EASTERN MICHIGAN UNIVERSITY
Southeast Michigan Criminal Justice Policy Research Project (SMART)

Analysis of Ann Arbor Police Department Traffic Stop Data, 2017-2019

July 2023

Report created by: Kevin Karpiak, PhD; Sara Srygley, MA; Grigoris Argeros, PhD; Mo Amer, MA

Front Matter

Acknowledgments: this report was made possible through the support of the Community Policing Innovations Initiative of the Community Foundation for Southeast Michigan and the College of Arts and Sciences at Eastern Michigan University. Jessica Drake, from CPII stands as an essential element of what is able to appear here. In addition, we would like to thank several individuals for their patience, generosity and goodwill throughout the development of this collaboration. In particular we would like to thank members of the Ann Arbor Police Department–including former AAPD Chief Michael Cox, current Interim AAPD Chief Aimee Metzer, Deputy Chief Jason Forsberg, Acting Deputy Chief Patrick Maguire, Officer Michael Dorsch, and Data Architect Jeff McDole–and members of Ann Arbor's Independent Community Police Oversight Commission (ICPOC)–especially former Chairs Lisa Jackson and Frances Todoro-Hargreaves as well as current Chair Stefani Carter and former ICPOC Commission Members Cynthia Harrison. We would also like to thank several members of City Commissions, including members of Ann Arbor's Human Rights Commission–especially Keta Cowan, Ayesha Ghazi Edwin, Aiden Sova, and Leslie Stambaugh–and Ann Arbor's Transportation Commission. Ann Arbor Transportation Manager Raymond Hess offered crucial guidance at critical junctures of this project. This collaboration also benefited from the expertise and experience of oversight practitioners and community members throughout Washtenaw County, including Allen Dobryden, William Lopez, Nevo Polonsky, Mary Ann Sarosi, Germaine Smith, Chuck Warpehoski, Glynda Wilks, Gail Wolkoff and members of the Michigan Roundtable for Diversity & Inclusion's Southeast Michigan Sankofa Alliance: Criminal Legal Accountability Project. Finally, this report benefited from the advice and wisdom of several subject matter experts, including members of the Vera Institute and the Washtenaw Equity Partnership, Dr. Avram Bornstein, Dr. Nick Camp, Dr. Robin Engle, and Dr. Scott Wolfe. Our own colleague at EMU, Dr. Paul Leighton, offered critical feedback on earlier drafts of this report. Any errors or omissions in this report are entirely the responsibility of the authors, and despite the sage advice of these individuals.

Suggested Citation:

> Karpiak, Kevin G., Sara Srygley, Grigoris Argeros, and Mo Amer. 2023. “Analysis of Ann Arbor Police Department Traffic Stop Data, 2017-2019.” Southeast Michigan Criminal Justice Policy Research Project.

List of Tables & Figures

Tables

Figures

Executive Summary

Overview

- The Southeast Michigan Criminal Justice Policy Research Project (SMART) at Eastern Michigan University conducted an analysis to identify potential evidence of racial disparities in traffic stops collected by Ann Arbor Police Department between January 1st 2017 and December 31st 2019.
- A further aim of the report was to serve as a useful resource and guide for how communities, policy and law makers, and Civilian Oversight Boards, may make use of such traffic stop data in order to guide their work and priorities.
- The resulting analysis represents the most comprehensive and nuanced analysis of traffic stop disparities in the history of AAPD and the city of Ann Arbor.

Methods

- In order to conduct this analysis, SMART examined several frequencies–the overall Frequency of Stops, the frequency of specific recorded Reasons for Contact, and the frequency of Searches.
- SMART then cross-tabulated those frequencies by Race and Gender in order to conduct a Benchmark Analysis designed to identify disparities.
- This methodological design offer an important new level of nuance to our understanding of the distribution of disparities in Ann Arbor traffic stops

Results

- Our analysis identified significant disparities across every dimension examined, with non-white motorists being Stopped and Searched more frequently and White motorists being Stopped and Searched less frequently than would be expected in every instance.
- These disparities were not uniform across racial categories nor across various Reasons for Contact.
- The largest disparities identified in this analysis involve Multi-Racial and African-American male drivers for stops initiated for Equipment Violations (which occurred 2.41x more likely than would be expected) as well as for Searches after the initial stop (which occurred between 5.4x to 3.65x more often than would be expected).

Recommendations

- SMART recommends that police administrators, elected officials and oversight practitioners use this analysis to inform their priorities, taking into account especially the Reasons for Contact and post-contact Outcomes which exhibit the largest disparities.
- SMART also offers specific recommendations for more consistent and robust data collection and publication practices, especially pertaining to post-stop outcomes, which would enable more nuance along this dimension in future analyses.

Introduction

This report consists of the findings resulting from our analysis of traffic stop data recorded by the Ann Arbor Police Department between January 1st, 2017 and December 31st, 2019, as well as several key recommendations resulting from that analysis. It is the product of a collaboration, spanning more than two years, between the City of Ann Arbor, the Ann Arbor Police Department (AAPD), Ann Arbor's Independent Community Police Oversight Commission (ICPOC) and Eastern Michigan University's Southeast Michigan Criminal Justice Policy Research Project (SMART). It was funded in part through support from EMU's College of Arts & Sciences as well as the Community Policing Innovations Initiative (CPII), a program of the Community Foundation for Southeast Michigan (CFSEM) which has as its goal "to provide the guidance and support necessary for local communities in partnership with local law enforcement, to develop community-driven, substantive, and pragmatic changes in the way that policing and public safety services are provided," which it primarily achieves through the provision of administrative support as well as Technical and Training Assistance (TTA)[1].

This work of relationship-building made the accomplishments of this report possible in material ways that should not be overlooked. It involved dozens of hours of work planning, education, and mutual orientation–between AAPD, ICPOC, SMART researchers and other community groups– that enabled shared learning and recognition of mutual goals between each partner. It is the kind of work made possible through sustained partnerships between public universities with regional commitments and those municipal and community organizations willing to trust and work with them. The result of this particularly extensive and sustained collaboration is the most robust and nuanced analysis of disparities in traffic stops in the history of the City of Ann Arbor. It is the first such analysis to incorporate several years of traffic stop data, which in turn enabled a more nuanced and specific analysis of existing disparities. Most specifically, this analysis is the first of its kind for the City of Ann Arbor that was able to cross-tabulate race and gender in order to highlight the specific dynamics of those dimensions. It is also the first such analysis that was able to offer a more nuanced lens onto disparities in particular types of stops and post-stop outcomes, additional dimensions that became essential for our overall conclusions.

Benchmarks

SMART conducted a series of what is known as "Benchmark" analyses in order to assess whether there was significant evidence of racial disparities in their frequency, type, or outcome of traffic stops. Benchmark measures are commonly used in the analysis of police traffic stop data. A benchmark test compares the rate of contact, or of a specific outcome, with what would be expected if there were no discrimination or disparities present (Neil and Winship 2019). It may also be done by comparing the outcomes for a demographic group to the outcomes for all stopped drivers. There are a variety of benchmark types that can be used; however, all rely on comparing actual rates to the expected rates based on a predetermined measuring standard. All benchmark tests share in common that they seek to determine a denominator, or an expected

[1] https://cfsem.org/initiative/innovative-policing/

rate of traffic stops, against which to measure recorded traffic stops in order to determine whether disparities exist.

While benchmark tests are among the most common methods of analysis for police data research, there are limitations to this approach that should be kept in mind when interpreting results. One such problem is the “denominator problem” where the results appear as though there are differences between racial groups in their treatment when the differences are actually in the base-group used for measurement (Ridgeway and MacDonald 2010). For evidence of disparities in traffic stops, the proper base-group requires demographic data about people who drive (and may thus be stopped), which is harder to measure than, say, the resident population. The benchmark tests described in this section all approach the denominator problem differently, and each of these tests has strengths and weaknesses. Because benchmark tests, particularly when oversimplified, have drawbacks but are also the most accessible method for conducting traffic data research, it is important to look at the data in a variety of ways and using multiple test methodologies, before reaching conclusions about disparities in policing.

