MICHIGAN STATE
UNIVERSITY

An Experimental Evaluation of the Michigan State Police Internal Traffic Stop Data Dashboard

September 2024

Scott Wolfe   Travis Carter

Jedidiah Knode   Grace Henry

Michigan Justice Statistics Center | School of Criminal Justice | Michigan State University

PREPARED BY THE MICHIGAN JUSTICE STATISTICS CENTER
cj.msu.edu

***Authors' Note*: This research was supported by the U.S. Department of Justice, Bureau of Justice Statistics under Grant #15-PBJS-21-GK-00021-BJSB. Points of view and opinions provided are those of the authors and do not represent the official position of the U.S. Department of Justice. We thank the Michigan State Police (MSP) for partnering on this project, especially the leadership and staff within MSP's Criminal Justice Information Center. We also thank the troopers and post commanders who agreed to be interviewed as part of this project. Their time and feedback are greatly appreciated. Additionally, we would like to thank Justin Nix and Jeff Rojek for reviewing earlier drafts of this manuscript. Please direct all correspondence to Scott Wolfe, Michigan State University, School of Criminal Justice, Baker Hall 510, East Lansing, MI 48824. Email: wolfesc1@msu.edu.

Recommended citation:

Wolfe, S. E., Carter, T., Knode, J., & Henry, G. (September 2024). An Experimental Evaluation of the Michigan State Police Internal Traffic Stop Data Dashboard. East Lansing, MI: Michigan Justice Statistics Center, School of Criminal Justice, Michigan State University.

Published September 2024

About the Authors

Scott Wolfe is a Professor, Associate Director, and Ph.D. Program Director in the School of Criminal Justice at Michigan State University. He is also Director of the Michigan Justice Statistics Center. Dr. Wolfe's research focuses on policing, and he has published on various issues including police-community relations, police training, officer safety and wellness, and police employee-supervisor relations. His recent work can be found in *Criminology*, *Justice Quarterly*, and *Journal of Research in Crime & Delinquency*. Dr. Wolfe holds a Ph.D. in criminology and criminal justice from Arizona State University.

Travis Carter is an Assistant Professor in the Department of Criminology at the University of South Florida. He received his Ph.D. from the School of Criminal Justice at Michigan State University where he was a Graduate Research Fellow for the Michigan Justice Statistics Center. His research centers on policing, with a focus on racial inequalities in police behavior and crime prevention, and developing evidence-based solutions through research partnerships with police agencies. Travis's research has appeared in *Criminology, Criminology & Public Policy, Justice Quarterly, Crime & Delinquency*, *Policing: An International Journal*, and *Journal of Criminal Justice*.

Jedidiah Knode is a doctoral student in the School of Criminal Justice at Michigan State University, and he currently serves as a Graduate Research Fellow for the Michigan Justice Statistics Center. His research focuses on policing, particularly factors of time and place as they relate to enforcement decisions, and queer perspectives in criminal justice systems. Jed has published in *Criminology, Criminology & Public Policy,* and *Justice Quarterly.*

Grace Henry received her master's degree from the School of Criminal Justice at Michigan State University. Her research focuses on policing, emphasizing psychological orientations and police behavior. As a Graduate Research Assistant, Grace has worked on the Michigan State Police Sexual Assault Kit Initiative, infographics for the Michigan Justice Statistics Center, and the Detroit and Jackson Project Safe Neighborhoods projects. She is currently a Project Coordinator at Michigan State University.

Michigan State University – School of Criminal Justice

The School of Criminal Justice at Michigan State University combines a legacy of leadership in criminal justice with cutting-edge research and education programs addressing the most pressing problems facing society. Whether you are a current or prospective student, a potential research collaborator, a proud alum, or a professional seeking to address a real-world problem, we invite you to learn more about these exciting research, education, and engagement activities.

Michigan Justice Statistics Center

The mission of the Michigan Justice Statistics Center (MJSC) is to advance knowledge about crime and justice issues in the state of Michigan while informing policy and practice. The MJSC conducts policy-relevant research and analysis to advance data-driven processes and the adoption of evidence-based strategies. Ultimately, the goal is to promote safe Michigan communities and support efforts to enhance the fairness, effectiveness, and efficiency of the criminal justice system through research, evaluation, and the dissemination of information.

Web: cj.msu.edu
Follow us on Twitter: @msu_scj

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report details an evaluation of a dashboard system rolled out by the Michigan State Police (MSP) in the Spring of 2022. MSP began exploring potential racial disparities in the agency's traffic stops in 2021. As part of that identification process, MSP developed a set of internal dashboards meant to bring awareness to MSP personnel about the racial composition of their traffic stops. After MSP piloted the dashboard with four posts, we designed an experimental evaluation where implementation of the dashboard in the remaining 26 posts was randomly assigned.

The evaluation consists of two main components: quantitative analyses of whether the dashboard reduced disparities or had unintended consequences during the study period (2019-2023), and semi-structured interviews focused on MSP employees' experiences with the dashboard. The main evaluation findings are as follows:

- The rate at which Black drivers were pulled over did not significantly change after the dashboard's rollout.
- Drawing on a host of racial disparity estimates, the evidence suggests that disparities did not significantly differ between posts that implemented the dashboard (i.e., treatment group posts) and those that continued business as usual (i.e., control group posts). Simply put, the dashboard did not influence racial disparities in traffic stop behavior.
- Rates of traffic stops, crashes, drug violations, and weapons violations did not significantly change after the dashboard was released. This suggests the dashboard did not lead to a pullback in traffic stops or negative public safety effects.
- Interview data revealed that almost all troopers reported not using the dashboard. Post commanders used the dashboard, but predominantly to compile monthly reports, and not as a mechanism to discuss the racial composition of traffic stops with their troopers. One-

one-one meetings between post commanders and troopers rarely occurred according to the interviewees.

- Many troopers and some post commanders expressed a lack of receptivity to the dashboard due to its perceived lack of utility and ambiguity about its overall purpose. This was frequently tied back to issues with the dashboard's rollout and departmental messaging.

We conclude the report with suggestions for improving future initiatives of this type and the implications for the dashboard moving forward.

INTRODUCTION

Traffic stops are widely used by police officers to enforce traffic laws and proactively police the public (Tapp & Davis, 2022), but these stops often disproportionately involve people of color (Grogger & Ridgeway, 2006; Pierson et al., 2020; Smith et al., 2021). The reasons for these disparities range widely, as do the negative consequences – from weakened trust in the criminal justice system (Kirk & Papachristos, 2011) to reluctance cooperating with police (Brunson & Wade, 2019). While there is ample research on the identification of disparities (see Neil & Winship, 2019; Ridgeway & MacDonald, 2010), the evidence base on remediation of disparities is scant (for a notable exception, see Worden et al., 2020). To minimize disparate outcomes in traffic stops, police and scholars must work together to find evidence-based solutions (Lum & Koper, 2017).

Accordingly, we used a group-randomized trial (GRT) as part of our partnership with Michigan State Police (MSP) to evaluate whether their implementation of an internal data dashboard reduced racial disparities in traffic stops across posts. MSP implemented two traffic stop data dashboards, one for post commander use and one for troopers, along with directives on how to use the data within a group of randomly assigned posts. The goal of the dashboard system was to educate troopers and their post commanders about the racial distribution of stops in their posts and increase their situational awareness about potential stop-related racial disparities. Ultimately, the agency anticipated that such efforts would reduce traffic stop disparities over time. We assessed the impact of dashboard implementation with a variety of racial disparity measures using MSP's traffic stop data while also examining whether the intervention led to any collateral consequences on troopers' behavior (e.g., pulling back in response to being monitored). To help contextualize our findings through understanding buy-in and receptivity (Wolfe et al., 2022), we individually interviewed a large group of troopers (N = 35) and post commanders (N = 9) who were part of the intervention. The information gleaned from these interviews speaks to treatment fidelity, participants' perspectives on the dashboard system, and opportunities for improvement moving forward. MSP's intervention and willingness to subject their effort to empirical scrutiny provides a unique opportunity to begin filling the gap in our understanding of what works in addressing traffic stop racial disparities.

Overview of the Michigan State Police Internal Traffic Stop Data Dashboard System

The Michigan State Police (MSP) has general law enforcement duties across the state of Michigan. The agency engages in traditional police functions in partnership with local municipalities and is responsible for highway patrol throughout the state. Accordingly, MSP conducts a large number of traffic stops, with an average of around 300,000 per year since 2020. In early 2021, MSP began to examine whether there was evidence of racial disparity in their troopers' traffic stop behavior. The agency partnered with a research team at Michigan State University to conduct independent analyses of their traffic stop data. This ongoing partnership has resulted in several reports revealing that racial disparities in traffic stops identified across the entire agency may be concentrated within a smaller set of patrol regions (i.e., posts) (Carter et al., 2022; Wolfe et al., 2021). As part of a response to the identification of these disparities, MSP developed a set of internal data dashboards to allow post commanders and their troopers to monitor the racial composition of their post's traffic stops.

The agency's traffic stop data were fed into two separate dashboards, one for troopers themselves (hereafter, the "trooper dashboard") and one for post commanders (hereafter, the "commander dashboard"). The trooper dashboard allowed troopers to see the racial composition of their individual traffic stops in comparison to their post's averages. The post commander dashboard included the same data, but MSP command staff required them to use the dashboard to complete monthly and quarterly reports specifying the average number of traffic stops involving White, Black, Hispanic, Asian, and "Other" drivers conducted by their troopers.[1] The commanders were then required to indicate the percentage of their troopers who conducted the average number of stops involving nonwhite drivers, as well as those who conducted an above and below-average number of stops of nonwhite drivers in their respective posts.

