code[15]: civil traffic, criminal traffic, petty civil and criminal. This year, only three categories of stops occurred during traffic stops: civil traffic, criminal traffic, and criminal. Civil traffic stops comprise the majority of stops that occurred. Civil traffic violations include violations in which the driver does not face jail time and instead may be subject to a fine. Examples of these include speeding, equipment violations, or seatbelt violations. Criminal traffic violations are traffic violations that result in a fine and involve possible jail time. These include criminal speeding, reckless driving, driving under the influence, or driving on a revoked or canceled license. Petty violations are criminal violations with less severe penalties that do not include the possibility of jail time. These include boating violations, park violations, and curfew violations. Criminal violations are non-traffic violations that involve possible jail time and typically are incident to the traffic stop, such as stopping an individual with an active warrant for criminal activity or identifying criminal activity not related to

[15] The Arizona Revised Statutes

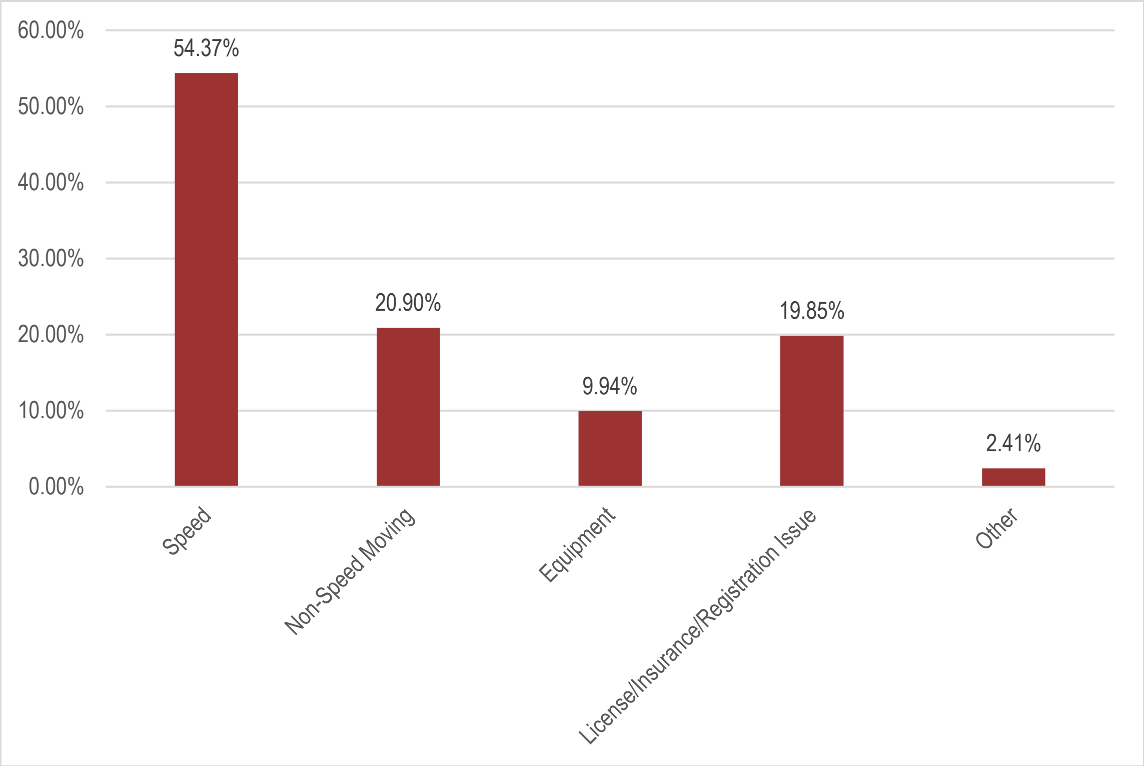
the stop. Of the traffic stops that result in a citation, over 96 percent result in a civil traffic classification, and approximately 3 percent result in a criminal traffic classification. The dataset contains 33 criminal classifications, accounting for less than 0.2 percent of the stop classification reasons.

Figure 7. Traffic stop classifications



MCSO categorized stops into five violation categories: speed, non-moving speed, 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., speeding, criminal speeding, speeding in a school zone, racing, or reckless driving). Non-speed moving violations included violations for which the vehicle was moving, such as turning, failure 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 included any violation 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, illegally modified vehicles, and opacity on window tint). Driving documentation violations included any violation associated with licensing (vehicle or driver), insurance, and registration. Examples include driving without a license, driving on a suspended/revoked license, expired registration, failure to possess insurance, or driving without license plates. 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, or littering. Speed violations were the most common violations, followed by non-speed moving offenses (Figure 8).