Benchmarks Used in this analysis

This report will make use of two types of Benchmark analyses. The main benchmark will be Collision or “Crash” data benchmarking. This benchmark uses demographic data on drivers who were in vehicle accidents. The benefit of this type of benchmark is that it provides demographic information about the actual driving population in the research area, while offering a sample determined by a fairly random event. One potential flaw of this approach is that there is a possibility of bias in the assignment of fault, which would mean that the sample of drivers may be skewed and not represent a complete picture of the people driving in the research area (Smith et al. 2021; Tillyer, Engel, and Cherkauskas 2010). In addition, for our analysis of Searches, we include a benchmark that compares this Outcome to the population of stopped drivers (as opposed to comparing Search rates to the population of overall motorists, as in the Collision benchmark). For more explanation for these methodological choices, see the sections Benchmark Analysis for Total Stops and Benchmark Analysis for Searches below.

However, in addition to the benchmark analyses used by SMART, several other types of benchmark analyses are discussed in this report. For that reason we’ve included a general primer on some of the most common such benchmarks, including their relative strengths and drawbacks.

Other Types of Benchmarks

Census Benchmarks

A common benchmark in early traffic data research was to use census data for the neighborhood or community and to compare the rate of traffic stops against the demographics of the local population (Smith et al. 2021). This data is easy and free to access, so using this method is both time and cost efficient. However, this method assumes that all residents have a car and drive. It may be additionally problematic in that drivers passing through the area may not be represented in the community's demographic information. Ann Arbor, for example, may have a variety of drivers who would not be reflected in census data including individuals coming from neighboring cities or townships to shop, eat, or work, as well as students and tourists or visitors from further away coming into and navigating around the city. As a result, the census data for Ann Arbor, and in many other communities, does not adequately the overall population of motorists in Ann Arbor.

Veil of Darkness Benchmarks

Veil of darkness benchmarks compare traffic stop data occurring in an "inter-twilight period," the hours in between the earliest and latest official civil twilight for a given location. The underlying rationale behind this mode of analysis is that stops that occur within this window of time operate as a kind of natural experiment, in which time of day is controlled for, in order to assess the impact of available light on the racial distribution of traffic stops. Consequently, the theory holds, such analyses can provide evidence for whether officers are purposefully targeting African-Americans or other racials groups; if such groups are overrepresented during daylight hours (when the race of the driver can potentially be seen) versus when they can not (Taniguchi et al. 2017; Wolfe, Carter, and Knode 2021; Grogger and Ridgeway 2006; Smith et al. 2021; Neil and Winship 2019). However, this method is best suited to large data sets that contain enough stops to take advantage of the start and end of daylight savings time (Pierson et al. 2020). An additional limitation of this method is that it assumes uniformity in traffic patterns, behaviors, and demographic distributions across the course of the days of the week and throughout the year. This is a very questionable assumption in the case of Ann Arbor, which experiences considerable variation in traffic patterns across the year, whether it be for large sporting events, summer festivals, or the academic semester associated with the University of Michigan. This benchmark has therefore not been included in the current report.

Internal Benchmarks

Internal benchmarks operate differently from other benchmark types. Rather than comparing drivers against other drivers to determine disparities, the internal benchmark compares officer activities against other officers to determine disparities in the traffic stop behavior of individual officers. Officers with similar work shifts and assignments were compared to identify outliers, or officers whose traffic stop behavior differs considerably from other similarly assigned officers (McLean and Rojek 2016). Because this approach only compares officers to one another, it cannot identify whether disparities are happening throughout the entire department, as officers who are similar to one another would not be flagged for disparity. Furthermore, this approach cannot account for any causes of differences between officers, such as events or requests for help that may occur randomly during their assigned shifts (McLean and Rojek 2016).

Observational Benchmarks

Observational Benchmarks, a method pioneered by John Lamberth[2] through a study of the New Jersey Turnpike (1994), are another attempt at finding a useful estimate of the motorist

[2] John Lamberth is also the lead author on one of the key traffic stop analyses conducted in Ann Arbor, discussed below (see 2004 Lamberth Consulting Report)

population. This method makes use of trained observers to tally the number of motorists at a given location for a set period of time. These tallies are then used as the benchmark against which to compare traffic stops rates in an analysis for disparities. While this benchmarking method is intriguing for several reasons, like all benchmarks it relies on a set of assumptions and brings with it accompanying limitations (McLean and Rojek 2016). One such limitation, long noted by researchers (Alpert, Smith, and Dunham 2004; Engel and Calnon 2004), is that the value provided when compared to the cost and labor above other benchmarks such as collision data remains unclear (Barnum, Miller, and Miller 2015). Additionally, they introduce several additional challenges: their targeted coverage (at only a small number of pre-selected intersections) may mean the analysis misses overall trends throughout the jurisdiction, and can also mean that key locations are excluded from the analysis (McLean and Rojek 2016).

Other Methods of Analysis

Hit Rates

Hit-rate measures have been used to explore potential bias in police vehicle-search activity during traffic stops. This test compares the rate of "hits," defined as the officer finding contraband of some kind, to the rate of searches to determine whether searches are being performed more than they are called for and whether racial disparities between those searched and those who are found to have contraband of some kind (Fridell 2005; Neil and Winship 2019). A similar concept could be applied to other traffic stop outcomes in order to identify discrepancies in the rate of citations, verbal warnings, or arrests compared with the racial demographics of those stopped.

Description of Data

Data Overview

The data used in this report is administrative data, collected and entered by the Ann Arbor Police Department, and includes all vehicle traffic stops conducted by the Ann Arbor Police Department for the years 2017, 2018 and 2019. The AAPD data includes date and time information for the beginning and end of traffic stops, demographic information for the driver of the vehicle, such as race, age, date of birth, gender, reasons cited by the officer for the stop, outcome of the stop, stop location, and a de-identified officer identification number for each stop.

It is important to make several notes on how the dimension of Race is recorded in Ann Arbor Police Department traffic stop data. Ann Arbor City Council Resolution R-50-2-00 of February 7, 2000 requires that AAPD collect and regularly share data on the Race, Age, and Gender (as well as select other information) of people stopped by AAPD[3]. However, the driver's license issued by the State of Michigan does not currently include a field to indicate the holder's racial identity; nor is it current AAPD policy to directly inquire about the motorist's race during the conduction of traffic stops. For these reasons, importantly, the variable of race as collected in this data relies on the officer's visual interpretation of the drivers' identity, not the driver's own self-identification. Additionally, it is important to note that for consistency of analysis this report makes use of the specific categories recorded by the AAPD, even though it is more common to treat "Hispanic" and "Middle Eastern" identities as ethnic, not racial categories, and for other categories, such as "Asian" and "Pacific Islander" to be combined[4]. The reported racial categories are therefore as follows:

[3] "R-50-2-00 APPROVED AS AMENDED RESOLUTION SUPPORTING AND PROVIDING GUIDELINES FOR THE COLLECTION OF DATA TO INFORM COMMUNITY CONCERNS ABOUT RACIAL PROFILING" in ANN ARBOR CITY COUNCIL MINUTES REGULAR SESSION - FEBRUARY 7, 2000 https://a2gov.legistar.com/View.ashx?M=M&ID=43337&GUID=7EC85D26-8C7C-4167-98F4-634C85841979

[4] For a further discussion of the use of racial categories in AAPD reporting, see Conclusions below)

- African American
- Asian
- Hispanic
- Middle Eastern
- Multi Racial
- Native American
- Pacific Islander
- White

Data Preparation

Upon receipt of the raw data provided by AAPD, some “cleaning” of the data was necessary in order to make it useful for this analysis. All 2020 cases have been excluded from this analysis because of the likelihood of differences in policing and driving traffic and behaviors related to the COVID-19 pandemic which are not representative of a typical year. Additionally, stops where the driver's age was less than 14 years old have been removed from the data. This is because a small number of cases (n= 19) included drivers under the age of 14, with some of those stops indicating ages as young as 1 year. This points to a likelihood that unusually young ages were the result of possible data entry errors. The result is a final sample of 34,631 cases.

Reasons for Contact (Reason for Stops)

The code options provided by the AAPD for the “Reason for Stop” variable include:

- Alcohol/Drugs
- Assist
- Belt/Restraint
- Crime BOL (Be On Lookout),
- Equipment Violation
- Other,
- Speed
- Traffic Violation
- Weighmaster

Of the 34,631 stops encompassed in this analysis, Traffic Violations were the most common reason for stops, accounting for 51.4% of all stops during the 2017-2019 period. This was the most common reason for stops for male and female drivers, and for all racial/ethnic groups. Speed and equipment violations also comprised a large share of the total stops (45%), with stops for weighmaster and alcohol or drugs being least common (see Table 1: Reason for Contact).