Additionally, post commanders were instructed to use troopers' stop data to guide monthly one-on-one meetings with all their troopers to discuss their statistics. To help normalize the system, these discussions were integrated into pre-existing monthly meetings that commanders routinely have with their troopers. This process was done to ensure the dashboard was not equated with targeted intervention or punishment of specific troopers. Rather, such meetings were intended to increase troopers' awareness of the racial distribution of their stops. To reach these ends, the commander-trooper meetings were designed to educate troopers on why such data tracking is important, to encourage those that showed no evidence of disparity, and to further discuss the stop behavior for each trooper. In many cases, this effort was intended to identify possible legitimate reasons for any observed disparity (e.g., patrol area, assignment time, shift). In no situation was the post commander instructed to initiate any form of reprimand. Post commanders received training on how to use the dashboard and expectations for the trooper meetings through several channels including meetings with MSP's Colonel and virtual trainings from the agency's data management unit that developed the dashboard.

An evaluation of MSP's internal data dashboard system can offer valuable insights into the existing racial disparity literature. Rather than serving as a steppingstone toward punishment, MSP's internal dashboard focused on transparency and objectivity through conversations between

[1] "Other" represents a catch-all category encompassing drivers where troopers could not identify their race/ethnicity.

troopers and their supervisors. This is important because Walker and colleagues (2007) argued that such systems work best when front-line supervisors can encourage desired behavior without relying on the threat of punishment. Such practices are the hallmark of organizationally-fair management practices that have been shown to produce a multitude of beneficial work-related behaviors among police officers, including behaving in manners that coincide with agency goals such as distributing fair outcomes to the public (Wolfe & Lawson, 2020).

MSP's internal dashboard system was designed to have post commanders and troopers self-monitor their traffic stop behavior and discuss the numbers during one-on-one meetings where troopers had a voice in the process. The potential benefits of maintaining an organizationally-fair climate through proactive monitoring and educational support, coupled with analyzing real-time data on officers' traffic stop behavior, make MSP's internal dashboard a potentially useful tool for reducing disparity in their traffic stops. Of course, whether such an intervention influences traffic stop racial disparity remains an empirical question. An evaluation of this effort will speak to whether the dashboard system is an evidence-based practice worth pursuing or how it could be changed in the future.

METHODS

Study Setting

The state of Michigan was home to 10,077,331 residents in 2020, most of whom identified as White (79%) and a smaller portion as Black or African American (14%). Michigan residents' per capita income was $34,768 in 2021, roughly 8% lower than the nationwide rate ($37,638). Operating as the state police department of Michigan, the Michigan State Police (MSP) is composed of seven operational districts and 30 patrol regions or "posts." Each district contains between two to seven posts, with an average of four posts per district. Working across these districts were 1,902 sworn personnel and 1,202 civilian staff in 2023. Of those sworn, slightly over half were "Troopers" who serve as the front-line officers for the agency. These troopers were predominantly White (90%) and male (91%).

Evaluation Design

MSP analysts developed the dashboards throughout 2021 and pilot-tested them within four posts during late 2021 and early 2022. We designed a group randomized trial (GRT) based evaluation of the dashboards where the remaining MSP posts (N = 26) were randomly assigned to treatment or control conditions. Posts assigned to the treatment condition provided all its troopers access to the trooper dashboard and required post commanders to use the commander dashboard to complete monthly and quarterly reports and conduct monthly one-on-one trooper meetings. The four pilot-test posts continued using the data dashboards throughout the intervention period but were omitted from the analysis to avoid biasing the results. Thus, there were 13 treatment posts and 13 control posts when the dashboard system went live on May 1, 2022. Control posts had no access to the dashboards and went about their assignments as they previously had done.

Treatment assignment was conducted at the post-level as opposed to the trooper level due, in part, to concerns with experimental reactivity. If some but not all troopers received the treatment in a post, there was concern that treatment troopers may feel targeted by command staff. Treatment assignment by post is not immune to this reactivity, but we reasoned that the chances were lower if every trooper in a post was assigned the treatment. Another concern stemming from treatment assignment was curbing the risk of treatment contamination. There is no way to control when troopers share dashboard information with their peers in the same post, which would diminish the validity of the randomized design if treatment assignment was at a trooper level. Assigning dashboard usage by post thus reduces this concern, assuming that communication among officers between posts is more limited.

Evaluation Outcomes

We assessed multiple quantitative outcomes to determine whether the program worked as intended. First, we measured monthly trends in the rate and variability of Black drivers stopped by MSP troopers in treated and control posts.[2] Changes in the racial composition of a trooper's stop behavior is a key outcome for the dashboard itself, while changes in the variability of each trooper's rate would indicate that the dashboard generates more predictable stop behavior.[3] This, of course, operates under the assumption that the dashboard works by enhancing troopers' self-awareness about their stop behavior and thus leads to more conscious stop decision making, which is the theoretical basis for the dashboard.

Next, we probed these changes in the racial composition of troopers' stop behavior further by constructing a set of measures to determine if the dashboard can mitigate racially disparate traffic stop behavior. We measured racial disparities in three ways to ensure our results were robust to measurement challenges with estimating disparities in police behavior (see Neil & Winship, 2019). First, we constructed an estimate of racial disparity by taking the difference in the proportion of Black drivers stopped by MSP troopers in each post every month to the proportion of the total residential population within each post that identifies as Black or African American according to 5-year American Community Survey Census estimates for the residential population (hereafter referred to as the "census benchmark"). We constructed another estimate of disparity by taking the difference in the proportion of Black drivers stopped by MSP troopers in each post to the percentage of Black drivers who were involved in two-vehicle crashes and designated as not-at-fault in those crashes on a monthly basis (hereafter referred to as the "crash benchmark").

---

[2] We restrict attention to Black citizens in our evaluation because they represent the largest racial minority group in Michigan.

[3] Variance is measured based on the rate of Black drivers stopped by each trooper in a given post for a given month.

Values for each of the benchmark estimates range from [-1, 1], with any value different from 0 indicating that Black drivers are under- or over-represented in traffic stops relative to their representation in each of the benchmark datasets. We simply take the absolute value of these estimates which allows us to assess whether the dashboard system implementation was associated with reducing disparities. This allows for an easy interpretation of results where any values larger than 0 indicate the magnitude of racial disparity based on each benchmark estimate (values closer to 0 indicated less disparity). The goal of the analyses is to determine if the dashboard reduces the magnitude of disparities observed through each benchmark.

The benchmark estimates work by assuming either the census population or crash population adequately represents those at risk of being stopped by MSP troopers on the road (Alpert et al., 2004; Smith et al., 2021). However, traffic commute times to work in Michigan average about 25 minutes per day, and range between 15 to 36 minutes in certain counties. This means the chances of a Michigander driving across county lines to go to work is high, which also means that the census benchmark may fail to reflect the true driving population at risk of being stopped by MSP troopers in any given county. Similarly, scholars have criticized traffic crashes as being unrepresentative of the population at risk of being stopped by the police (Ridgeway & MacDonald, 2010), while others have further shown at-fault determination in traffic crashes may be race-based (West, 2018). Even after accounting for who is at fault in a crash, the crash benchmark may thus fail to represent the true population at risk of being stopped by MSP troopers.

Given these limitations, we also measured racial disparities in traffic stop behavior using the ‘veil-of-darkness’ (VOD) method for each post in the treated and control conditions (Grogger & Ridgeway, 2006; Knode et al., 2023). However, unlike the previous two disparity measures, the VOD method requires an entire year of data to construct a valid estimate of disparity at the post level. Accordingly, we created a VOD estimate of disparity for each post annually, once per year of the study period (2019-2023, T = 4). Specifically, the disparity we track over time is measured as the odds of a Black driver being stopped on the roadway during daylight compared to darkness, holding constant important factors that may predict racial variation in drivers on the road and enforcement initiatives over time. If this odds ratio exceeds 1, the results suggest that Black drivers are disproportionately stopped by police on the basis of the visibility of their race. For our study, we converted this odds ratio into a standardized regression coefficient for each post to allow comparisons between posts and to facilitate a more intuitive discussion of the findings.

Much like the previous benchmark estimates, the standardized disparity values generated from the VOD range between [-∞, ∞], where values different from 0 indicate Black drivers have a lesser or greater chance of being pulled over in daylight than darkness. We take the absolute value of these estimates to ensure that any values larger than 0 indicate a raw magnitude of racial disparity. This allows us to determine if the dashboard reduced the magnitude of racial disparities based on our VOD-informed estimates of disparity. See Appendix A.1 for more information on how stops were selected and disparity estimates were generated at the post-level using the VOD method.

In a series of supplemental analyses, we considered whether there were any unintended consequences of implementing the dashboard. Research shows that the initial move toward greater oversight can lead officers to pullback, or de-police, from activities that are being monitored (Devi & Fryer, 2020), which may be cause for concern given that others have shown pullbacks in traffic stops can translate into increased traffic crashes and crime (DeAngelo & Hansen, 2014; but see also Cho et al., 2021; Nix et al., 2023). To explore this possibility, we measured the monthly rate of traffic stops and traffic crashes in each post between treated and control conditions. We then explored whether the dashboard had an impact on crime. Past traffic enforcement initiatives such as the Data-Driven Approaches to Crime and Traffic Safety (DDACTS) initiative have purported that traffic stops are designed to prevent traffic crashes and the transportation of firearms and illicit drugs that could be used in future criminal events (McClure et al., 2014). If the rate of traffic stops decreased due to the dashboard, this could lead to increases in weapons-related and drug-related crimes in their jurisdictions. Alternatively, if the dashboard does not yield a pullback, but does elevate trooper awareness of their stop decision making, this could make troopers more selective in their stop decisions and reap potential decreases in said crimes. Accordingly, we tracked monthly rates of weapons- and narcotics-related offenses at the post-level between treated and control posts to see if trends changed following the dashboard's implementation.

Data

Quantitative Data

Data for this study come from four sources: MSP traffic stops, Michigan traffic crashes, U.S. Census, and National Incident-Based Reporting System (NIBRS) data. The study period for data collection spans January 1st, 2019 to May 1st, 2023. Data on traffic stops come from MSP, which were tracked electronically through daily activity logs by MSP troopers. These data were measured at the incident level monthly for each post, and contain information on stop characteristics (e.g., time, reason for stop), driver demographics (e.g., race, sex), and patrol assignment information.[4] To make traffic stop counts comparable between large and small posts, stops were rated using county residential population data and aggregated to the post level using 2022 5-year U.S. Census estimates.