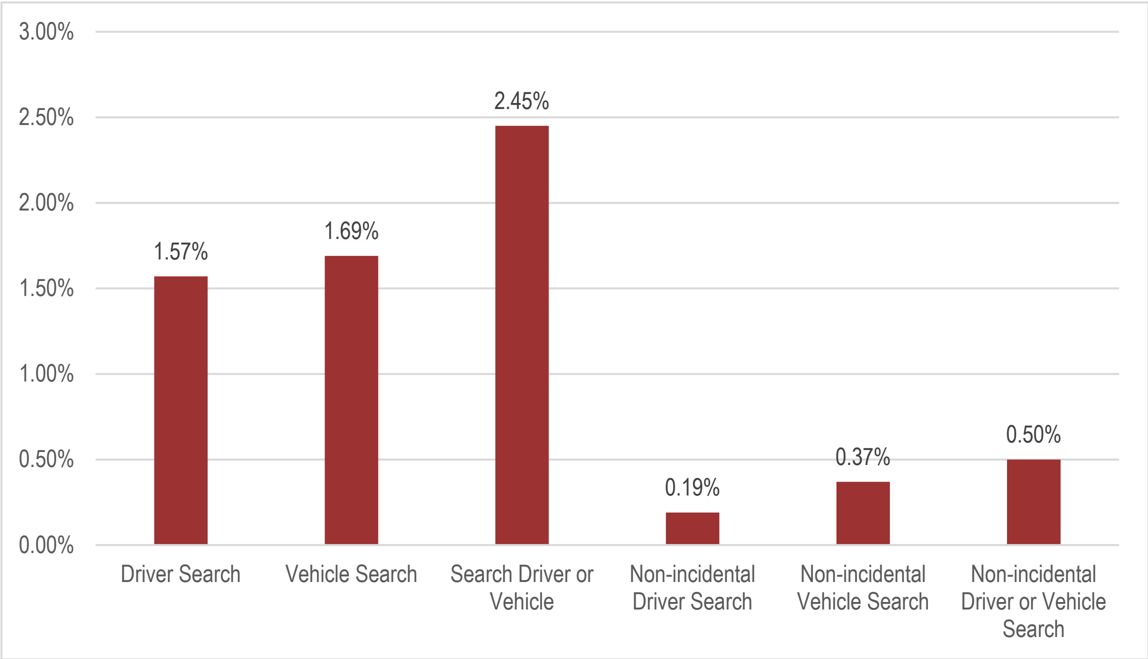
Figure 8. Violation categories[16]



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

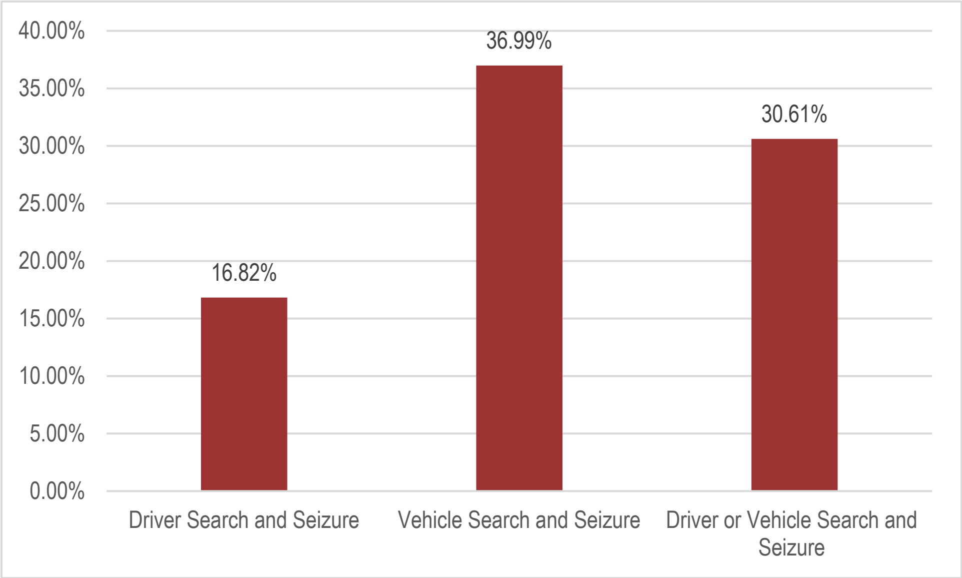
Figure 9 presents information about searches. MCSO policy dictates that deputies search all arrested drivers and search all towed vehicles; these searches are not discretionary on the part of the deputy. Non-incidental searches refer to searches that are not connected to arrests or tows, meaning they are discretionary searches conducted by deputies. As Figure 9 shows, the majority of searches of drivers occurred incident to arrest. For the comparative analysis, we considered non-incidental searches of drivers or vehicles as a search outcome; there were slightly more searches conducted by MCSO deputies of vehicles than searches of drivers.

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). Overall, 30.61 percent of non-incidental searches resulted in seizures (Figure 10).