Table 1: Reason for Contact

Reason for Contact	n	%
Alcohol/Drugs	21	0.1%
Assist	58	0.2%
Belt/Restraint	112	0.3%
Crime BOL	97	0.3%
Equipment Vio	4,869	14.1%
Other	927	2.7%
Speed	10,746	31.0%

Table 1: Reason for Contact

Reason for Contact	n	%
Traffic Vio.	17,791	51.4%
Weighmaster	10	0.0%
Grand Total	**34,631**	**100.0%**

Searches

In addition to “Reason for Contact”, the data provides several fields of information for post-stop outcomes. One such field is for the Type of Search conducted, which also includes the possibility of “No Search.” The complete code options provided by the AAPD for the “Reason for Stop” variable include:

- After Arrest
- Consent
- Impound
- K-9
- No Search
- P/C
- Plain View

Of the 34,631 traffic stops included in this analysis, the vast majority did not include a search of any kind (see Table 2: Searches). One measure analysts use to describe and analyze such outcomes is by calculating a Search rate, which describes the percent of stops which result in a search of any kind. The overall Search Rate of AAPD for the time period covered by this analysis was 2.4% (816 searches divided by 34,631 total stops). Although there is no compulsory and universal comprehensive state or national database against which to compare these statistics, a comparison against previous studies can be illustrative to put this number in context. For example, one of the most comprehensive studies found an average search rate of 3.37% in their analysis, which encompassed six states and 132 agencies, including state agencies, local agencies, sheriff departments, and specialized agencies (Baumgartner et al. 2017, 32). In a further study of agencies in North Carolina, those same researchers found a state-wide average Search rate of 3.41% (Baumgartner, Epp, and Shoub 2018, 143). This would suggest that AAPD’s search rate is relatively low. However, it is also important to note that Baumgartner et al. also emphasize that such rates can be widely variable, as they are subject to individual officer and agency discretion, with the highest North Carolina agencies exhibiting a Search Rate between 18.74% and 10.95% and the agencies with the lowest Search Rates exhibiting rates between 0.2% and 1.79% (Baumgartner, Epp, and Shoub 2018, 141–43).

Table 2: Searches

Search Type	n	%
Searches (All types)	816	2.4%
No Search	33,815	97.6%
Grand Total	**34,631**	**100.0%**

In addition to the overall Search Rate, the 816 stops involving a search during the time period of analysis can be further broken down into the type of search conducted (See Table 3: Search Type).

.

Table 3: Search Type

Search Type	n	%
After Arrest	222	27.2%
Consent	183	22.4%
Impound	184	22.5%
K-9	67	8.2%
P/C	126	15.4%
Plain View	34	4.2%
Grand Total	**816**	**100.0%**

Analysis

Beyond the above preparation of a set of descriptive statistics on AAPD Traffic Stop behavior, SMART also conducted a series of benchmark analyses that can provide indication of the presence of racial disparities within the frequency of stops, reason for stops, and frequency of searches. The first form of analysis is what is known as Collision or "Crash" Data Benchmarking, and was used to analyze potential disparities in the overall Frequency of Stops as well as the reason for those stops. This benchmark is described more fully in the next section (see Benchmark Analysis for Total Stops). In order to assess whether there were disparities in Searches, we additionally used a benchmark that compared specific rates of these phenomena to the overall population of stopped motorists. For more detail on this second Benchmark, see Benchmark Analysis for Searches below.

Benchmark Analysis for Total Stops

Collision Data Benchmarking is one method for addressing the "denominator problem," that is, of trying to develop a picture of the racial demographics of the traffic population so that it can be assessed against the demographic distribution of recorded stops. This method compares the demographic attributes of individuals involved in collisions within the same geographic area during the same period of time and takes that as a useful estimation for the distribution of the traffic population as a whole. It then compares that distribution to the recorded stops and assesses to what degree any differences in distribution are statistically significant.

The collision benchmark analysis essentially compares the rate by which a group is involved in traffic collisions to the rate in which they are involved in traffic stops. We then calculated what is known as an "odds ratio", which, in simple terms, compares the odds of an occurrence in one group (# of traffic collisions) versus the odds of an occurrence in another group (# of traffic stops) . Specifically, we report the degree to which the observed frequency of stops exceeds or falls short of the expected frequency, as predicted by that group's representation in traffic collisions. Finally, we tested the odds ratio for statistical significance, using the z-statistic, to determine the degree of confidence in which we can conclude that the ratio of observed to expected numbers is not the result of chance. The z-statistic takes into account not just the odds ratio, but also the number of cases used to calculate it.

Table 5: Stops by Race & Gender below shows the overall rate of traffic stops conducted by AAPD between 2017-2019, broken down by Race and Gender; Table 6: Collisons by Race & Gender shows the rates of collisions reported to AAPD during that same time period, also broken down by the Race and Gender of the driver. A collision benchmark analysis makes use of the rates in Table 6 as a best-guess estimate of the overall motorist population in Ann Arbor during that time period; it is the "expected rate" that we would see in other phenomena if there were no disparities present. Combining the data in Tables 5 and 6 allows us to calculate the odds ratios for Stops, broken down by Race and Gender. These results are presented in Table 7: Collision Benchmark Analysis of Stops.

Table 5: Stops by Race & Gender

Race	*Gender*	n	%
African American	Female	2,199	6.3%
	Male	3,159	9.1%
African American Total		5,358	15.5%
Asian	Female	1,291	3.7%
	Male	1,696	4.9%
Asian Total		2,987	8.6%
Hispanic	Female	367	1.1%
	Male	540	1.6%
Hispanic Total		907	2.6%
Middle Eastern	Female	619	1.8%
	Male	1,576	4.6%
Middle Eastern Total		2,195	6.3%
Multi Racial	Female	163	0.5%
	Male	242	0.7%
Multi Racial Total		405	1.2%
Native American	Female	35	0.1%
	Male	54	0.2%
Native American Total		89	0.3%
Pacific Islander	Female	17	0.0%
	Male	35	0.1%
Pacific Islander Total		52	0.2%
White	Female	10,386	30.0%
	Male	12,252	35.4%
White Total		22,638	65.4%
Grand Total		**34,631**	**100.0%**