[4] It is important to note that Michigan drivers' licenses do not include driver race. MSP policy further prohibits troopers from asking drivers to self-report their race. As such, race information is collected during traffic stops based on troopers' visual assessment of the driver's race.

Traffic crash data come from the State of Michigan Traffic Crash Reports, known as the UD-10 by troopers and officers across the state. All Michigan police agencies are required to collect and submit traffic crash data, which are organized and housed by MSP. These data contain driver-level information (e.g., race, gender, at-fault), which was used to create the crash benchmark. More specifically, we measured the proportion of drivers who identified as Black who were designated as "not-at-fault" in a two-vehicle collision.[5] This information was collected monthly for treated and control posts to compare against the racial distribution of stops.

We also used the traffic crash data as part of our public safety assessment of the dashboard, which involved analyzing the incident-level information on crashes. More specifically, we measured the monthly count of traffic crashes at the county level, which were aggregated to each post and rated per 10,000 residents in treated and control conditions. When collecting these data, we aggregated county-level population counts to the post level because post jurisdictional boundaries followed county lines. Lastly, we measured monthly weapons and narcotics violations at the county level and aggregated up to posts using (NIBRS) data. We used this data to consider the potential collateral consequences of dashboard implementation on crime trends.

Qualitative Data

For our qualitative analyses, we created a semi-structured interview template for both troopers and post commanders. Questions were designed to understand MSP employees' experiences with the dashboard, including how they perceived its intention and utility and opinions on its rollout. We worked with MSP to generate a random sample of three troopers from each of

[5] We analyze this subset of traffic collisions because traffic engineering research suggests that a two-vehicle not-at-fault crashes is a more random occurrence than a single or multi-vehicle crash (three or more vehicles). These latter forms of crashes can vary as a function of drivers' experience, age, and other non-random factors (Chandraratna & Stamatiadis, 2009). Thus, the two-vehicle not-at-fault subset of crashes provide a more valid estimate of the driving population.

the 17 treatment posts and 10 of the 17 post commanders.[6] We emailed the 61 potential participants an invitation to be interviewed about their perspectives on the dashboard, following up our initial requests at the one-week and two-week mark. Forty-four of the potential participants completed an interview (72%) for a total of 35 troopers and nine post commanders. Nearly 89% of interviewees were male (11% female) and 94% were White (6% Black). Interviewed troopers had served an average of eight years in MSP) while post commanders had served an average of 22 years. We used constructivist grounded theory (Charmaz, 2006) to analyze the interviews with two independent coders to ensure interrater reliability. These codes were subsequently sorted into emergent themes and broader overarching categories.

Analytic Strategy

We used a difference-in-differences (DiD) framework to quantify the effect of the dashboard on traffic stop racial disparity, traffic crashes, and crime. Conceptually, the DiD estimator is derived from the formula:

$$DiD = \left(\overline{Treated_{post}} - \overline{Treated_{pre}}\right) - \left(\overline{Control_{post}} - \overline{Control_{pre}}\right) \qquad \text{Eq. (1)}$$

where, each outcome is measured at the post level, and posts are nested within *Treated* and *Control* conditions. Here, *Treated* indicates posts assigned to the dashboard system, and *Control* indicates those not assigned. Taking the difference in an outcome between the pre- and post-intervention periods and within *Treated* and *Control* conditions can account for time-varying and time-invariant confounders (Cunningham, 2021). Accordingly, the DiD estimator provides an optimal approach to estimating causal effects in this evaluation and has been used routinely in criminal justice program evaluations (see McLean et al., 2020; Smith & Petrocelli, 2019).

[6] We interviewed troopers at each post that had the dashboard in place, including those with the pilot program (n = 4). While these pilot posts were excluded from quantitative analyses due to having different pre- and post-treatment dates, we felt troopers at these posts could still offer valuable insight regarding the implementation and utility of the dashboard itself.

To identify the DiD value, a regression model was estimated based on the general form:

$$Y = \beta_0 + \beta_1 Treated + \alpha_g + \alpha_t + \varepsilon \qquad \text{Eq. (2)}$$

where, Y is the outcome of interest (e.g., disparity, traffic stops, crashes, and crime), α_g captures differences in the outcome between treated and untreated posts, α_t captures any differences in the outcome before and after the intervention, and *Treated* is a binary variable indicating whether a post was in the treatment condition after the dashboard was released in May of 2022. Here, β_1 represents the DiD value obtained through Equation (2) and thus the causal effect of the dashboard on each outcome. The regression models were estimated using ordinary least squares (OLS) for all outcomes. To ensure our standard errors were robust in the OLS regression models, we estimated them using 10,000 cluster wild bootstrap replications (Cameron, Gelbach, & Miller, 2008; Canay, Santos, & Shaikh, 2021).[7]

To further ensure our estimates were robust, we calculated all *p*-values using randomization inference based on 1,000 simulations through the user-written package, *ritest*, in Stata 18 (Heß, 2017). The intuition behind randomization inference is that we can test whether the observed treatment effect would have significantly exceeded a distribution of placebo effects based on a thousand different permutations of randomly assigning posts to treatment and control conditions (Keele, McConnaughy, & White, 2012). This operates in the same spirit as other exact tests of a null hypothesis in that we can derive a *p*-value from the distribution of placebo treatment effects using a conventional significance criterion (α = .05) to determine what proportion of that distribution the observed treatment effect lies beyond. The primary benefits of this method are that it offers more powerful tests of a treatment effect in small sample settings and ensures that the results are robust to the random assignment procedure (MacKinnon et al., 2021). One additional

[7] Results were the same using non-robust standard errors albeit these standard errors were slightly smaller.

benefit to randomized inference is that it makes no distributional assumptions, which means that its results are robust to potential non-normality in any regression model residuals. Therefore, we used randomization inference to provide further robustness to the main findings and follow the lead of a growing number of criminal justice program evaluations and randomized experiments that have done the same (e.g., Chalfin et al., 2022; Ham et al., 2022; Nix et al., 2019).

Robustness Checks

We conducted a post hoc power analysis using Stata 18 to determine the risk of low statistical power, which can inhibit our ability to identify significant causal effects if the null hypothesis should be rejected (Weisburd et al., 2022). Assuming a small effect size of .20 for our outcome analyses, and a conventional significance criterion of .05 based on a two-tailed test with multiple regression, we estimated a power level of .72 for this evaluation. Larger effect sizes above .30 yielded more commonly accepted levels of statistical power. The smaller sample size attributed to using annual data for our VOD estimates of racial disparity yielded much lower power. Without evidence to draw from when determining the expected effect size for an intervention of this nature, we reported the results according to two-tailed tests based on a .05 and .10 significance threshold to reduce concerns of Type-II error. Moreover, the use of randomized inference will help alleviate any residual concerns about statistical power.

We assessed the fidelity of the randomization procedure in two ways. First, we conducted a balance check to determine if there were any systematic differences in observed outcomes and potential confounders between treated and control posts. We did this by regressing each study outcome on a bivariate treatment indicator while sub-sampling the data to only the pre-intervention period. These models were estimated using OLS with robust standard errors clustered at the post-level to account for the repeated observations within posts. Second, we assessed whether trends in stops, crashes, and weapons and drug violations were parallel leading up to the intervention. Key to the DiD framework is that parallel trends must exist between the treated and untreated conditions in the main outcomes of interest (Cunningham, 2021). This should hold under random assignment, but we interrogated this assumption by conducting a joint test of statistical significance in the linear trends between treated and untreated conditions for the pre-intervention period using monthly data. If the trends statistically deviated from one another in the pre-intervention period, the parallel trends assumption would be violated, and the main results could be biased.

RESULTS

Descriptive Statistics

From January of 2019 to May of 2022, MSP posts conducted an average of 692 traffic stops per month, with some conducting as few as 10 and others conducting as many as 4,294 stops monthly. Meanwhile, MSP posts saw an average of 585 crashes, 46 drugs/narcotics violations, and 24 weapons violations. Black drivers made up about 9% of all traffic stops conducted by MSP posts on a monthly basis, with percentages ranging between 0% and 48% in specific posts.

Table 1 reports the means for the treated and control posts on each study outcome when measured as a rate per 10,000 residents. We report p-values from a bivariate regression model that estimated the difference in each of these rates between treatment and control posts during the pre-intervention period. In general, the results show that treatment and control posts were evenly balanced on observables with no statistically significant differences.

Also presented in Table 1 are F-statistics corresponding to joint tests of the linear trends in each outcome variable between treated and control conditions to determine whether such trends run parallel in the pre-intervention period. As noted by the non-significant F-statistics, all but one outcome trended in parallel during the pre-intervention period between treated and control conditions. Post-level traffic stop rates diverged between August 2020 and January 2022, which is represented by the significant F-statistic (F = 7.06, $p \leq 0.01$) and appears to be influenced by two posts. These posts experienced comparatively high changes in volumes of tourism through 2021 and into 2022 while having low annual residential population estimates, thereby generating large changes in their traffic stop rates compared to other posts. Indeed, when conducting a joint test of the linear trends of unrated traffic stop counts, the results indicate non-significant differences in their linear trends (F = 0.06). Accordingly, in a set of sensitivity analyses, we re-estimated our main models while omitting these posts to ensure our findings are not influenced by these outliers (see Appendix A.2). Coupling this with earlier evidence that indicated balance on observables between treatment and control conditions, the randomization procedure was largely successful.