Figure 10. Seizures during non-incidental searches



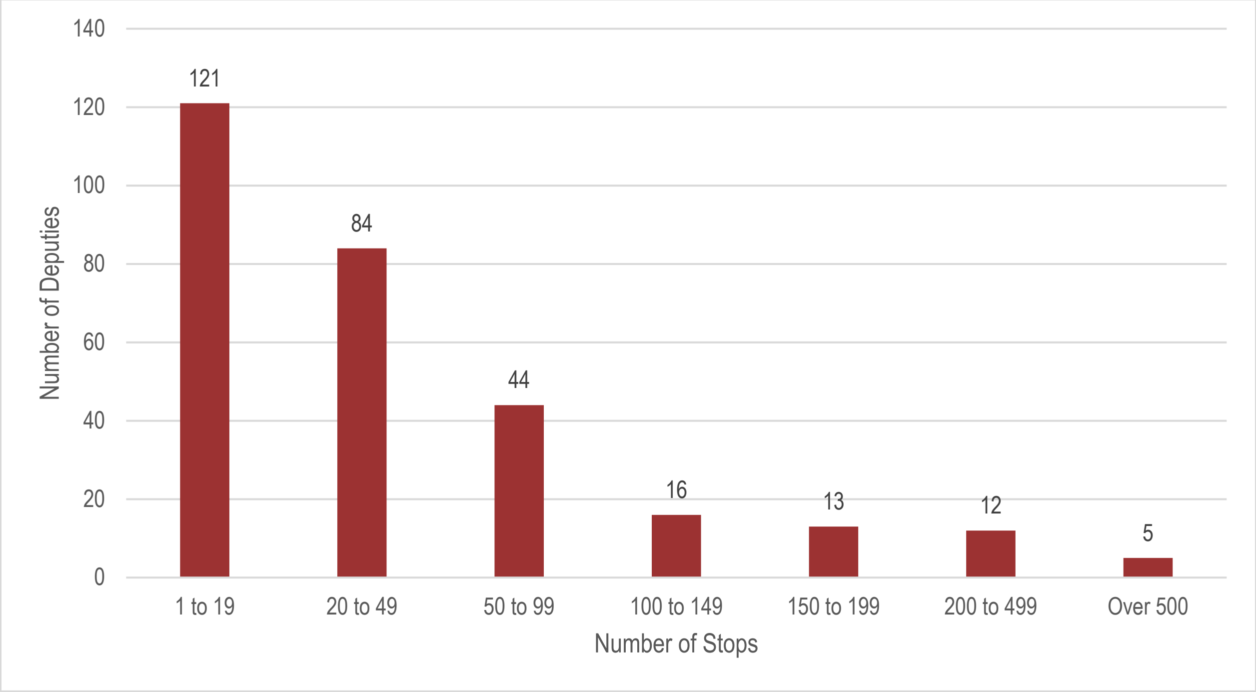
Deputies use the driver arrest variable field to document whether arrests are classified as cite and release (i.e., citation in lieu of detention) or bookings. Depending on the charges against the driver, deputies can use their discretion to choose between the two options. For example, a deputy arresting an individual for driving under the influence may use their discretion regarding whether the individual is too impaired to be released on their own recognizance or to another sober, licensed driver or should be booked for the night. Arrests of drivers are rare among traffic stops, with cite and release arrests representing 4.01 percent of total traffic stops, and less than 1 percent of stops ending in a custodial arrest (Figure 11).

Figure 11. Arrests during traffic stops

Deputy characteristics

The dataset includes stops made by 295 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 49 stops during this period, but a few deputies made over 500 stops in the same period. This finding is similar to trends in 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 supplemental appendices, which are as follows:

- 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 each specification.

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, this resulted in a critical t-statistic of 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 assumed for valid estimation, meaning that all observations contain a positive probability of being in the condition of interest or not based on the probability score (p-score) (Khandker et al., 2010). Balance evaluates the effectiveness of the matching procedure in reducing observable differences between observations within 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.

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

Stop length

The analysis team investigated differences in stop length between Hispanic and White drivers, Black and White drivers, and all racial and ethnic minority and White drivers. As noted above, deputies can indicate whether they experienced specific circumstances that extended the length of a stop beyond their control, which include technical issues (e.g., a printer failure), a language barrier, a DUI stop, training, calling for a tow, documentation issues, and other. These analyses removed all extended stops based on MCSO's traffic stop quarterly report, 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 2022 was approximately 14.81 minutes (or approximately 14 minutes and 49 seconds). Removing the extended stops, the average stop length for White drivers in 2022 was approximately 11.55 minutes (or approximately 11 minutes and 33 seconds). Table 4 summarizes the findings from this analysis. **Our analysis found statistically significant differences in stop lengths between Hispanic and White drivers and between Black and White drivers, with the average stop lengths for Hispanic drivers being about half a minute longer than the average stop lengths for White drivers, and the average stop for Black drivers and all racial and ethnic minority being slightly longer than that.**

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

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.48	2.36	Yes
Black vs. White drivers	0.72	2.54	Yes
Minority vs. White drivers	0.83	4.48	Yes

As noted in the methodology, we also considered the effect of including special assignments to assess whether the inclusion of this matching variable tangibly affects the results. As seen in Table 5, the results are consistent.

Table 5. Propensity score matching results for stop length, extended stops removed, including special assignments

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.71	3.70	Yes
Black vs. White drivers	1.13	3.51	Yes
Minority vs. White drivers	1.02	4.68	Yes

We also present the results from three alternate specifications: an analysis of all stops (including extended stops), an analysis using extended stop indicators as second stage control variables, and an analysis using extended stop indicators as matching variables. Table 6, Table 7, and Table 8 present those results. Based on the findings in the *Traffic Stop Quarterly Report: Extended Traffic Stop Indicator Use*, the results without the extended stops likely reflect the disparities experienced by the community the most accurately, absent unusual circumstances that extend stop length.