Table 6: Collisons by Race & Gender

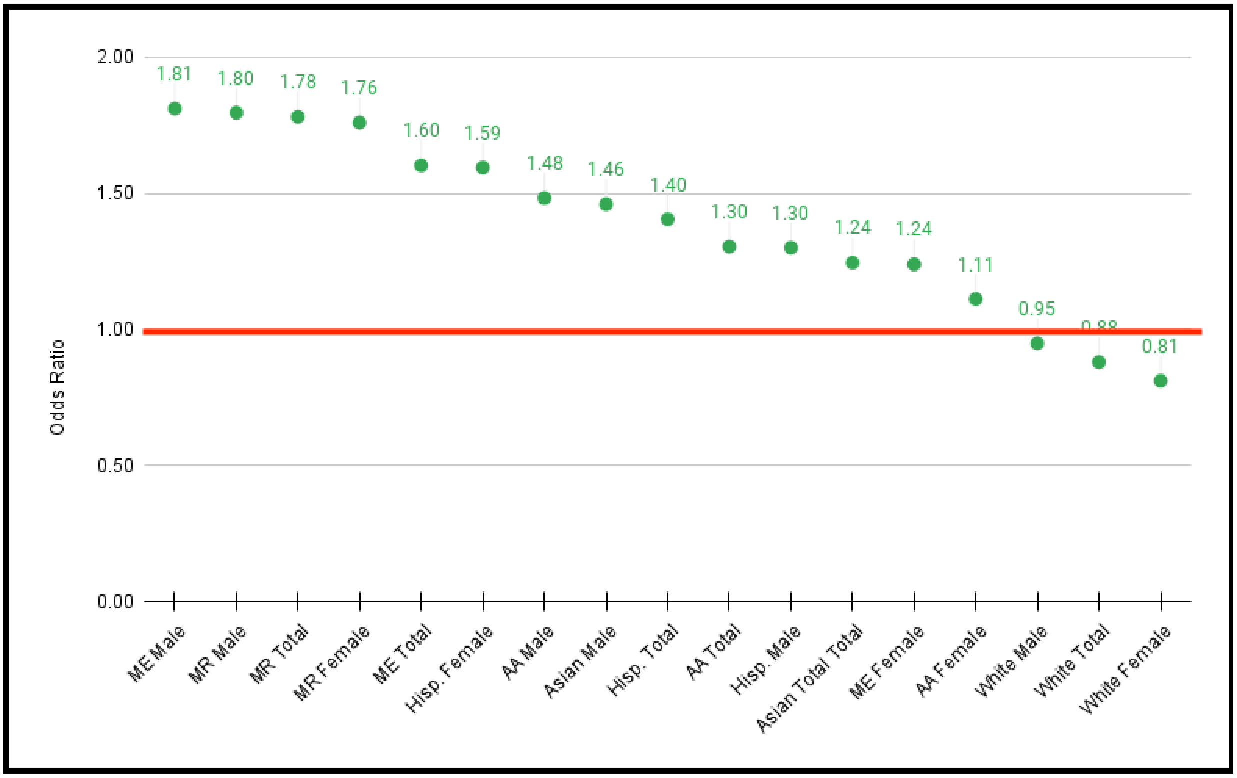
Race	*Gender*	n	%
African American	F	705	5.7%
	M	759	6.2%
African American Total		1,464	11.9%
Asian	F	441	3.6%
	M	414	3.4%
Asian Total		855	6.9%
Hispanic	F	82	0.7%
	M	148	1.2%
Hispanic Total		230	1.9%
Middle Eastern	F	178	1.4%
	M	310	2.5%
Middle Eastern Total		488	4.0%
Multi racial	F	33	0.3%
	M	48	0.4%
Multi racial Total		81	0.7%
Native American	F	19	0.2%
	M	15	0.1%
Native American Total		34	0.3%
Pacific Islander	F	5	0.0%
	M	9	0.1%
Pacific Islander Total		14	0.1%
White	F	4,565	37.0%
	M	4,604	37.3%
White Total		9,169	74.3%
Grand Total		**12,335**	**100.0%**

Table 7: Stops Odds Ratios

Race	Gender	Odds Ratio	Significance
African American	Female	1.11	**
	Male	1.48	***
AA Total		1.30	***
Asian	Female	1.04	
	Male	1.46	***
Asian Total		1.24	***
Hispanic	Female	1.59	***
	Male	1.30	**
Hispanic Total		1.40	***
Middle Eastern	Female	1.24	**
	Male	1.81	***
ME Total		1.60	***
Multi Racial	Female	1.76	**
	Male	1.80	***
MR Total		1.78	***
Native American	Female	0.66	
	Male	1.28	
NA Total		0.93	
Pacific Islander	Female	1.21	
	Male	1.39	
PI Total		1.32	
White	Female	0.81	***
	Male	0.95	***
White Total		0.88	***

Our results here are reported using progressive degrees of confidence: one asterisk (*) indicates p<0.05, or in other words, that we can conclude with a 95% confidence that the result is not the result of random chance. This is the standard bar of certainty for analyses of this type (see the section Comparisons with previous studies below). However, we've also indicated higher degrees of confidence, with two asterisks (**) indicating a p<0.01, or 99% confidence, and a three asterisks (***) indicating a 99.9% degree of confidence. Importantly, none of these results indicate a particular causal relationship, only the existence of differential frequencies–in other words, this analysis establishes the *existence* of disparities that are unlikely to be the result of chance but does not indicate the *cause* of those disparities. For a further discussion of how to interpret these results, see the section [Conclusions and recommendations]() below.

Figure 1: Significant Total Stops Odds Ratios for Collision Benchmark by Race and Gender



This analysis shows statistically significant disparities in the frequency of Traffic Stops across many racial categories. [Figure 1]() shows the statistically significant odds ratios from [Table 7: Stops Odds Ratios](), ranked from the highest odds ratio to the lowest. As a reminder, any value above 1 indicates that the frequency of stops for that group is higher than would be expected compared to the benchmark, while any number lower than zero indicates that the frequency of stops for that group is lower than would be expected. All non-white racial groups for which we were able to obtain a significant result indicate an increased frequency of being subject to traffic stops than would be expected based on their representation in the population of total motorists, as estimated by the Collision Benchmark. The highest such disparity is among Middle Eastern Male drivers, who have an odds ratio of 1.81 indicating that they are stopped 81% more often than would be expected based on their representation in the population of motorists as estimated through collision data. They are followed closely by Multi Racial Male drivers (80% more likely to be stopped), Multi racial drivers as a total group (78% more likely), and Multi Racial Female drivers (76% more likely). Conversely, White drivers as a total group are stopped 11% *less* often than would be expected based on their representation in the total traffic population estimate, with White Female drivers being stopped 19% less often than would be expected (.81 odds ratio).

Benchmark Analysis of Reason for Contact (Reason for Stops)

In addition to the overall rate of stops, benchmark analyses can be conducted for a variety of other dimensions, including the Reason for Contact (the reason for the initial stop) as well as various Outcomes of that encounter, such as whether a Search was conducted. Conducting an analysis of disparities along the dimension of Reason for Contact can be useful in that it can add nuance to the disparities identified in the frequency of Overall Stops. For example, it may indicate particular types of Stops as having larger rates of disparities or it may indicate that the disparities evident for different demographic groups in Overall Stops are related to different initial Reasons for Contact. Understanding these patterns can offer tools for addressing observed disparities (see Discussion below). For reasons of brevity, our analysis here will focus on the three most frequently indicated Reasons for Contact: Equipment Violations, Speeding Violations, and Traffic Violations.