Table 1. Summary Statistics

	Control	Treatment	*p*	*F*
Traffic Stop Behavior				
% Black Traffic Stops	7.0%	10.5%	0.413	0.66
Variance in % Black Traffic Stops	0.014	0.024	0.405	3.92
Racial Disparity				
Census Benchmark	0.368	0.420	0.448	0.01
Crash Benchmark	0.368	0.413	0.544	0.00
VOD	0.182	0.187	0.917	0.15
Public Safety Outcomes				
Traffic Stop Rate	0.975	1.112	0.685	7.06**
Traffic Crash Rate	0.474	0.511	0.348	1.09
Drug Violations	0.130	0.063	0.434	1.53
Weapons Law Violations	0.076	0.034	0.579	1.37

Notes: * $p \leq .05$, ** $p \leq 0.01$. Census Benchmark refers to a racial disparity estimate based on the absolute difference between the percent of residential population that is Black in a patrol post to the percent of stops involving Black drivers. Crash Benchmark refers to a racial disparity estimate based on the absolute difference between the percent of Black drivers involved in two-vehicle not-at-fault crashes in a patrol post to the percent of stops involving Black drivers. VOD refers to a racial disparity estimate measured as an absolute standardized effect based on the log odds of a Black driver being pulled over during daylight compared to darkness within an intertwilight period. All other outcomes were rated per 10,000 residents using data from the U.S. Census 5-year average estimates from 2022. All p-values reported in the table are calculated using randomization inference and are based on 1,000 simulations from the "ritest" package in Stata (Heβ, 2017). All F-statistics correspond to joint tests of the linear trends in each outcome variable between treated and control conditions with 25 degrees of freedom (except for the VOD which has only one degree of freedom).

Main Findings

In the first panel of Table 2, we regressed monthly estimates of racial disparity at the post level using the census and crash benchmarks, respectively, on our binary treatment variable. For treatment posts, racial disparities increased by 0.15% after the dashboard was implemented according to the census benchmark and by 2.94% based on the crash benchmark, though neither of these were statistically significant.[8] Meanwhile in the control posts, disparities dropped by 6.38% based on the census benchmark and by 6.05% based on the crash benchmark (neither of which were statistically significant). Taken together, change in these disparities between treatment and control groups were found to be not statistically significant. We also regressed annual estimates of racial disparity derived from the VOD method on the treatment variable to further examine whether there was an intervention effect. Across the treatment posts, we saw an increase in the magnitude of racial disparity by 14.97% while we saw a 15.97% drop in the magnitude of disparity across the control posts. As such, taking the difference in these differences yielded a positive net effect on traffic stop disparities according to our VOD estimates. This is not a statistically significant effect due, in part, to the large variability in these estimates across treatment and control posts.

In the second panel of Table 2, we report the results of our difference-in-differences estimates when regressing the monthly proportion of traffic stops involving Black drivers and the variability of these rates on a binary treatment variable indicating the MSP posts that were randomly assigned to the treatment condition. Across treatment posts, we see the monthly rate at which Black drivers were stopped by troopers increased slightly by 1.3%. In contrast, we see a less than 1% drop in the monthly rate at which Black drivers were stopped by troopers across control posts. Overall then, the results show the monthly rate at which Black drivers were pulled over in posts with the dashboard increased compared to those without the dashboard after its implementation, but this estimate was not statistically significant. The results indicated a marginally significant increase in the variability of the rate at which Black drivers were stopped in

[8] All percent changes referenced here are derived from estimating the percentage change for each outcome between pre- and post-intervention periods for treated and control groups.

Table 2. Impact of Dashboard on Traffic Stop Behavior, Disparities, and Public Safety Outcomes (N = 1,352)

	Treatment	Control	
	Pre-Post Difference	Pre-Post Difference	Difference-in-Differences
Racial Disparity			
Census Benchmark			
Coef.	0.001	-0.024	0.025
95% CI	[-0.064,0.065]	[-0.085,0.037]	[-0.080,0.128]
p	0.985	0.438	0.617
Crash Benchmark			
Coef.	0.012	-0.022	0.034
95% CI	[-0.054,0.078]	[-0.086,0.041]	[-0.072,0.140]
p	0.719	0.490	0.517
Veil of Darkness			
Coef.	0.027	-0.030	0.058
95% CI	[-0.089,0.144]	[-0.141,0.081]	[-0.117,0.234]
p	0.633	0.586	0.498
Traffic Stop Behavior			
% Black Traffic Stops			
Coef.	0.002	0.000	0.002
95% CI	[-0.023,0.026]	[-0.014,0.014]	[-0.009,0.012]
p	0.899	0.972	0.717
Variance in % Black Traffic Stops			
Coef.	0.003	-0.001	0.004
95% CI	[-0.004,0.009]	[-0.005,0.002]	[-0.001,0.010]
p	0.400	0.425	0.092
Public Safety Outcomes			
All Traffic Stops			
Coef.	0.030	-0.030	0.060
95% CI	[-0.156,0.216]	[-0.141,0.082]	[-0.208,0.327]
p	0.748	0.603	0.646
Crashes			
Coef.	0.014	0.013	0.001
95% CI	[-0.019,0.047]	[-0.016,0.043]	[-0.018,0.020]
p	0.404	0.378	0.942
Drug Violations			
Coef.	-0.005	-0.022	0.016
95% CI	[-0.019,0.008]	[-0.060,0.017]	[-0.005,0.042]
p	0.446	0.268	0.202
Weapons Law Violations			
Coef.	0.002	-0.001	0.003
95% CI	[-0.009,0.012]	[-0.033,0.031]	[-0.005,0.011]
p	0.747	0.938	0.474

Note: Census Benchmark refers to a racial disparity estimate based on an absolute difference between the percent of residential population that is Black in a patrol post to the percent of stops involving Black drivers. Crash Benchmark refers to a racial disparity estimate based on absolute difference between the percent of Black drivers involved in two-vehicle not-at-fault crashes in a patrol post to the percent of stops involving Black drivers. VOD refers to a racial disparity estimate measured as an absolute standardized effect based on the log odds of a Black driver being pulled over during daylight compared to darkness within an intertwilight period, where N = 104. All other confidence intervals were estimated using 10,000 cluster wild bootstrap replications. All other outcomes were rated per 10,000 residents using data from the U.S. Census 5-year average estimates from 2022. All *p*-values reported in the table were calculated using randomization inference and were based on 1,000 simulations from the "ritest" package in Stata (Heβ, 2017).

a post each month (b = 0.004, p = 0.092). Rather than making troopers' stop behavior more predictable, the results suggest that the dashboard made them less predictable. We re-emphasize that these outcomes reflect changes in troopers' stop behavior in general—they do not indicate changes in potential racial disparities.

Finally, we present results for the public safety outcomes in the third panel, where monthly rates of traffic stops, crashes, drug violations, and weapons violations were regressed on the treatment variable, respectively. The results show that while traffic stops increased after the dashboard's implementation, traffic crashes decreased during the same period –each of which were non-significant. Meanwhile, the monthly rate of drug violations and weapons violations increased, albeit neither were significant. The lack of statistically distinguishable effects suggests that troopers did not pullback in their traffic stop behavior when the dashboard system was implemented, and there were not subsequent changes in traffic crash or crime trends from the pre- to post-intervention periods. Simply put, the dashboard system did not cause troopers to conduct fewer stops and did not lead to negative effects on public safety.

In addition to the robustness checks previously conducted, we performed additional sensitivity analyses (not presented here) to ensure the results were robust to slight non-normality in the model residuals, different data sources in which we rated the study outcomes, and potential outliers in the data.[9] For reference, Appendix A.2 presents estimates under the original model specification while omitting two potential outlier posts in the treatment and control conditions. Appendix A.3 presents model estimates when each original outcome rate was log-transformed due to their positive skews.[10] Lastly, Appendix A.4 presents model estimates when each outcome was rated per 100,000 vehicle miles traveled using annual county-level estimates aggregated to the

[9] Although bootstrapping and randomized inference is not new and offer a reliable approach to dealing with violations of distributional assumptions, we present this evidence to satisfy any readers who remain skeptical of this approach.
[10] We did not log-transform the census and crash benchmark estimates because almost 39% of values were negative in each benchmark. These would be undefined in the transformation because they have no logarithms.

post-level from the Michigan Department of Transportation. In all cases, the results were substantively similar, and the conclusions remain the same.

Qualitative Results

Our interviews with MSP personnel may offer insight into why the dashboard failed to reduce racial disparities. We observed three broad categories that suggested potential barriers to using the dashboard based on the interviews: 1) opinions on racial disparities, 2) factors impacting receptivity, and 3) goal ambiguity. We conclude with a summary of the dashboard's implementation and usage by troopers and post commanders at the time the interviews were conducted. All percentages refer to the full set of respective interviewee groups (35 troopers [TRs], 9 post commanders [PCs]) rounded to the nearest whole number.

Why Race "Doesn't Matter": Opinions on Racial Disparities

This category consists of those interviewed who either conflated disparity with discrimination or criticized others (the media or the department) for doing so (97% TRs, 89% PCs). A post commander pointed out that troopers feel they are "being accused of racially profiling," and yet, "from day one, that was dispelled, [they] were told that was not the case. But still people continue to think that way." Race was framed as "big" in policing, especially "recently" (54% TRs, 33% PCs), with some explicitly referencing George Floyd's murder (17% TRs, 22% PCs) or the expansion of the Black Lives Matter movement (9% TRs) in the summer of 2020. The nationwide narrative on police was often mentioned, with a trooper suggesting:

> I think a lot of people feel very attacked by some of the narrative that's going around at all geographical levels. So perhaps a natural and not fully educated reaction to something like this would be okay, well this is just a way to have an "aha" moment with somebody who works in law enforcement.

Others referenced further back near the time period surrounding Ferguson and Michael Brown, with another trooper saying:

> We're all cynical as police officers, right? Especially probably in the last eight to ten years. It's gotten worse, and then you think the media is out to get us, and nobody, you know, it's just that perception I think, it feels there's more weight to it on us than there was, say, maybe like 15 years ago.

To justify their decision-making processes and actions, interviewees emphasized that they do not stop drivers because of race (74% TRs, 89% PCs), but rather due to violations (60% TRs, 67% PCs), public safety concerns (20% TRs), the reactions of drivers (9% TRs), and to be "proactive instead of reactive" (9% TRs). Many troopers further argued that they rarely can see the race of a driver in the cars they stop (60% TRs, 22% PCs), often because they work at night (46% TRs, one PC), and that they cannot control who they pull over (20% TRs) since the law is "black and white" (26% TRs, 22% PCs).