Table 6. Propensity score matching results for stop length, all stops (baseline model 2020 and prior)

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic vs. White drivers	1.58	3.38	Yes
Black vs. White drivers	1.49	2.06	Yes
Minority vs. White drivers	1.21	2.41	Yes

Table 7. Propensity score matching results for stop length, controlling for extended stop indicators

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.25	0.50	No
Black vs. White drivers	1.35	2.00	Yes
Minority vs. White drivers	0.54	1.44	No

Table 8. Propensity score matching results for stop length, including extended stop indicators as matching variables

Model	Difference (in minutes)	t-statistic	Statistically significant?
Hispanic vs. White drivers	1.05	1.94	Yes
Black vs. White drivers	0.53	0.98	No
Minority vs. White drivers	0.50	1.02	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 and ethnic minority and White drivers. To provide context and a comparison point, approximately 52.93 percent of stops involving White drivers ended in a citation. Table 9 summarizes the findings from this analysis. **For similarly situated stops, compared to White drivers, Hispanic drivers were 3.68 percentage**

points more likely to receive citations rather than warnings or other stop outcomes. No statistically significant difference is observed for Black or all racial and ethnic minority drivers. These findings are consistent with those from the previous annual report in terms of statistical significance for Hispanic drivers.

Table 9. Propensity score matching results for citations

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	3.68	3.21	Yes
Black vs. White drivers	0.27	0.12	No
Minority vs. White drivers	1.03	1.12	No

As noted in the methodology, we also considered the effect of including special assignments to assess whether the inclusion of this matching variable tangibly affects the results. As seen in Table 10, the results are generally consistent.

Table 10. Propensity score matching results for citations, including special assignments

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	2.57	2.19	Yes
Black vs. White drivers	-1.40	0.63	No
Minority vs. White drivers	1.54	1.62	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 and ethnic minority and White drivers. To provide context and a comparison point, approximately 0.31 percent of stops of White drivers involved a search. Table 11 summarizes the findings from this analysis. **Search rates had statistically significant differences (higher rates) for Hispanic drivers when compared to White drivers.**

Table 11. Propensity score matching results for non-incidental searches

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.65	3.36	Yes
Black vs. White drivers	0.25	0.86	No
Minority vs. White drivers	0.24	1.65	No

As noted in the methodology, we also considered the effect of including special assignments (e.g., DUI Task Force, Aggressive Driver, and Click-it-or-Ticket) to assess whether the inclusion of this matching variable tangibly affects the results. As seen in Table 12, the results are generally consistent, though the coefficient for Black drivers (as compared to White drivers) is too small to be reportable.

Table 12. Propensity score matching results for non-incidental searches, including special assignments

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.57	3.24	Yes
Black vs. White drivers	<0.0001	0.00	No
Minority vs. White drivers	0.40	2.38	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 and ethnic minority and White drivers. To provide context and a comparison point, approximately 4.28 percent of stops involving White drivers ended in an arrest. Table 13 summarizes the findings from this analysis. **We found no statistically significant differences in arrest rates.** These findings differ from previous annual reports. The findings were consistent across all alternate specifications of the main propensity score matching model with replacement.

Table 13. Propensity score matching results for arrests

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.59	1.18	No
Black vs. White drivers	0.77	0.89	No
Minority vs. White drivers	0.35	0.87	No

As noted in the methodology, we also considered the effect of including special assignments to assess whether the inclusion of this matching variable tangibly affects the results. As seen in Table 14, the results are consistent for Black and all racial and ethnic minority drivers, but the comparison between Hispanic and White drivers is statistically significant.

Table 14. Propensity score matching results for arrests, including special assignments

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	1.21	2.40	Yes
Black vs. White drivers	0.36	0.32	No
Minority vs. White drivers	0.63	1.65	No

We also compared booked arrests to all other arrests, booked arrests to cite-and-release arrests, and non-warrant arrests to all arrests (Table 15, Table 16, and Table 17). During booked arrests, the individual is processed in a jail facility, while during cite-and-release arrests, individuals receive a criminal citation and future court date but are free to leave the stop. For comparison, across all drivers, booked arrests accounted for approximately 0.63 percent of stops, compared with cite-and-release arrests (among White drivers, booked arrests account for approximately 0.43 percent of stops). **We found statistically significant differences in arrest rates for booked drivers for all racial and ethnic minority drivers, but not for Hispanic or Black drivers.**

Table 15. Propensity score matching results for booked arrests, compared to all other stops

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	0.22	1.55	No
Black vs. White drivers	0.63	1.32	No
Minority vs. White drivers	0.41	2.56	Yes

We also compared booked arrests to cite-and-release arrests. For this comparison, **we found statistically significant differences in booked arrests for Black drivers, but not for Hispanic drivers or all racial and ethnic minority drivers.**

Table 16. Propensity score matching results for booked arrests, compared to cite-and-release arrests