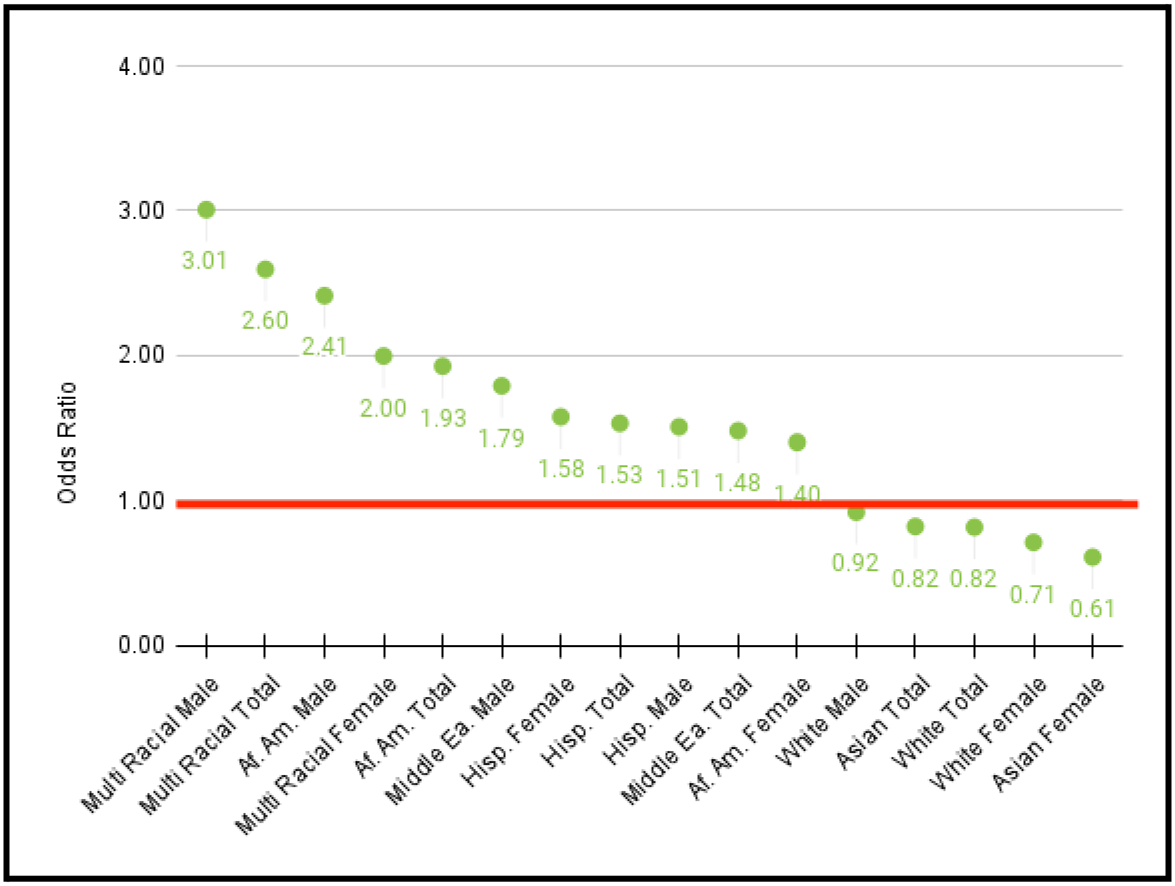
Equipment Violations

Table 8: Equipment Violations Odds Ratios demonstrates that, as with Overall Stops, there are significant disparities in stops for Reason of Equipment Violation. The largest such disparity is for Multi Racial Male drivers, who are stopped for Equipment Violations 3x more often than would be expected. They are followed by African-American Male drivers (2.41x more likely), and Multi Racial female drivers (2x more likely). Conversely, Asian female drivers are stopped for Equipment Violations 39% *less* often than would be expected based on the representation in the overall population, while White Female drivers are stopped 29% less often and White Male drivers 8% less often.

Table 8: Equipment Violations Odds Ratios			
Race	**Gender**	**Odds Ratio**	**Significance**
Af Am	Female	1.40	***
	Male	2.41	***
	Total	1.93	***
Asian	Female	0.61	***
	Male	1.05	
	Total	0.82	**
Hispanic	Female	1.58	**
	Male	1.51	**
	Total	1.53	***
Middle Ea.	Female	0.94	
	Male	1.79	***
	Total	1.48	***
Multi Racial	Female	2.00	**
	Male	3.01	***
	Total	2.60	***
Native Am.	Female	0.53	
	Male	1.01	
	Total	0.75	
Pacific Isl.	Female	1.52	
	Male	1.69	
	Total	1.63	
White	Female	0.71	***
	Male	0.92	***
	Total	0.82	***

Figure 2 includes the statistically significant odds ratios for Equipment Violations, listed in rank order. The red line highlights the Odds Ratio of 1.00, the expected rate for such stops based on the given demographic's representation in the overall motorist population.

Figure 2: Significant Equipment Violations Odds Ratios by Collision Benchmark



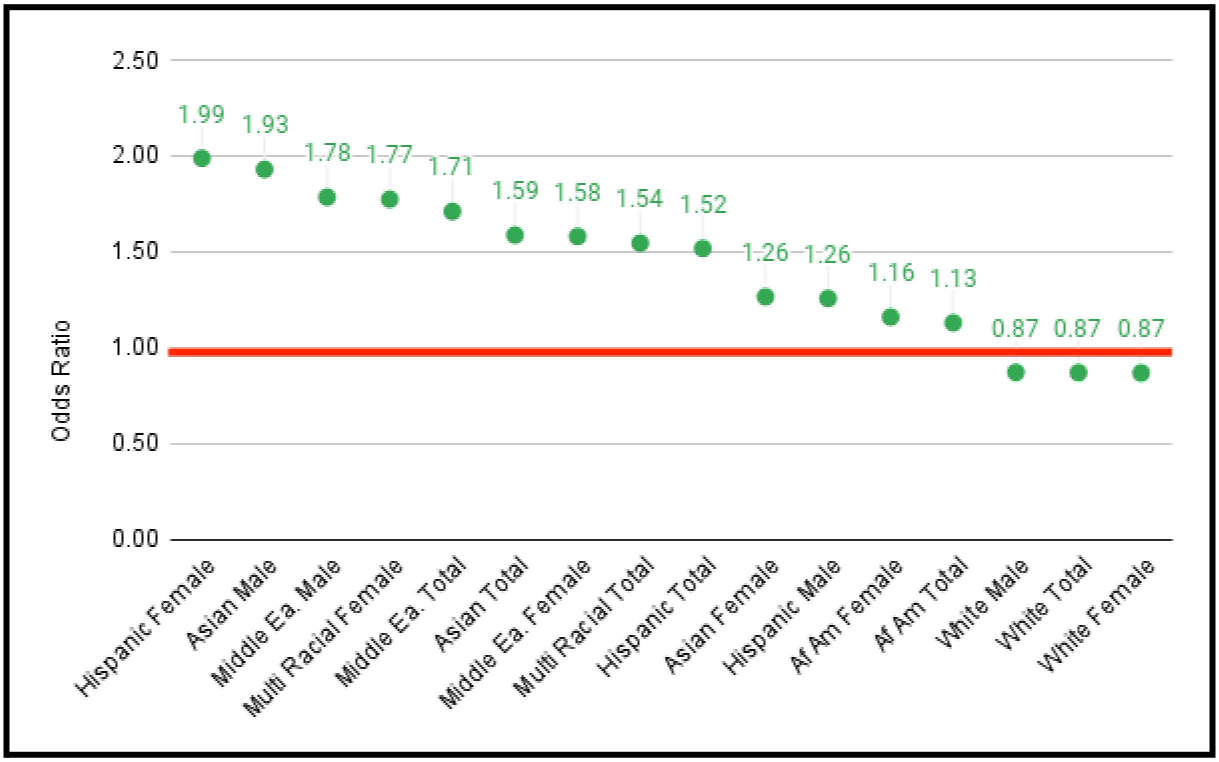
Speeding Violation

As with Equipment Violations, a benchmark analysis can be conducted for stops in which the reason for contact is listed as Speeding Violation. As with Overall Stops and Equipment Violations, White drivers are stopped significantly less often than would be expected, with an Odds ratio of .87 indicating both White Male and White Female drivers are stopped for this reason 13% less often than would be expected. Conversely, Hispanic females (99% more likely), Asian Males (93% more likely), and Middle Eastern Males (78% more likely) are the groups most likely to be stopped for this reason.

Table 9: Speeding Violations Odds Ratios			
Race	**Gender**	**Odds Ratio**	**Significance**
Af Am	Female	1.16	**
	Male	1.10	
	Total	1.13	***
Asian	Female	1.26	***
	Male	1.93	***
	Total	1.59	***
Hispanic	Female	1.99	***
	Male	1.26	**
	Total	1.52	***
Middle Ea.	Female	1.58	***
	Male	1.78	***
	Total	1.71	***
Multi Racial	Female	1.77	**
	Male	1.39	
	Total	1.54	**
Native Am.	Female	0.72	
	Male	0.92	
	Total	0.81	
Pacific Isl.	Female	2.07	
	Male	1.02	
	Total	1.39	
White	Female	0.87	***
	Male	0.87	***
	Total	0.87	***

Figure 3 includes the statistically significant odds ratios for Speeding Violations, listed in rank order. The red line highlights the Odds Ratio of 1.00, the expected rate for such stops based on the given demographic's representation in the overall motorist population.