Interviewees also asserted that the areas they worked and general deployment decisions explained the presence or absence of disparity (80% TRs, 89% PCs). A common argument was that high-crime areas tend to have a higher proportion of nonwhite residents, and as patrol deployment is often concentrated in high-crime areas, this leads to pulling over more nonwhite drivers. One trooper described her stops as taking place in "a lot of inner city, predominantly Black urban areas. So our numbers, you see probably 90% Black traffic stops, because that's kind of just how the city works."

There was also a prevailing narrative around how all professions have "good and bad guys," with the mindset that "bad apples" were driving negative perceptions surrounding policing (40% TRs, 78% PCs). Some framed this as something that policing had in common with "any other profession" (29% TRs, 22% PCs). But most interviewees were also adamant that no one at their specific post was doing anything wrong (60% TRs, 100% PCs), with many troopers specifically referencing their own behavior not being discriminatory (40%). Post commanders were particularly critical of outside entities scrutinizing trooper traffic stop behavior. One post commander explicitly said, “I just want to ensure that what they're doing and their practices are not being scrutinized. There's nothing wrong with [what they’re doing].” These attitudes help frame many of the interviewees’ lack of receptivity to the dashboard rollout. As one trooper said, “what's the point of looking to see who we're pulling over if it's not biased in any way?” If most troopers and post commanders did not believe there is a problem with racial disparities in their traffic stops, it is not surprising they would view the dashboard as unnecessary.

“They Want to Be Transparent With Everybody Else But Us”: Factors Impacting Receptivity

Given that MSP’s dashboard focused on racial disparities, the majority of interviewees associated it with the release of a recent report in 2021 identifying disparities in MSP traffic stops (91% TRs, 100% PCs). This likely had a dire impact on receptivity toward the dashboard, as interviewees described the messaging around the report as poor (60% TRs, 56% PCs). A trooper exemplified this, saying, “when it came out, there was a lot of dissent for it, but I don't necessarily think it was for the actual dashboard or the study. I think it was just the way that our department handled it, originally.” A post commander further illustrated this, saying:

> Cops aren't trusting by nature. If the people at the top would have said, "hey, we're going to take a look at our stats, we're going to ask professionals to review it to see if there's anything we can do better", guys would have gone, "yeah, hold me accountable.” But because of the way it rolled out, that's not how it hit.

Interviewees suggest that troopers felt “blindsided” by the report investigating racial disparities (26% TRs, 33% PCs), with some saying they learned of the report from media coverage (23% TRs). A trooper described this coverage as “pretty much saying how MSP was racist, and I know that kind of caught us by surprise.” Other troopers echoed this, with one saying that “a lot of guys figured out [the report] happened when they're watching the news the next morning getting ready for work.” A trooper further detailed this:

> It gets posted on Facebook, "MSP shows disparities in traffic stops". You know, total negative lighting […] I thought that was an awful headline. It could have been presented a little better that, you know, "MSP took initiative to find out if there's traffic disparities." Simple like that, and you can still say what the end results are, but just wording it a little better would have been nice. I know just that headline gave a lot of troopers' negative outlooks on it before they even were exposed to it, so they were not happy. I mean, I'm one of them as well, I thought the headline was awful.

As this suggests, interviewees described feeling upset and stressed as a result (77% TRs, 78% PCs), and many believed the dashboard data would likewise be used against MSP (43% TRs, one PC). Many believed the statistics compiled by using the dashboard were unreliable or "skewed" due to poor design or lack of context (69% TRs, 78% PCs). This was often because interviewees believed the analyses did not account for assignment type, COVID, or general data limitations (57% TRs, 78% PCs). For example, a trooper mentioned:

> We have specialty teams […] like our Hometown Security Teams. They're a drug interdiction team. I have secured city projects that are in [redacted cities], all over in the state. You know, they're, that's all they're doing is stopping cars. And I don't know if those numbers skew a lot of the other numbers or not."

Many interviewees mentioned that race was becoming an influential factor in decision making on the road (29% TRs, 67% PCs), including trooper reporting practices. Because of this, there was some concern that analyses could underestimate disparities, with one trooper saying:

> I know I've heard a lot of people say, "I'm just gonna mark everybody white." And it's like, so that makes me think okay, it's not going to be more accurate even though we have this dashboard. People are going to just put in you know, white, even if they pulled over a black guy, it's like it's not going to be 100% accurate ever.

In a similar vein, another trooper explicitly framed reporting practices as a way to avoid scrutiny:

> I can tell you that I know a lot of troops have said, if they want to go down this road […] everyone is gonna become an unknown race. Because we're not going to give them the ammunition to shoot us. If they're going to be making these claims that we are racist in our traffic enforcement, we'll just quit reporting on race. And guess what? No one will ever be racist again.

These perceptions underscore general distrust in the dashboard and the agency leadership responsible for its rollout. While we did not ask questions specific to departmental relationships, and interviewees were complimentary of post-level command staff, most expressed a general disconnect with those in headquarters (80% TRs, 78% PCs). This manifested as resistance to the dashboard, with one trooper saying:

> If somebody like you or your team [the research team] or whatever were to present this dashboard to us, as opposed to our command – and we love our command and everything like that, but it's automatic, we don't want to listen to them or whatever, because they haven't been on the road for this many years or whatever – like to see it from a civilian's perspective, and, you know, I think we would get down to the why, so I think that would be beneficial.

Many of those interviewed felt that troopers had been “thrown under the bus” by command staff (34% TRs, one PC), especially during the report release (which translated into negative views of the dashboard). Because of this disconnect, many chose not to communicate their dashboard concerns with command staff when provided with the opportunity (54% TRs, 67% PCs). Some undermined the importance of the dashboard by referencing command staff being overly concerned with optics (15% TRs, one PC), with a trooper saying the dashboard was a way to say, “look how great we are, we're going to do you know, whatever you want. We're going to try to appease you, look how hip and cool we are by doing all this.” Some post commanders were aware of this and contested that perspective, with one saying:

> It's not always about our optics. It's about the optics of the people on the receiving end. So we need to be very cognizant about what we do, what we say, and how we treat people.

Interviewees gave many reasons for not wanting to express their opinion, such as having a lower rank (23% TRs, one PC), and apathy related to the belief command staff would not listen to concerns (34% TRs, 56% PCs). A trooper exemplified this attitude, saying, “The department is going to do what the department wants to do. They're really not going to listen to the troops, so what's the point of having a say?” Some mentioned fears of internal discipline (34% TRs, 22% PCs) that may result from either criticizing the dashboard or being singled out by the dashboard. A trooper said:

> To be honest with you, this department, the way that it's going, in my opinion, you're not allowed – well, you are allowed to discuss your displeasure, but that, you know, they say they won't come down on you, but you'll find yourself not getting promoted, not getting other opportunities within the agency.

This perception extended to fears that the dashboard itself would be used as a disciplinary tool, in spite of command staff insisting it would not. This was directly tied to distrust in command staff, with a trooper saying that command staff “did say [the dashboard would not be used for discipline], but people say a lot of things. That's why putting it in a contract where you know, it's written and you could go to court if they tried to do it, as opposed to, you know…people say a lot of things and then they go back on it at the same time.”

Interviewees described the dashboard training as a single virtual meeting. This training was regarded as either ineffective or with apathy by most interviewees, particularly troopers (91% TRs, 44% PCs). While some troopers spoke in positive generalities regarding the training (17% TRs), most believed it was short, insufficient, or could not remember what it entailed (74% TRs, 33% PCs). This was related to perceptions that the dashboard’s implementation was unclear and confusing (69% TRs, 33% PCs), with many uncertain about how to access the dashboard (51% TRs) or what information was available on it (37% TRs). Many admitted that they simply did not care about the dashboard (66% TRs), but others found that the virtual training was less appropriate, desiring a more “hands-on” approach. One trooper said the training would be better if “we did it in small groups at each worksite where you could actually like, flip through it more.” Another trooper said:

> They do it on one computer, and we all watched them do it. So it's certainly not because I don't want [to], what I need is to click, I need to do it myself for me to really get a good grasp. But that's typically not how the department does it. It's typically; you watch somebody do it. We'll talk about it for a couple hours and show you where different things are. And then good luck. Good luck next month when you need to use it.

It is evident that negative predispositions toward the dashboard and those responsible for implementing it influenced interviewees' receptivity to its training and use.

"I Have No Idea What It's About": Goal Ambiguity

An important hindrance to the adoption of the dashboard was ambiguity surrounding the dashboard's purpose. For example, while most interviewees understood the dashboard tracked statistics on racial groups that troopers stopped, there was confusion surrounding who the dashboard was meant to be used by. Most viewed the dashboard as a mechanism to appease external audiences (86% TRs, 100% PCs), whereas fewer interviewees believed the dashboard was useful for administrative and departmental decision making (51% TRs, 78% PCs), and few interviewees believed it was meant to be used by troopers (26% TRs, 22% PCs). This is problematic because the dashboards were intended to reduce racial disparities by helping troopers become aware of their traffic stop behavior. Instead, many viewed the dashboard as something developed by and for researchers (43% TRs, 78% PCs), perhaps indicating that MSP was not projecting ownership of the dashboard. A post commander stated that "we were told that the entire purpose of this was, that it's for you guys [the research team], and I guess we're paying you a shit ton of money to do the dashboard study." Likewise, many believed MSP was "basically asking MSU to do a secondary study" and that the dashboard's main intention was facilitating that study (29% TRs, 44% PCs). This was often tied to skepticism surrounding the 2021 report on MSP's traffic stops, with a post commander stating:

> It's to make sure that you as a group doing research actually understand how we actually do traffic policing, and what our datasets are that you actually need that are measurable so you can actually make an analysis, because quite honestly, the first report that was given, you guys were kind of unfair. I'm a statistics geek, so they basically took a couple of years of just things we record, threw it at somebody and said, make an analysis (chuckles). Now we're at the point with the dashboard, where we're saying, what measurables do you actually need, what data set do you need for actual empirical data, so you can actually make a legitimate analysis in the format of a real research study.