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	1.73	0.69	No
Black vs. White drivers	15.32	2.28	Yes
Minority vs. White drivers	2.40	0.75	No

Finally, we compared non-warrant arrests to all other arrests. For comparison, across all drivers, non-warrant arrests accounted for approximately 3.64 percent of all traffic stops. **We found statistically significant differences (higher rates than White drivers) in the rates of non-warrant arrests for Hispanic drivers.**

Table 17. Propensity score matching results for non-warrant arrests, compared to all stops

Model	Difference (percentage points)	t-statistic	Statistically significant?
Hispanic vs. White drivers	1.08	2.15	Yes
Black vs. White drivers	0.22	0.25	No
Minority vs. White drivers	0.44	1.16	No

Racial or ethnic differences in arrest activity for booked, cite-and-release, and warrant arrests were explored in the *Traffic Stop Quarterly Report* (TSQR) that was released at the end of the second quarter of 2022. That report shed additional light on the disparities evidenced here and continues to help MCSO better address these disparities.

Seizures

The analysis team investigated differences in seizure rates, predicated on non-incidental searches, by the race or ethnicity of the driver. Deputies made 116 stops involving non-incidental searches during the analysis period. Table 18 presents the breakdown of searches with and without seizures by the race or ethnicity of the driver. The chi-square test of homogeneity returned χ^2=0.218, p=0.8595, and the Fisher's exact test returned p=0.892, 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 18. Seizures during non-incidental searches by the race or ethnicity of the driver

Race or ethnicity of the driver	Number of Searches	Percentage of searches without seizures	Percentage of searches with seizures
Asian	1	100	0
Black	8	62.5	37.5
Hispanic	49	71.4	28.6
Native American	0	N/A	N/A
White	40	67.5	32.5
Overall	98	69.4	30.6

CONCLUSION

The MCSO and the CNA analysis team **conclude that there is evidence of disparate citation rates, search rates, and stop length by driver race or ethnicity in 2022 traffic stops in the baseline analyses**. This finding is consistent with past studies of traffic stop outcomes in other agencies (as noted in this report's introduction), as well as with previous traffic stop analyses within the MCSO under the Court Order. Stops involving Hispanic drivers were more likely to be longer and to result in a citation or search than stops involving White drivers. Stops involving Black drivers were also more likely to be longer but not more or less likely to end in a citation, search, or arrest than stops involving White drivers. Similar to stops involving Black drivers, stops of all racial and ethnic minorities were not more or less likely to result in a citation, arrest, or search than stops involving White drivers; however, they were more likely to be longer. Although results this year still indicate statistically significant findings in stop length, MCSO is encouraged that there is less than a minute difference in stop length for all minority groups.

We identified disparities in many, but not all, stop outcomes, but the disparities are not as large as those observed in prior years (as indicated in Table 19).[17] Notably, we did not identify statistically significant differences for any group in the baseline analysis for the most severe outcome—arrest. Note that the calculated differences for each year cannot necessarily be assumed to represent statistically significant changes over time; this information is purely descriptive. In the table below, red check marks represent statistical significance, and green null symbols represent a lack of statistically significant differences between the identified group and White drivers.

[17] 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, including differences in special assignments, 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 was found in Table 7 in previous TSAR reports.

Table 19. Comparison of statistical significance and differences across TSARs

Outcome	2017–2018 Finding		2019 Finding		2020 Finding		2021 Finding		2022 Finding	
	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
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
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
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
Seizures	ø		ø		ø		ø		ø	

Red check marks represent statistical significance, and green null symbols represent a lack of statistically significant differences between the identified group and White drivers

Figure 13 visually depicts the trend in the difference in the average length of traffic stops across the last five TSAR reports. This difference, for Hispanic and all racial and ethnic minority drivers, has consistently been under two minutes. The MCSO remains concerned about these disparate outcomes. Through *Traffic Stop Quarterly Report 3: Extended Stop Indicators* and *Traffic Stop Quarterly Report 4: Long Non-Extended Traffic Stops*, MCSO identified the need to collect additional data regarding stop length because unmeasured characteristics of stops have not been accounted for in this or previous TSARs. In May of 2022, MCSO began collecting information that identifies delays in traffic stops due to drivers not having the required "driving documentation" (license, registration, and insurance) with them at the time of the stop as well as 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 calling for a tow. The MCSO is committed to identifying the contributing factors to these differences in traffic stop length and to taking steps to combat them.