Figure 3: Significant Speeding Violations Odds Ratios by Collision Benchmark



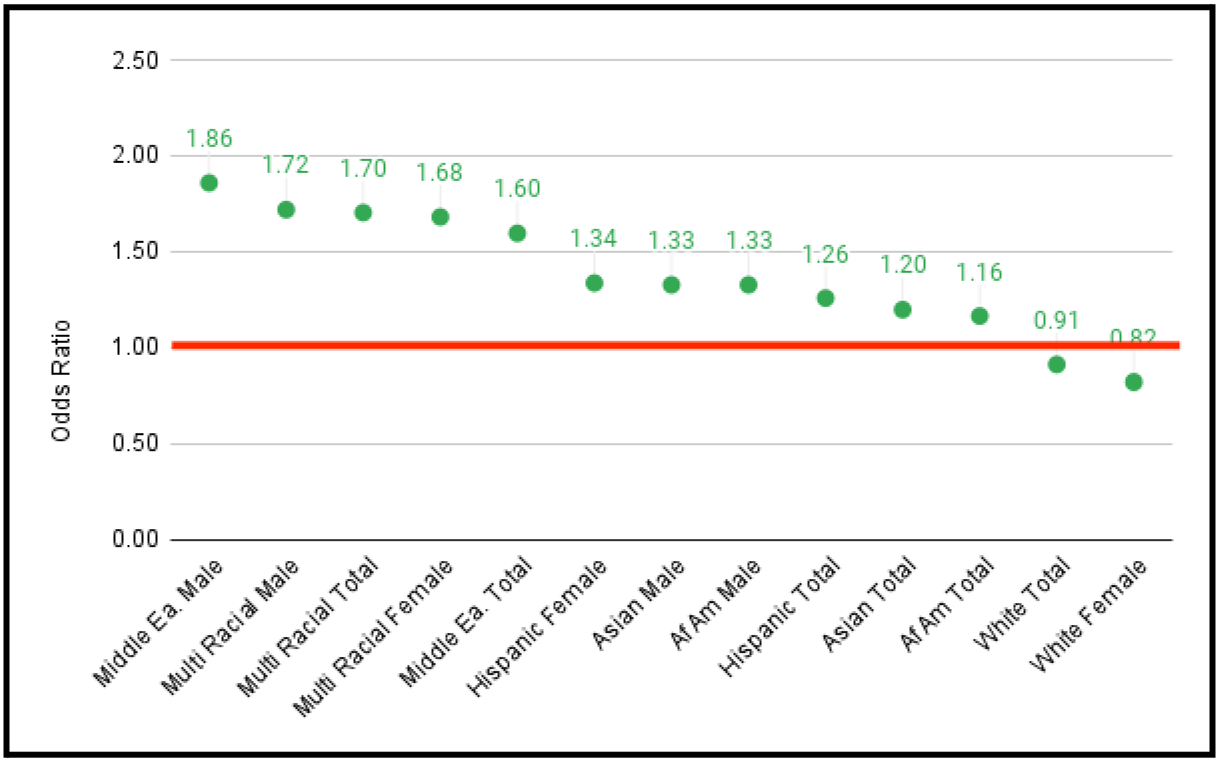
Traffic Violation

As with Equipment Violations and Speeding Violations, a benchmark analysis can be conducted to identify potential disparities in traffic stops initiated for reason of Traffic Violation. White Drivers (odds ratio 0.91, or 9% less likely), and specifically White Female drivers (18% less likely) are the only group to be stopped less often than would be expected based on their representation in the overall population of drivers. However several other groups are significantly *more* likely to be stopped for this reason than would be expected, with Middle Eastern Male drivers (86% more likely), Multi Racial Male drivers (72% more likely), and Multi Racial Female drivers (68% more likely) showing the largest significant disparities.

Table 10: Traffic Violations Odds Ratios			
Race	**Gender**	**Odds Ratio**	**Significance**
Af Am	Female	0.99	
	Male	1.33	***
	Total	1.16	***
Asian	Female	1.08	
	Male	1.33	***
	Total	1.20	***
Hispanic	Female	1.34	**
	Male	1.21	
	Total	1.26	**
Middle Ea.	Female	1.14	
	Male	1.86	***
	Total	1.60	***
Multi Racial	Female	1.68	**
	Male	1.72	**
	Total	1.70	***
Native Am.	Female	0.66	
	Male	1.53	
	Total	1.04	
Pacific Isl.	Female	0.69	
	Male	1.54	
	Total	1.24	
White	Female	0.82	***
	Male	1.00	
	Total	0.91	***

Figure 4 includes the statistically significant odds ratios for Speeding Violations, listed in rank order. The red line highlights the Odds Ratio of 1.00, the expected rate for such stops based on the given demographic's representation in the overall motorist population.

Figure 4: Significant Traffic Violations Odds Ratios by Collision Benchmark



Benchmark Analysis for Searches

Finally, we conducted a Benchmark Analysis of Searches. These were reported in a separate data field than the other Outcomes (Verbal Warning, Citation, Arrest, etc.), allowing for the possibility of a reliable analysis[5]. An analysis of at least some dimension of post-stop outcomes can be helpful for several reasons. For one, they capture a key moment of discretion in officer behavior; as several researchers have shown (cf. Epp, Maynard-Moody, and Haider-Markel 2014; Baumgartner, Epp, and Shoub 2018), officers make several key decisions *after* the stop that greatly influence the overall experience of the motorists stopped. For another, including an analysis of post-stop outcomes here allows for a further dimension of understanding what elements of traffic stops may be contributing to the overall disparities associated with criminal justice system contact. This understanding can be especially relevant for policy makers and police administrators interested in reducing such disparities.

For analyzing Searches it is possible to make use of a Stop Benchmark in addition to the Collision Benchmark used for the previous analyses in this report. While a Collision Benchmark estimates the expected rate of stops, reasons for stop, or outcome of stop based on the frequency in which a given demographic group is represented in *collisions* (which are as a close

[5] See Conclusions & Recommendations as well as the Outcome of Contact (Outcome of Stops) section in the Appendix for a further explanation of the limitations of an Outcomes analysis based on the data provided.

approximation of the overall motorist population), a Stop Benchmark instead compares those rates against the population of *stopped motorists*, which serve as the “benchmark” for that analysis.

As with all benchmarks, there are benefits and drawbacks to each of these two choices. A Collision benchmark is a very powerful tool to identify whether the overall rate of traffic stops illustrates evidence of disparities. However, it is only one such indicator. Analyses can be more robust if multiple benchmarks are used, which allows researchers to apply what Ross et al (2020) call the “preponderance of the evidence” method of analysis. The main drawback of using Stops as a benchmark for post-stop analysis is that, in cases such as Ann Arbor’s in which there is evidence that there are significant disparities in overall traffic stops, using the population of people who are stopped as a benchmark “bakes in” those disparities and, rather than identifying them as disparities, sets them as a basis for “expected” rates the phenomenon in question. This can have the unfortunate result of diminishing, or even eliminating, evidence of disparities. Despite this rather significant drawback, a Stop Benchmark can still be useful in identifying specific areas in which disparate outcomes are evidenced, once stops are made. For example, it may be the case that Searches evidence different frequencies or patterns of disparities; or, perhaps, in contrast to Overall Stops, none at all. An analysis using a Stop Benchmark may be able to identify such cases and, especially when used with other Benchmarks such as Collision benchmarks, may help researchers identify overall patterns in disparities. For that reason, the following analysis of Searchesmakes use of both Collision and Stop Benchmarks.

In other words, a Collision Benchmark retains the overall picture of disparities experienced by motorists, but may “drown out” the impact of additional post-stop outcomes and decisions that may involve additional dynamics. Conversely, a Stop Benchmark may erase or “drown out” the overall picture of disparities, but may be useful in adding nuance to the specific dynamics of post-stop outcomes. For that reason, our analysis places the Collision benchmark alongside a Stop benchmark for the post-stop analysis.

With that rather important framing, Table 11: Searches Odds Ratios below reports the odds ratio for Searches, tabulated by Race & Gender.

Table 11: Searches Odds Ratios

		Collision Benchmark		Stop Benchmark	
Race	Gender	Odds Ratio	Significance	Odds Ratio	Significance
Af Am	Female	1.22		1.10	
	Male	5.42	***	3.65	***
	Total	3.40	***	2.61	***
Asian	Female	0.14	***	0.13	***
	Male	0.33	***	0.23	***
	Total	0.23	***	0.18	***