The most common perspective was that the dashboard's main audience was the public (77% TRs, 78% PCs), particularly as a way of being proactive and transparent (43% TRs, 33% PCs). This was also viewed in a negative light with fears of dashboard data being weaponized against troopers (43% TRs). One trooper exemplified this, saying, "I think it just makes it easier for the public to get whatever numbers they want to maybe hold up against you." Some were disgruntled over the dashboard because they believed it was just another way to surveil and scrutinize police activity (34% TRs, 22% PCs). As one trooper said:

> So it's like, every year they're coming out with another way to monitor you, you know? And it's just, it's never a good feeling to constantly, like everything you do, like your whole workday, just video cameras, you know, my camera in here running 24/7. You know, anyone at a job wouldn't like that, just being recorded 24/7. It gets old.

Those who did believe the dashboard was meant for internal use largely believed it was "another admin thing for us that has no bearing on what we do in our profession." In this way, the dashboard was at best tangential and at worst unrelated to "real police work" (37% TRs, 33% PCs).

Many interviewees, especially troopers, admitted they were uncertain of the dashboard's ultimate purpose (63% TRs, 22% PCs). The primary belief was that the dashboard would be used to investigate disparities to show MSP troopers are not racist (77% TRs, 78% PCs). Likewise, some interviewees thought the dashboard could help combat bad publicity (54% TRs, 44% PCs). For example, a trooper said they appreciated the dashboard because:

> It helps that public perception of, “oh, they're only out there stopping these individuals.” But now we have a dashboard that shows, you know, “hey, we're actually, this post is stopping white people more often at this percentage.”

Many also thought the dashboard could operate as a stop statistics database (69% TRs, 33% PCs). While the dashboard can serve many functions, it is noteworthy that no one believed the dashboard’s purpose was to reduce disparities. Some suggested that reducing disparities didn’t even make sense with the available information, with a post commander saying:

> If you say, “there's a disparity, we're going to look into it and find out what it is,” that's great. But when you're leading an organization, and you say, “there's a disparity, we don't know what it is, but we commit to making changes and doing better.” All your people go, “we don't know what it is, what are we going to change and make better?” You're already telling us we're doing something wrong, and you just told the whole world we're doing something wrong.

Ultimately, many interviewees said the dashboard was unnecessary because it had no utility (71% TRs, 67% PCs). Those interviewed fell into two camps, with some overlap – those who actively disliked or distrusted the dashboard (57% TRs, 44% PCs) and those who did not buy in because they deemed the dashboard irrelevant (57% TRs, 56% PCs). Harkening back to interviewee perceptions of disparity, those who stated that race “doesn’t matter” in policing (49% TRs and 44% PCs in full sample) almost always perceived the dashboard as unnecessary. As one trooper said:

> As far as my area … I don't really see how you can…we have so much stuff that we already work on. You can already go in and see your statistics on like the number of stops, the number of arrests and citations issued. You can already see all that. I don't know why you, there's not a draw to me to see how many people I pulled over this ethnicity or that, and that's not a draw to me. So, I don't think there would be a reason for anyone to really want to use it for that purpose.

This was also tied into perceptions of command staff, with a trooper saying:

> Their [command staff] thing is race, race, race, race, race, I, to be honest with you, I'm tired, I'm tired of hearing about it, I'm tired. That's how I operate, I don't care what you look like or what kind of background, it doesn't matter.

In the same line of reasoning as the dashboard and race not mattering, most troopers framed the performance of their job as superseding that of any information the dashboard provided (57% TRs). A trooper stated that the dashboard is "not going to change the way I do my job. I have a job to do, regardless of statistics or anything along those lines." This directly relates to the dashboard's lack of adoption among MSP employees and other challenges related to the dashboard rollout.

"I Don't Think Anybody Would Use It": Implementation Challenges

All but one trooper said they rarely or never used the dashboard. The main reason given by both troopers and post commanders was that viewing the dashboard could alter trooper behavior (69% TRs, 100% PCs). One trooper said that "I wasn't going to look at it because I was nervous it would impact how I would do my job, and it would make me fearful of how I do my job." The dashboard's intention to help troopers become more aware of the racial demographics of their stops was recognized by many interviewees (46% TRs, 78% PCs), but it was most frequently regarded as negative. A post commander said:

> Almost every trooper's mentioned [heightened awareness of driver race]. That they do now, from their own statement, they do now take that into account. You know, when it's on the forefront, because the awareness is out there. I mean, this has been mentioned to us. I mean, you hear about it in media, not only from our department, but other sources, and they say, absolutely. When they deal with someone of a different race than themselves, they think about this, absolutely.

Likewise, some troopers explicitly referenced changing their traffic stop activity not because it may be biased, but to avoid having their dashboard numbers stand out. A trooper gave an example:

> I work in seatbelt detail. I stop twelve cars, three minorities – blacks – the rest were whites. I had to let, well, I didn't have to, but I watched five minorities drive by me, clearly with no seatbelts on, so I didn't upset the numbers.

Another reason for not using the dashboard was that interviewees believed it failed to provide information besides race (51% TRs, 67% PCs). Specifically, troopers expressed interest in knowing more detail on stop locations (11%), stop outcomes (20%), shift and time-of-day information (14%), the reason for their stops (6%), and seasonal differences in stops (14%). As one trooper said:

> When they first said that something was coming out like [the dashboard], I took it [to mean] it was gonna go more into detail, which I was excited for. Because I would like to see the different traffic stops, and possibly like, what was the outcome on the stop, and why I stopped them? I think that'd be cool, and I was excited to see that, but once we had our first training and meeting with our, I think it was our captain, and it just showed race and traffic stops, people were pretty upset, I would say.

Many troopers also didn't use the dashboard because they felt they were already familiar with the racial demographics of those they were stopping (26% TRs, 44% PCs). Others believed the dashboard was an unwarranted time commitment or that they were too busy with other police duties (29% TRs, 56% PCs).

While most troopers indicated they did not use the dashboard, all but one post commander said they were familiar with the supervisor dashboard, and 67% said they regularly used it to compile reports. However, one-on-one meetings with troopers were infrequent, and 77% of troopers were uncertain how the supervisor dashboard was to be used. Post commanders used meetings to discuss data with troopers who were over certain post metrics, but 67% of post commanders also mentioned that they immediately reassured their troopers that they were not doing anything wrong. This was largely because post commanders were resistant to the idea of any troopers changing their behavior. A post commander said:

> I don't want an enforcement member going out there and changing their enforcement activity because of the dashboard. That's exactly what I don't want to happen. Now, do I think that's happening? I can't say 100% yes. But I guarantee those wheels are turning in the back of enforcement members heads of, hey, I was called into my boss's office, this was said, they said I stopped too many people of "A" so now I'm not even going to look at "A" and I'm going to focus my efforts on "B" to maybe pull my numbers back down.

Reassuring troopers and telling them not to change their behavior makes sense from an organizational perspective to put troopers at ease, but in tandem with the other barriers to implementation, this may have reduced the dashboard's potential effect on anyone it was attempting to reach. The implications of these findings are discussed next.

DISCUSSION

Police leaders and academics have spent considerable time over the past few decades trying to better understand the extent and nature of traffic stop racial disparities. While disparities may or may not be indicative of racial bias, disparity can be undesirable and negatively impacts how Black Americans perceive and interact with the police (Gau & Brunson, 2010; Nam et al., 2023; Tyler & Wakslak, 2004). The problem is that we have little empirical evidence on "what works" in mitigating these disparities once they are identified. The Michigan State Police (MSP) implemented a data dashboard system aimed at addressing traffic stop racial disparities by increasing awareness among troopers about their traffic stop characteristics. We partnered with MSP to evaluate whether the dashboard system led to reductions in disparities, and the results offer guidance for future attempts to reduce racial disparities in traffic stops.

Our main findings are based on the veil-of-darkness (VOD) methodology. If the dashboard impacted racially-disparate stop behavior, we would expect our standardized VOD estimates to get closer to 0 in posts with the dashboard, indicating similar likelihoods between daylight and darkness of pulling over Black drivers. However, the VOD results indicated the magnitude of disparity did not significantly differ based on use of the dashboard. This lack of difference was corroborated by our other disparity estimates using the census and crash benchmarks. Altogether, our results suggest the dashboard system had no effect on existing traffic stop racial disparities during the study period. We also examined whether trooper pullback occurred, as well as whether reduced deterrent effects resulted in crime and traffic crash increases (see Nix et al., 2023), but we found the dashboard did not result in any of these undesirable outcomes.

Our qualitative findings suggest explanations for why the dashboard system did not have a significant effect. Most importantly, there was a lack of receptivity and buy-in among interviewees, so much so that the vast majority almost never engaged with the dashboard. This was largely due to the belief that race should be irrelevant in policing, with many officers taking a "colorblind" approach to their traffic stop behavior (Welsh et al., 2021). While this approach is grounded in the principle of treating all drivers equally, it carries with it the perception that racial disparities in policing are only important if they result from discrimination. Most interviewees did not see utility in the dashboard because they did not believe their colleagues engaged in discriminatory stop practices. What is more, the prevailing narrative was that the dashboard was primarily implemented to satisfy the public in light of increased nationwide police scrutiny.

This perspective on racial disparities in policing was connected to interviewees feeling they were unfairly treated by command staff during the dashboard rollout. In 2022, command staff publicly distributed the report we conducted (Wolfe et al., 2021) as part of an overall plan to address traffic stop racial disparity within the agency. While command staff told employees they did not necessarily believe racial profiling was occurring, many troopers felt "thrown under the bus" by the decision to go public with the report without first notifying employees internally. Interviewees perceived the media's interpretation of the report to be that MSP was racist, and command staff's response focusing on "change" was perceived as admitting to the existence of that problem. This contributed to distrust in command staff, apathy toward expressing concerns, and reluctance to engage with the dashboard. In short, MSP troopers and post commanders perceived a lack of organizational justice with the report and dashboard rollout. In response to this, MSP has already taken strides to prevent similar situations in the future. We conducted a follow-up racial disparity report for the agency and command staff had several meetings with post commanders, troopers, and support staff to discuss the report's findings and seek their feedback (Carter et al., 2022). We believe integrating employees into the process early will help future agencies who deploy similar dashboard systems while leading to greater treatment fidelity.