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

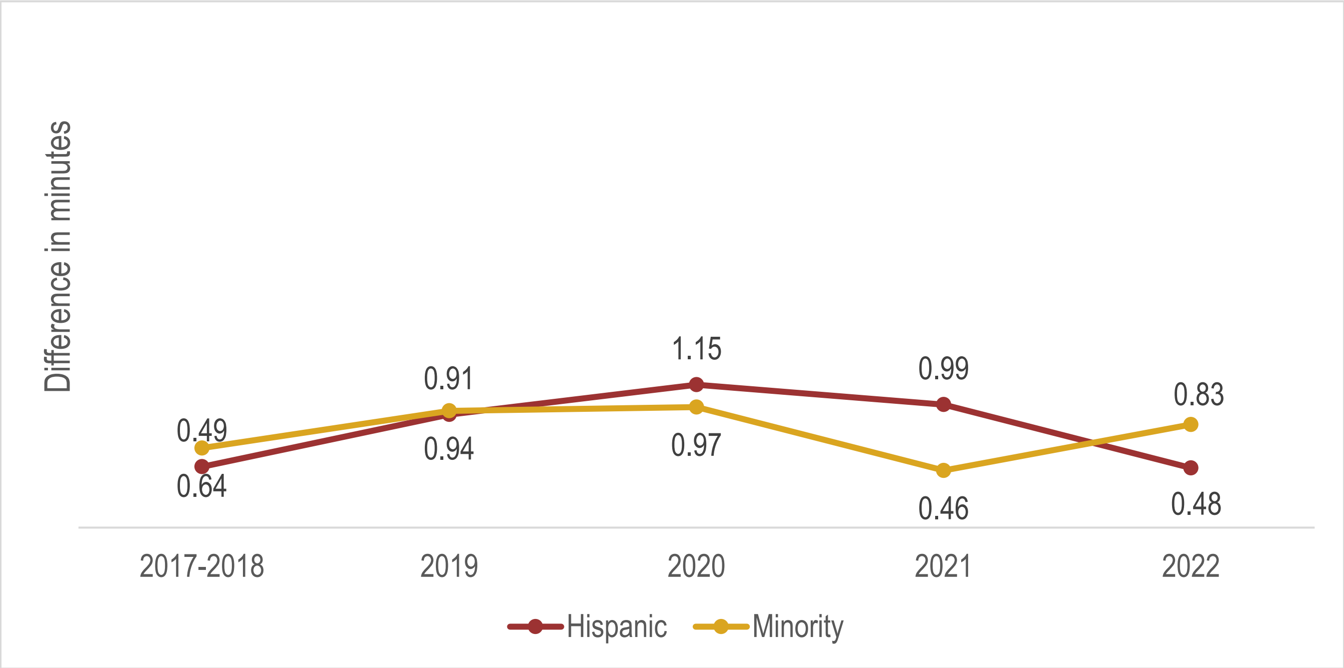


Figure 14 presents an upward trend in differences in citation rates for Hispanic and all racial and ethnic minority drivers compared with White drivers, which decreased in 2021 but increased again in 2022. The MCSO completed a quarterly report (*Traffic Stop Quarterly Report #6: Citations and Warnings*) analysis to further investigate the cause. Examining the types and number of violations that result in citations and warnings helped provide insight into the cause of these disparities, and therefore how to target efforts to combat them. This year, the offense type of violations was used 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 matched, the direction of the disparity changed, indicating that Hispanic drivers were more likely to receive a citation.