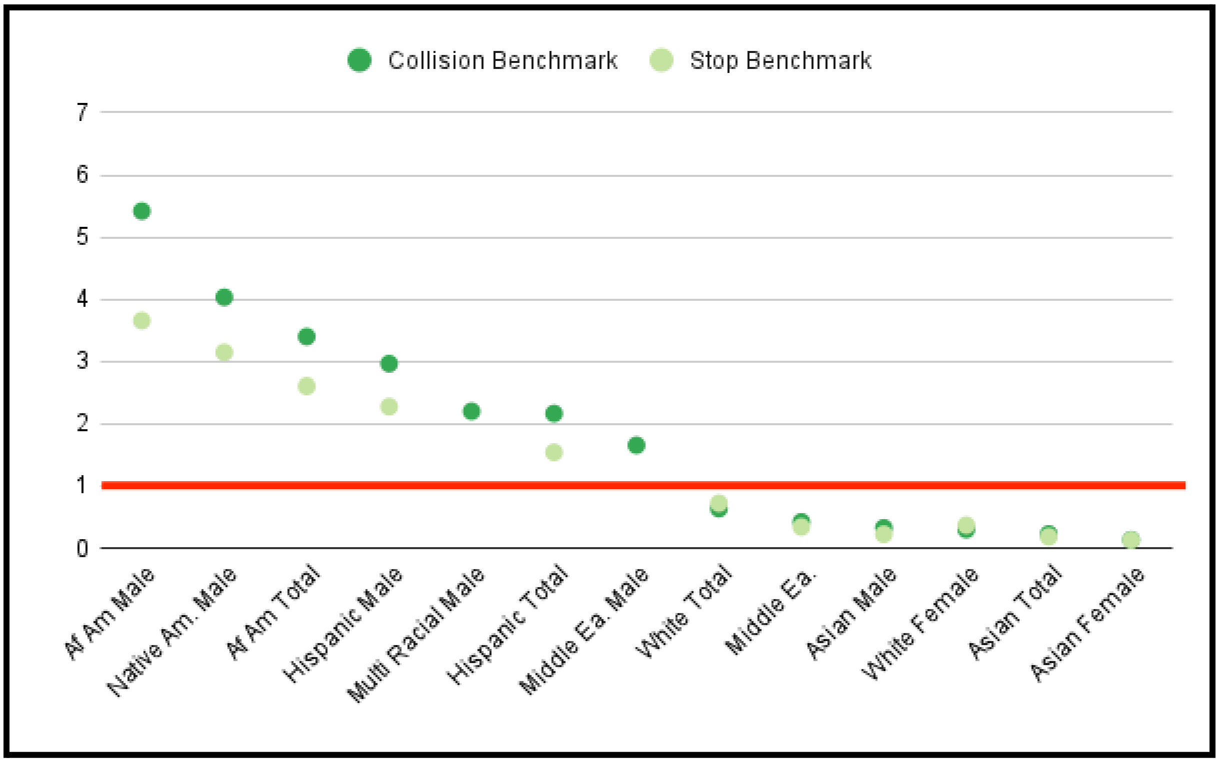
Table 11: Searches Odds Ratios					
		Collision Benchmark		Stop Benchmark	
Race	Gender	Odds Ratio	Significance	Odds Ratio	Significance
Hispanic	Female	0.74		0.46	
	Male	2.96	***	2.28	***
	Total	2.17	***	1.54	**
Middle Ea.	Female	0.42	**	0.34	**
	Male	1.66	**	0.92	
	Total	1.21		0.75	
Multi Racial	Female	0.46		0.26	
	Male	2.20	**	1.23	
	Total	1.49		0.84	
Native Am.	Female	0.80		1.21	
	Male	4.03	**	3.14	**
	Total	2.22		2.38	
Pacific Isl.	Female	0.00		0.00	
	Male	3.36		2.43	
	Total	2.16		1.63	
White	Female	0.30	***	0.37	***
	Male	0.97		1.03	
	Total	0.64	***	0.73	***

Figure 5, below, illustrates the statistically significant odds ratios for Searches, listed in rank order. The red line highlights the Odds Ratio of 1.00, the expected rate for such stops based on the given demographic's representation in the overall motorist population. As expected, the Stop benchmark closely mirrors the Collision benchmark, although it does indeed reflect a lower degree of disparity across the board. However it is important to understand that the Stop Benchmark frequencies do not reflect disparities in stops–the stop benchmark measures only disparate outcomes *after* the initial stop, not in the stop itself. Taken together, therefore, the Collision and Stop Benchmarks indicate a similar but distinct mechanism producing disparities.

With that understanding in place, our analysis indicates the largest disparity is for African American male motorists, who are searched at a rate between 5.42 times (for the Collision Benchmark) to 3.65 times (for the Stop Benchmark) more often than would be expected, based on each respective benchmark. They are followed by Native American Male motorists (between 4.03x to 3.14x more often than expected), Hispanic Male motorists (between 2.17x to 1.54x more often than would be expected), and Multi-Racial Male motorists (between 2.2x to 1.23x times more often than would be expected). Middle Eastern Male motorists show a significant result for the Collision Benchmark only, which indicates they are stopped 66% more often than would be expected.

Conversely, White motorists as a whole (36% to 27% less likely) and White Female motorists in particular (70% to 63% less likely), are Searched significantly less often than would be expected. They are joined by Asian Female motorists (77% to 82% less likely), Asian Male motorists (67% to 77% less likely), and Middle Eastern Female motorists (58% to 66% less likely) as being less likely to be Searched than would be expected based on the respective Benchmarks.

Figure 5: Significant Searches Odds Ratios by Collision & Stop Benchmarks



Discussion

Overview of Results

Before a more detailed summary of the above findings, it is important to return to the original goals and limitations of this analysis. This study was designed to identify and give conceptual nuance to disparities in Ann Arbor Police Department traffic stops. It was not designed to evaluate causal relationships, and can therefore neither affirm nor negate hypotheses as to the causal determinants of the disparities documented herein. Specifically, this should be understood to mean that it is neither *evidence for* nor *exoneration from* any racial bias, explicit or implicit, as a causal factor in the results; disparities can be evidence of racially-motivated hostile police practices, but they can also be caused by factors other than racial animus (Warren et al. 2006). Furthermore, no single data point, frequency, or odds ratio can serve as a definitive description of the existence nor contours of such disparities. For this reason, we have followed the “preponderance of the evidence” approach (Ross, Kalinowski, and Barone 2020) to our analysis, which endeavors to give narrative and conceptual clarity to what can be “noisy,” complex, or even contradictory data, so that policy makers and administrators may have tools to guide their own efforts.

Having said that, the existence of disparities in AAPD traffic stops and the differential frequencies of specific policing practices are consistently and clearly evident. Our analysis identified significant disparities across every dimension examined, with non-white motorists being Stopped and Searched more frequently and White motorists being Stopped and Searched less frequently than would be expected in every instance. These disparities were not uniform across racial categories nor across various Reasons for Contact, however every racial group except White motorists showed some evidence of being the target of police intervention at a higher rate than would be expected.

In our analysis of Overall Traffic Stops, the disparities between White motorists and all other drivers were especially apparent. All non-white racial groups for which we were able to obtain a significant result indicated an increased frequency of being subject to traffic stops than would be expected based on their representation in the population of total motorists, as estimated by the Collision Benchmark. The highest such disparity was for Middle Eastern Male drivers, who were stopped 81% more often than would be expected. When taken as a total, all racial groups exhibited an increased likelihood of being stopped, although this picture was somewhat complicated by cross-tabulating the frequencies for both race and gender.

Similarly, our analysis of particular Reasons for Contact (reason for the traffic stop) offers both a stark portrait of disparities in traffic stops in overview and important differences between groups when analyzed in further detail. As with Overall Traffic Stops, White motorists (taken as a whole) and White Female drivers, especially, were stopped less often than would be expected for all three Reasons analyzed. The same was the case for White Male motorists, with the exception of Traffic Violations, which did not yield a significant result. Conversely, the only groups stopped more frequently than would be expected across all three Reasons were non-White motorists, although the frequencies with which various racial groups were represented across the three Reasons for Contact were not uniform. The largest disparity identified in this section of the analysis involved stops for reason of Equipment Violation, for which Multi Racial Male motorists were stopped 3.01x more frequently than would be expected, African-American male drivers were stopped 2.41x more likely than would be expected, and Multi Racial Female motorists were stopped 2.00x more frequently than would be expected. However remarkable disparities existed for other racial groups for other Reasons for Contact: Asian Male motorists are stopped 93% more frequently than would be expected for Speeding Violations, and Middle Eastern Male motorists are stopped for Traffic Violations 86% more frequently than would be expected.

However, the largest disparities were evident in Searches, the lone example of a post-stop outcome that we were able to analyze. Such post-stop outcomes are important because they represent a crucial set of opportunities for officer discretion and department-specific policy and practice guidelines (Baumgartner, Epp, and Shoub 2018; Epp, Maynard-Moody, and Haider-Markel 2014; Warren et al. 2006). Across the board, non-while Male motorists were subject to Searches more frequently than would be expected based on either a Collision or Stop benchmark. The largest such disparities involved African-American Male drivers, which occurred between 5.4x to 3.65x more often than would be expected. They are followed by Native American Male motorists (between 4.03x to 3.14x more often than expected), Hispanic Male motorists (between 2.17x to 1.54x more often than would be expected), and Multi-Racial Male motorists (between 2.2x to 1.23x times more often than would be expected).