The last major barrier to implementation was that the overarching goal of the dashboard was ambiguous in the eyes of MSP personnel. Troopers and post commanders were uncertain about how they were supposed to use the dashboard system, particularly given contradictions between early messaging of the dashboard as a way to reduce racial disparities and later messaging of the dashboard as a way to explore and understand disparities. This goal ambiguity was tied to the lack of receptivity and limited use of the system across the agency, harming the dashboard system's overall treatment fidelity. As mentioned before, troopers almost never used the dashboard, and some post commanders likewise indicated inconsistent use of one-on-one meetings with troopers that were key to the intervention. Moreover, post commanders who met with troopers often immediately reassured them they were not doing anything wrong before examining their stop activity, possibly undermining treatment effects related to introspecting on stop behavior.

Lack of treatment fidelity, particularly inconsistent use of the one-on-one trooper-commander meetings, likely played a role in undermining a treatment effect. These meetings were intended to provide an opportunity for troopers and their commanders to develop a deeper understanding of the racial composition of their stops and discuss why tracking such information is valuable. Consistent with the managerial principles of organizational justice (Colquitt et al., 2000), troopers were to be given a voice and opportunity to correct misunderstandings that may come from their stop statistics. If there was greater treatment fidelity (e.g., post commanders having routine meetings with troopers) or receptivity (e.g., through understanding the dashboard's utility), there may have been a greater chance for changes in disparity estimates to be observed. Decades of organizational behavior research have shown clear communication of an intervention's purpose and subordinate participation in the process improve receptivity to organizational initiatives (Wolfe & Lawson, 2020). The dashboard meetings were intended to do this, but they do not appear to have occurred at a level necessary to impact troopers' buy-in to the system.

We hope this evaluation, the first to empirically examine a dashboard's effects on racial disparities, motivates similar research to continue building the evidence base. Moving forward, more targeted interventions may be more effective at reducing disparities, as crime-related research has found focused and problem-oriented strategies provide the most return on investment (Lum, 2020; Ratcliffe, 2018; Weisburd et al., 2020; Weisburd et al., 2010). Perhaps like policing interventions aimed at reducing crime, strategies that target racial disparities should focus on specific, micro problems. Rather than focusing on all traffic stops (the analogy being "all crime"), it may be necessary to pinpoint a specific type of traffic stop (e.g., targeted enforcement stop activities) within a specific location (e.g., patrol areas of posts identified as contributing to an overall disparity problem). Sherman's (2007) "power few" may apply to racial disparities—perhaps disparities observed across an entire agency are concentrated in specific types of behaviors, among certain officers, and in specific locations (Ridgeway & MacDonald, 2009). Acknowledging this possibility will require us to revisit how we identify racial disparities on the path toward developing strategies to reduce them. We commend MSP for leading efforts to help begin building an evidence base related to addressing traffic stop racial disparities. Their efforts allowed this study to occur, and the results provide a solid foundation for future work. Research that builds on our efforts and stems from strong practitioner-researcher partnerships (Alpert, Rojek,

& Hansen, 2013; Hansen, Alpert, & Rojek, 2014; Rojek, Martin, & Alpert, 2014; Rojek, Smith, & Alpert, 2012) will be necessary as we seek evidence-based solutions to traffic stop racial disparities.

REFERENCES

Adams, I. T., McCrain, J., Schiff, D. S., Schiff, K. J., & Mourtgos, S. M. (2022). Public Pressure or Peer Influence: What Shapes Police Executives' Views on Civilian Oversight? *Preprint*.

Alpert, G. P., Smith, M. R., & Dunham, R. G. (2004). Toward a Better Benchmark: Assessing the Utility of Not-at-Fault Traffic Crash Data in Racial Profiling Research. *Justice Research and Policy*, *6*(1), 43–69.

Ang, D. (2020). The effects of police violence on inner-city students. *The Quarterly Journal of Economics*, *136*(1), 115–168.

Bor, J., Venkataramani, A. S., Williams, D. R., & Tsai, A. C. (2018). Police killings and their spillover effects on the mental health of Black Americans: A population-based, quasi-experimental study. *The Lancet*, *392*(10144), 302–310.

Burch, P., & Heinrich, C. J. (2016). *Mixed methods for policy research and program evaluation*. SAGE.

Brunson, R. K., & Wade, B. A. (2019). "Oh hell no, we don't talk to police" Insights on the lack of cooperation in police investigations of urban gun violence. *Criminology & Public Policy*, 18(3), 623-648.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, *90*(3), 414–427.

Canay, I. A., Santos, A., & Shaikh, A. M. (2021). The Wild Bootstrap with a "Small" Number of "Large" Clusters. *The Review of Economics and Statistics*, *103*(2), 346–363.

Carter, T., Knode, J., & Wolfe, S. E. (2022). *Exploring Racial/Ethnic Disparities in Michigan State Police Traffic Stops Using the Veil of Darkness Methodology.* East Lansing, MI: Michigan Justice Statistics Center, School of Criminal Justice, Michigan State University.

Chalfin, A., Kaplan, J., & LaForest, M. (2022). Street Light Outages, Public Safety and Crime Attraction. *Journal of Quantitative Criminology*, *38*(4), 891–919.

Chandraratna, S., & Stamatiadis, N. (2009). Quasi-induced exposure method: Evaluation of not-at-fault assumption. *Accident Analysis & Prevention*, *41*(2), 308–313.

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis* (Repr). Sage.

Cho, S., Gonçalves, F., & Weisburst, E. (2021). Do police make too many arrests? The effect of enforcement pullbacks on crime. *Preprint*.

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

DeAngelo, G., & Hansen, B. (2014). Life and Death in the Fast Lane: Police Enforcement and Traffic Fatalities. *American Economic Journal: Economic Policy*, *6*(2), 231–257.

Devi, T., & Fryer, R. (2020). Policing the Police: The Impact of "Pattern-or-Practice" Investigations on Crime. *Preprint*.

Dunham, R. G., Alpert, G. P., & McLean, K. (2021). Early Intervention Systems. In *Critical issues in policing: Contemporary readings* (8th ed., pp. 655–667). Waveland Press.

Fliss, M. D., Baumgartner, F., Delamater, P., Marshall, S., Poole, C., & Robinson, W. (2020). Re-prioritizing traffic stops to reduce motor vehicle crash outcomes and racial disparities. *Injury Epidemiology*, *7*(1), 3.

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.

Gullion, C. L., & King, W. R. (2020). Early intervention systems for police: A state-of-the-art review. *Policing: An International Journal*, *43*(4), 643–658.

Ham, A., Maldonado, D., Weintraub, M., Camacho, A. F., & Gualtero, D. (2022). Reducing Alcohol-Related Violence with Bartenders: A Behavioral Field Experiment. *Journal of Policy Analysis and Management*, *41*(3), 731–761.

Heß, S. (2017). Randomization Inference with Stata: A Guide and Software. *The Stata Journal: Promoting Communications on Statistics and Stata*, *17*(3), 630–651.

Jones, J. M. (2021, July 14). *In U.S., Black Confidence in Police Recovers From 2020 Low*. Gallup.Com.

Keele, L., McConnaughy, C., & White, I. (2012). Strengthening the Experimenter's Toolbox: Statistical Estimation of Internal Validity: The experimenters toolbox. *American Journal of Political Science*, *56*(2), 484–499.

Knowles, J., Persico, N., & Todd, P. (2001). Racial Bias in Motor Vehicle Searches: Theory and Evidence. *Journal of Political Economy*, *109*(1), 203–229.

Legewie, J., & Fagan, J. (2019). Aggressive Policing and the Educational Performance of Minority Youth. *American Sociological Review*, *84*(2), 220–247.

Lum, C. M., & Koper, C. S. (2017). *Evidence-based policing: Translating research into practice* (First edition). Oxford University Press.

MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2021). Wild Bootstrap and Asymptotic Inference With Multiway Clustering. *Journal of Business & Economic Statistics*, *39*(2), 505–519.

McCarthy, J. (2022, May 27). *Americans Remain Steadfast on Policing Reform Needs in 2022*. Gallup.Com.

McClure, D., Levy, J., La Vigne, N., & Hayeslip, D. (2014). *DDACTS Evaluability Assessment: Final Report on Individual and Cross-Site Findings* (NCJ 247889; p. 167). National Institute of Justice.

McLean, K., Wolfe, S. E., Rojek, J., Alpert, G. P., & Smith, M. R. (2020). Randomized controlled trial of social interaction police training. *Criminology & Public Policy*, *19*(3), 805–832.

Neil, R., & Winship, C. (2019). Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing. *Annual Review of Criminology*, *2*(1), 73–98.

Nix, J., Pickett, J. T., & Mitchell, R. J. (2019). Compliance, noncompliance, and the in-between: Causal effects of civilian demeanor on police officers' cognitions and emotions. *Journal of Experimental Criminology*, *15*(4), 611–639.

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research*, *14*(1), 2.

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

Ratcliffe, J. H. (2022). *Evidence-Based Policing: The Basics* (1st ed.). Routledge.

Ridgeway, G. (2018). Policing in the Era of Big Data. *Annual Review of Criminology*, *1*(1), 401–419.

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*. New York University Press.

Smith, M. R., & Petrocelli, M. (2019). The Effect of Concealed Handgun Carry Deregulation in Arizona on Crime in Tucson. *Criminal Justice Policy Review*, *30*(8), 1186–1203.

Smith, M. R., Tillyer, R., Lloyd, C., & Petrocelli, M. (2021). Benchmarking Disparities in Police Stops: A Comparative Application of 2nd and 3rd Generation Techniques. *Justice Quarterly*, *38*(3), 513–536.

Tapp, S. N., & Davis, E. (2020). *Contacts Between Police and the Public, 2020* (Special Report NCJ 304527; Contacts between Police and the Public, p. 25). Bureau of Justice Statistics.

Tyler, T. R., & Fagan, J. (2008). Legitimacy and cooperation: Why do people help the police fight crime in their communities. *Ohio St. J. Crim. L.*, 6, 231.