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

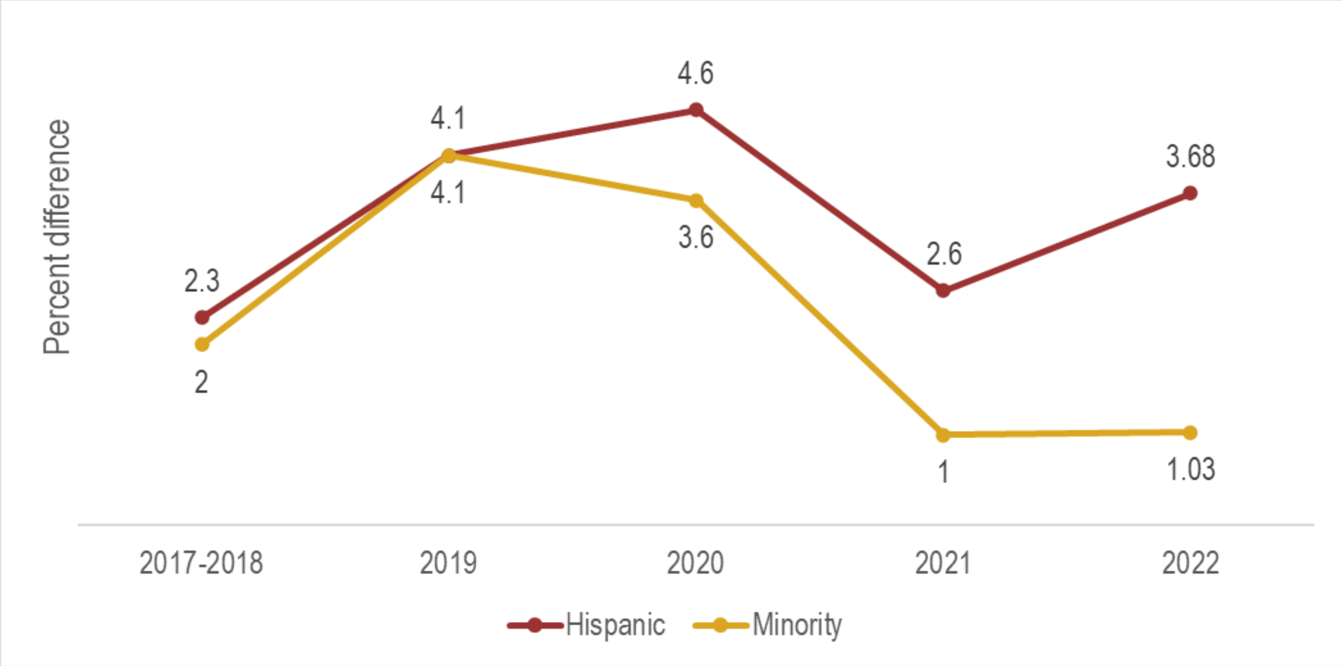


Figure 15 shows the difference between search rates for Hispanic and all racial and ethnic minority drivers compared with White drivers. Although the differences for Hispanic drivers were not statistically significant in the previous annual reports, they closely track with the observed differences for all racial and ethnic minority drivers, which decreased slightly in the 2019 analysis, increased in 2020, and decreased again in 2021 and 2022.

Figure 15. Difference in search rates by race or ethnicity (compared to White drivers)

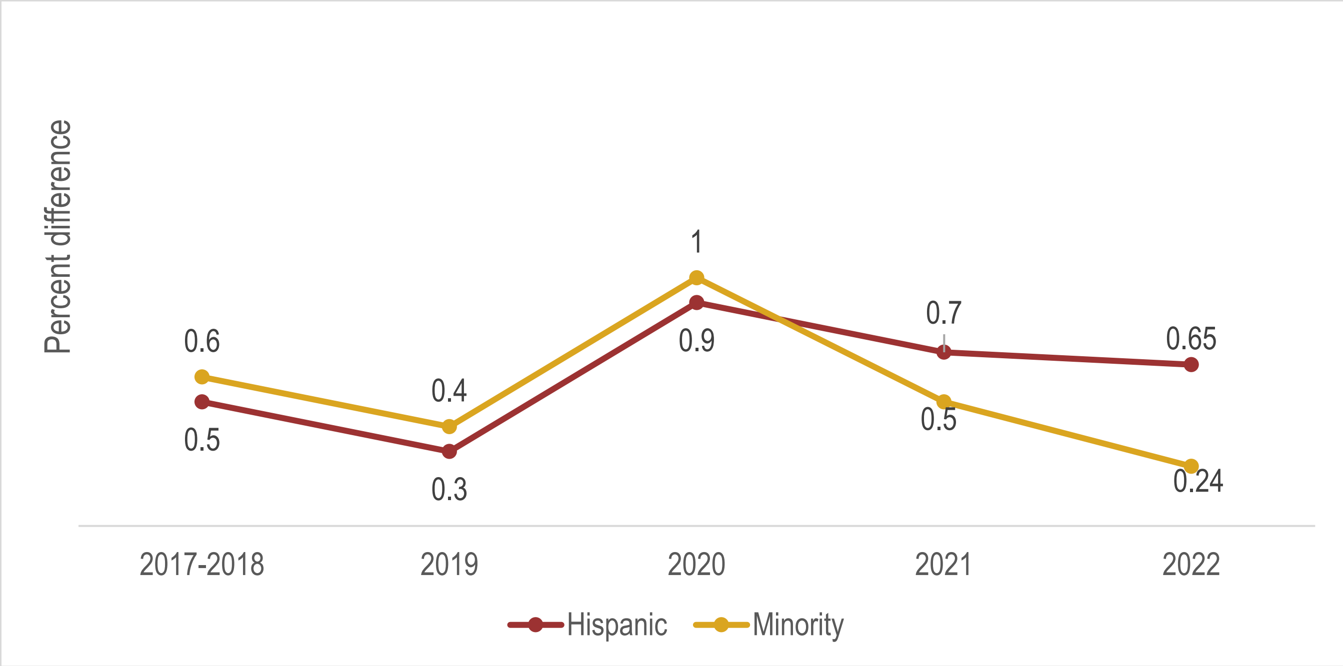
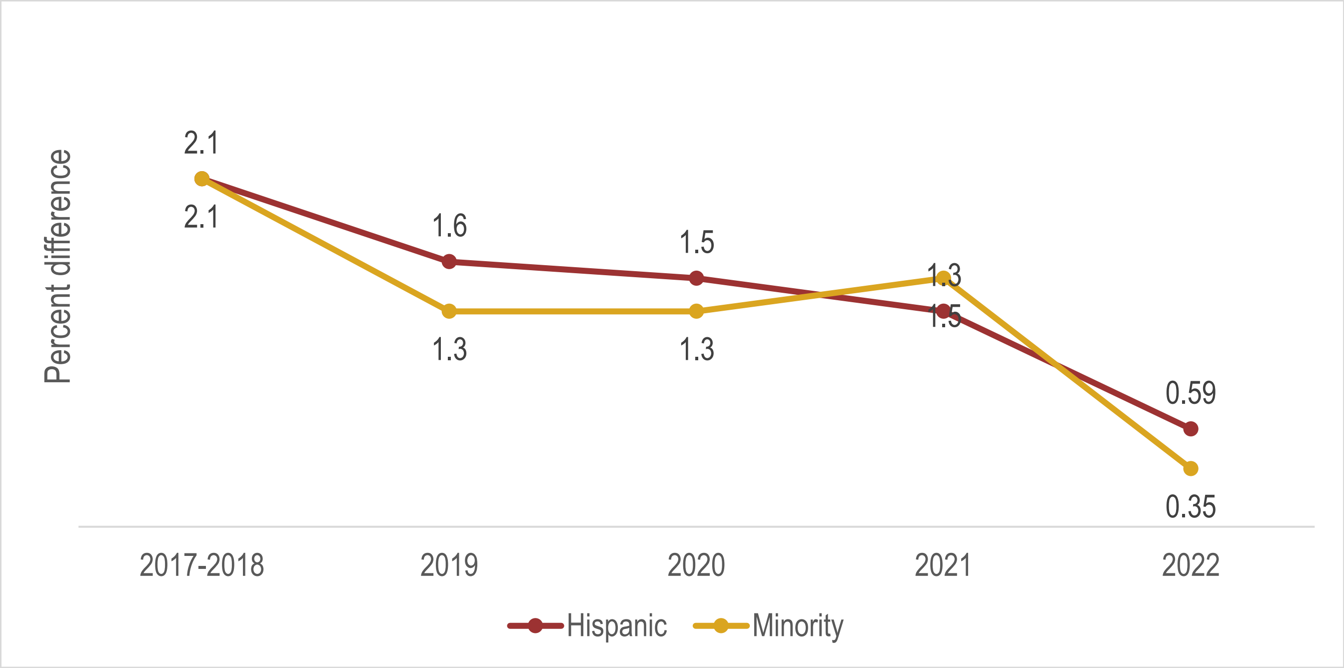