Tyler, T. R., & Wakslak, C. J. (2004). Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority. *Criminology*, *42*(2), 253–282.

Walker, S., Alpert, G. P., & Kenney, D. J. (2001). *Early Warning Systems: Responding to the Problem Police Officer* (p. 8) [Research Brief]. National Institute of Justice.

Walker, S., Milligan, S. O., & Berke, A. (2007). *Supervision and Intervention Within Early Intervention Systems: A Guide for Law Enforcement Chief Executives* (NCJ 212299; p. 54). Office of Justice Programs.

Weisburd, D., Wilson, D. B., Wooditch, A., & Britt, C. (2022). *Advanced Statistics in Criminology and Criminal Justice*. Springer International Publishing.

West, J. (2018). Racial Bias in Police Investigations. *Preprint*, 37.

Wolfe, S. E., Carter, T.M., & Knode, J. (2021). *Michigan State Police Traffic Stop External Benchmarking: A Final Report on Racial and Ethnic Disparities.* East Lansing, MI: School of Criminal Justice, Michigan State University.

Wolfe, S. E., & Lawson, S. G. (2020). The organizational justice effect among criminal justice employees: A meta-analysis. *Criminology*, *58*(4), 619–644.

Wolfe, S. E., McLean, K., Rojek, J., Alpert, G. P., & Smith, M. R. (2022). Advancing a Theory of Police Officer Training Motivation and Receptivity. *Justice Quarterly*, *39*(1), 201–223.

Worden, R. E., McLean, S. J., Engel, R. S., Cochran, H., Corsaro, N., Reynolds, D., Najdowski, C. J., & Isaza, G. T. (2020). *The impacts of implicit bias awareness training in the NYPD* (p. 188).

Wu, X., Lum, C., & Koper, C. (2021). Do everyday proactive policing activities reduce vehicle crashes? Examining a commonly held law enforcement belief using a novel method. *Journal of Criminal Justice*, *76*, 101846.

APPENDICES

Appendix A.1: Veil of Darkness Measurement Procedure

The veil of darkness (VOD) is a leading methodology for evaluating racial disparities related to the visibility of driver characteristics in traffic stops (Grogger & Ridgeway, 2006). Our principal assumption in a VOD analysis is that it is easier to identify driver race or ethnicity when it is light outside (before sunset) than when it is dark (after dusk). Because daylight's presence naturally varies over the course of the year, this acts as a natural experiment. For example, in Lansing, MI, it is dark outside at 7 PM in February but light outside at 7 PM in April. Note that at certain times of day, all stops will take place during daylight (e.g., noon) or during darkness (e.g., midnight), meaning if the racial distribution of our driving population varies by time, we might generate misleading results. We account for this by restricting our analysis to stops taking place between the earliest dusk and the latest sunset over the course of the year. During this intertwilight period (ITP), all stops can theoretically take place during daylight or darkness. Additionally, we exclude stops between sunset and dusk on individual days, as during this period, it is ambiguous whether stops are occurring during daylight or darkness.

While the underlying assumptions of this method are intuitive, it is important to rule out other explanations that explain driver race that might vary with daylight (for a comprehensive explanation, see Knode, Wolfe, & Carter, 2023). First, we exclude stops with more than one vehicle and those missing driver race information. Next, we account for seasonal variation by generating weights based on the proportion of daylight and darkness for individual stops within the ITP. After this, we include theoretically relevant controls within our multivariable regression. We control for day of week via fixed effects with Sunday as our reference group, and we control for time of day with smooth natural splines (see Grogger & Ridgeway, 2006). We also controlled for the levels of officer discretion (as high, low, or uncertain) and officer assignment type, both with fixed effects. Lastly, as disparities may be concentrated within certain posts, we control for the post each traffic stop occurs within. By analyzing stops within our ITP and holding these other stop characteristics constant, we isolate the reason for disparity to whether daylight (and by proxy, visibility of driver race) is present. If stops taking place during daylight are more likely to involve Black drivers than those taking place during darkness, that may indicate the presence of racial profiling.

Appendix A.2 Impact of Dashboard on Traffic Stop Behavior, Disparities, and Public Safety Outcomes Without Outliers (N = 1,248)

	Treatment Pre-Post Difference	Control Pre-Post Difference	Difference-in-Differences
Racial Disparity			
Census Benchmark			
Coef.	0.014	-0.027	0.041
95% CI	[-0.067,0.094]	[-0.095,0.040]	[-0.064,0.148]
p	0.739	0.424	0.425
Crash Benchmark			
Coef.	-0.023	-0.049	0.026
95% CI	[-0.102,0.056]	[-0.116,0.018]	[-0.081,0.137]
p	0.572	0.151	0.614
Veil of Darkness			
Coef.	-0.017	0.022	-0.040
95% CI	[-0.167,0.133]	[-0.128,0.173]	[-0.256,0.176]
p	0.821	0.764	0.688
Traffic Stop Behavior			
% Black Traffic Stops			
Coef.	0.001	0.000	0.001
95% CI	[-0.026,0.029]	[-0.014,0.014]	[-0.010,0.012]
p	0.929	0.972	0.793
Variance in % Black Traffic Stops			
Coef.	0.001	-0.001	0.004
95% CI	[-0.026,0.029]	[-0.005,0.002]	[-0.001,0.010]
p	0.929	0.425	0.092
Public Safety Outcomes			
All Traffic Stops			
Coef.	-0.033	-0.030	-0.007
95% CI	[-0.145,0.080]	[-0.141,0.082]	[-0.258,0.245]
p	0.568	0.603	0.954
Crashes			
Coef.	0.019	0.013	0.005
95% CI	[-0.013,0.050]	[-0.016,0.043]	[-0.014,0.025]
p	0.250	0.378	0.569
Drug Violations			
Coef.	-0.007	-0.022	0.015
95% CI	[-0.023,0.009]	[-0.060,0.017]	[-0.008,0.040]
p	0.390	0.268	0.422
Weapons Law Violations			
Coef.	0.002	-0.001	0.003
95% CI	[-0.011,0.015]	[-0.033,0.031]	[-0.005,0.012]
p	0.760	0.938	0.479

Note: Census Benchmark refers to a racial disparity estimate based on an absolute difference between the percent of residential population that is Black in a patrol post to the percent of stops involving Black drivers. Crash Benchmark refers to a racial disparity estimate based on an absolute difference between the percent of Black drivers involved in two-vehicle not-at-fault crashes in a patrol post to the percent of stops involving Black drivers. VOD refers to a racial disparity estimate measured as an absolute standardized effect based on the log odds of a Black driver being pulled over during daylight compared to darkness within an intertwilight period. All other confidence intervals were estimated using 10,000 cluster wild bootstrap replications. All other outcomes were rated per 10,000 residents using data from the U.S. Census 5-year average estimates from 2022. All *p*-values reported in the table were calculated using randomization inference and were based on 1,000 simulations from the "ritest" package in Stata (Heβ, 2017).

Appendix A.3 Impact of Dashboard on Traffic Stop Behavior, Disparities, and Public Safety Outcomes with Log-Transformed Outcomes

	Treatment Pre-Post Difference	Control Pre-Post Difference	Difference-in-Differences
Traffic Stop Behavior			
% Black Traffic Stops			
Coef.	0.036	-0.040	0.072
95% CI	[-0.209,0.281]	[-0.278,0.197]	[-0.113,0.253]
p	0.773	0.739	0.414
Variance in % Black Traffic Stops			
Coef.	0.098	-0.154	0.246
95% CI	[-0.265,0.461]	[-0.520,0.211]	[-0.139,0.637]
p	0.596	0.407	0.218
Public Safety Outcomes			
All Traffic Stops			
Coef.	0.015	-0.039	0.054
95% CI	[-0.136,0.165]	[-0.161,0.083]	[-0.166,0.271]
p	0.849	0.529	0.596
Crashes			
Coef.	0.026	0.023	0.003
95% CI	[-0.036,0.087]	[-0.039,0.085]	[-0.038,0.044]
p	0.409	0.464	0.904
Drug Violations			
Coef.	-0.005	-0.015	0.011
95% CI	[-0.017,0.008]	[-0.043,0.012]	[-0.004,0.026]
p	0.463	0.274	0.191
Weapons Law Violations			
Coef.	0.002	-0.001	0.002
95% CI	[-0.008,0.011]	[-0.024,0.023]	[-0.005,0.009]
p	0.753	0.955	0.572

Note: All confidence intervals were estimated using 10,000 cluster wild bootstrap replications. All other outcomes were rated per 10,000 residents using data from the U.S. Census 5-year average estimates from 2022. All *p*-values reported in the table were calculated using randomization inference and were based on 1,000 simulations from the "ritest" package in Stata (Heβ, 2017).

Appendix A.4 Impact of Dashboard on Traffic Stop Behavior, Disparities, and Public Safety Outcomes with Outcomes Rated per 100,000 Vehicle Miles Traveled

	Treatment	Control	
	Pre-Post Difference	Pre-Post Difference	Difference-in-Differences
Public Safety Outcomes			
All Traffic Stops			
Coef.	-0.011	-0.013	0.002
95% CI	[-0.113,0.091]	[-0.117,0.090]	[-0.239,0.239]
p	0.830	0.798	0.983
Crashes			
Coef.	0.011	0.012	-0.002
95% CI	[-0.014,0.035]	[-0.017,0.042]	[-0.018,0.015]
p	0.387	0.415	0.858
Drug Violations			
Coef.	0.000	0.000	0.000
95% CI	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
p	0.355	0.479	0.543
Weapons Law Violations			
Coef.	0.000	0.000	0.000
95% CI	[0.000,0.000]	[0.000,0.000]	[0.000,0.000]
p	0.883	0.977	0.927

Note: All confidence intervals were estimated using 10,000 cluster wild bootstrap replications. All other outcomes were rated per 10,000 residents using data from the U.S. Census 5-year average estimates from 2022. All *p*-values reported in the table were calculated using randomization inference and were based on 1,000 simulations from the "ritest" package in Stata (Heβ, 2017).