Figure 16 shows the trends in arrest rates for Hispanic and all racial and ethnic minority drivers compared with White drivers. The figure documents the overall downward trend in the difference in arrest rates across the last five TSAR reports. This difference for Hispanic drivers was highest in the 2017–2018 analysis and lower in the following years. The MCSO remains concerned about these disparate outcomes. At the same time, we are encouraged that the differences are getting smaller and no longer are statistically significant in the baseline analyses.

Figure 16. Difference in arrest rates by race or ethnicity (compared to 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. In April 2021, the MCSO began piloting 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 at the length of stop, citation rate, search rate, seizures, and arrest rate benchmarks using both comparative and descriptive analyses. The patrol activity of deputies identified by these analyses is extensively reviewed. This process allows for 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. In 2022, this pilot became a permanent part of MCSO's efforts to combat potentially biased policing.

The MCSO remains vigilant and committed to addressing disparities because they indicate possible systemic racial bias and because of the effect on the community. For context, based on propensity score matching estimates, Hispanic drivers in 2022 experienced less than a 30-second difference in stop length compared to the average stop length of 11 minutes and 33 seconds for White drivers, and Hispanic drivers had an estimated 3.68 percentage point higher citation rate compared to an average 52.93 percent citation rate for White drivers. A small number of traffic stops involved searches not incident to arrest; however, Hispanic drivers had an estimated non-incidental search rate that was 0.65 percentage points higher compared to the average of 0.31 percent for White drivers. Although these disparities indicate possible systemic bias, they may have other causes that are not accounted for in this study. 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 deployed a survey for community members who have interacted with its deputies, which is designed to capture how well deputies adhere to the principles of procedural justice. The MCSO has implemented and utilizes its early intervention system to track any disparities identified in deputies' traffic stops and conducts interventions when deemed necessary through the TSMR and review process. The MCSO produces several reports each year examining disparities in a continual effort to improve disparate outcomes for the community members it serves, including the TSAR, the TSQRs. In addition, the MCSO is continually evaluating and reviewing policies and procedures. It conducts 19 different inspections and publishes the results to ensure compliance. Additionally, the MCSO has implemented training courses over several years designed to improve cultural competencies, reduce implicit bias, and ensure bias-based policing does not occur.

The MCSO remains dedicated to its efforts to reduce bias through training, policy, and practice improvements; however, it notes that the disparities identified in this report may have causes outside of MCSO's control, and other potential causes than deputy or organizational bias must be acknowledged. The MCSO will take reasonable steps to investigate and closely monitor itself based on these results. 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 MCSO will use these analyses and other forthcoming analyses to better understand deputy behavior during traffic stops and better serve the residents of Maricopa County.

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

This page intentionally blank.

Appendix A. References

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

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

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

Baumgartner, F. R., Bell, K., Beyer, L., Boldrin, T., Doyle, L., Govan, L., ... & 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.

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

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 Council on Policing Reforms & Race: 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).

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

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

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

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

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

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

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

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

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

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*, 0011128720926122.

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

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

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

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

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

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

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*, 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! In *Final Project Report for Racial Profiling Data Collection and Analysis*.

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

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

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

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). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *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.

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

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

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

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

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

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

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*. Arlington, VA: CNA.

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

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

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

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

Simoiu, C., Corbett-Davies, S., & Goel, S. (2017). The problem of infra-marginality in outcome tests for discrimination. *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. Available at SSRN 3915038.

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
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