Comparisons with Previous Studies

It may be helpful to place the above results in the context of previous studies which have conducted similar traffic stop data analysis, both at the local and national scale. For example, Warren et al (2006) compared the experience of Black and White drivers in North Carolina, tabulating the results by Age group in addition to race and gender. For stops conducted by local (as opposed to State or Highway patrol agencies) they found Odds ratios of between 1.36 (for African American Male motorists aged 18 to 22) to 1.90 (for African American Male motorists aged 50 and older). Similarly, Seguino et al (2021) found an odds ratio of 1.81 for Black motorist traffic stops in Vermont using a crash data benchmark. The odds ratio observed for African American male motorists of this study was 1.48.

However, it may be more helpful to contextualize the above analysis through a comparison of similar such analyses conducted at the local and state level. Over the past 20 years, there have been three such analyses conducted: a 2004 analysis of AAPD traffic stop conducted by Lamberth Consulting (J. C. Lamberth and Lamberth Consulting 2004), a 2018 analysis of AAPD traffic stops conducted by Dolan Consulting (Dolan Consulting Group 2018), and a 2021 analysis of state-wide Michigan State Police traffic stops conducted by a research team at Michigan State University (Wolfe, Carter, and Knode 2021). While none of these previous studies follow exactly the methodology of this analysis, a comparison of the different approaches, and their various conclusions, can offer a framework to understand the current analysis.

2004 Lamberth Consulting Report

In 2004, the national consulting group Lamberth Consulting conducted an analysis of Ann Arbor Police Department Traffic stops, led by a team including Lamberth Consulting President Karl Lamberth, PhD, and future Washtenaw County Sheriff Jerry Clayton. This study was designed to address several key questions: "Is there evidence of racial profiling in the AAPD? Which minority groups (i.e., Blacks and Hispanics), if any, are targeted? In which locations is profiling likely to occur? Are there special circumstances that might be interpreted as biased policing?" (J. C. Lamberth and Lamberth Consulting 2004, 2). In order to address these questions, Lamberth Consulting conducted an Observational Benchmarking analysis, with trained observers stationed at several key intersections throughout Ann Arbor[6]. It then used this benchmark as a basis for comparison with actual AAPD traffic stops at each of the observed locations for which there were a sufficient number of stops to conduct the analysis[7].

Lamberth Consulting found that Blacks were stopped at a higher rate in each of the 7 instances they were able to analyze, ranging from 20% more likely to be stopped (at the intersection of S. University & State) to 70% more likely to be stopped (at the intersection of Stadium & Washtenaw). However, despite a Chi Square test indicating that five of these seven odds ratios could be established with a level of statistical significance between $p<0.001$ and $p<0.03$, the authors concluded that odds ratios of 1.5 and below are "benign" and not evidence that AAPD is targeting Black motorists (J. C. Lamberth and Lamberth Consulting 2004, 28). They concluded the report, without citation or reference, with the assertion that the "overall odds ratio of 1.5 for Black motorists is one of the lower odds ratios with regard to Blacks that we have seen in our analysis of jurisdictions around the country" and reiterated their conclusion that the overall results provided "no evidence overall that the AAPD is targeting Black motorists for stops" (J. C. Lamberth and Lamberth Consulting 2004, 30).

There are several key differences between the Lamberth study and the SMART analysis. One key difference between the two studies is the method of establishing benchmarks against which to compare traffic stop rates. We have discussed benchmark analyses in our brief overview of benchmarks (See Observational Benchmarks, above). In summary, such benchmark analyses can be an interesting, if labor intensive, method for estimating the demographics of a motorist population. However the additional benefits of this cost and labor above other benchmarks such as collision data remain unclear. Additionally, they introduce several additional challenges, each of which apply to the Lamberth Consulting study: their targeted coverage (at only a small number of pre-selected intersections) may mean the analysis misses overall trends throughout the jurisdiction. It can also mean that key locations are excluded from the analysis. Additionally, the current SMART analysis provides an analysis across several Racial categories as well as Gender, whereas the Lamberth study does not address Gender–either as an isolated factor or

[6] These observers were stationed at S. University & State, S. University & Washtenaw, Fourth & Huron, Hubbard & Huron Parkway, Stadium & Washtenaw, Stadium & Main, Eisenhower & State, Miller & Newport, and Seventh & Pauline.

[7] Although observational benchmarking was conducted at the intersections of both Miller & Newport and Seventh & Pauline, these were eventually excluded from the analysis because the total number of stops conducted at these locations did not enable an appropriate level of statistical significance (J. C. Lamberth and Lamberth Consulting 2004, 27).

its cross-tabulation with Race–nor did it find, at the time, that the number of Hispanic, Middle Eastern motorists were sufficient to conduct an analysis (J. C. Lamberth and Lamberth Consulting 2004, 29), nor are Asian motorists addressed in the report. Finally, the Lamberth report includes no analysis of the stated reason for Stops, nor of post-stop outcomes.

Despite these differences, the underlying results for which comparison is possible–namely the odds ratio for Black motorists being stopped– are remarkably similar. The Lamberth study found overall that Black drivers are 50% more likely to be stopped than would be expected based on their representation in the benchmark population. The SMART analysis finds that African American drivers are, overall, 30% more likely to be stopped than would be expected.

Despite these similar results, the conclusions Lamberth Consulting draws differ from SMART's own in important ways. First, Lamberth consulting concludes that odds ratios below 1.5 are "benign" and "not evidence that Black drivers are being targeted". However, there is no such "benign" standard in the contemporary literature, nor can such a thing be mapped onto an objective indicator. Such analyses can only indicate (1) whether a particular group is stopped at the rate that would be expected given their representation in some benchmark population; and (2) with what degree of confidence we can conclude any observed variations in traffic stop rates are *not* a result of random chance. The question of whether a statistically significant rate of disparity is "benign" or acceptable is a subjective question best addressed by policy-making bodies in partnership with the communities they represent.

Another key difference between the Lamberth study and the current SMART analysis pertains to the underlying research question, or aim. The Lamberth study aims to find evidence of whether certain groups are being targeted, a question to which they conclude that there is insufficient evidence to assert affirmatively. This is not merely an analysis of the *existence* of disparities but also their *causes*. There are two main problems with framing the research question in this manner: (1) It is very difficult to make causal conclusions derived from traffic stop data collected by police. Such data records very little information about officer's reasoning (either stated or not) behind their actions and decisions, nor does it test for various degrees of personal bias (either implicit or explicit). While certain analytic tools (for example the Veil of Darkness method and various versions of Internal Benchmarking) may achieve something closer to evidence of active targeting, the form of analysis conducted by both the Lamberth study and by SMART is mainly designed to assess the *existence* of disparities, not their causes. In this respect, both the Lamberth and current SMART report unequivocally provide evidence of disparities in traffic stops by the Ann Arbor Police Department.

2018 Dolan Group Report

In 2018, the Dolan Consulting Group was contracted by the City of Ann Arbor to examine the motor vehicle stops made by the AAPD to look for "a pattern of practice" that may reveal "any signs of biased policing" (Dolan Consulting Group 2018, 1). Their analysis examined the 2017 traffic stop and collision data collected by the AAPD against two key benchmarks: traffic collisions and criminal suspect descriptions. Dolan conducted this analysis by disaggregating this data into 4 Administrative groups (corresponding to the four AAPD Patrol Areas), the stops in each of these Administrative Units were then disaggregated into two types of stops ("traffic violations" and "criminal investigations"). Traffic violation stops were then further disaggregated into four distinct time slots (5:01am-11am, 11:01am-5pm, 5:01pm-11pm, & 11:01pm-5:00am) while "Criminal Investigation" stops were disaggregated into two time slots (5:01am-5:00pm and 5:01pm-5:00am). These 22 units were then analyzed separately by comparing the rate of traffic stops against benchmark data along both the dimensions of Race and Gender. As a result of this analysis, Dolan Consulting concluded that there was no evidence or signs of biased policing:

> "Overall, group disparities within vehicle stops by the Ann Arbor Police Department were very small. The greatest disparity occurred among male drivers, who were estimated to have been 6% more likely to be stopped than expected if no bias was present. The second greatest disparity was revealed among African-American drivers, who were estimated to have been 1% more likely to have been stopped than expected if no bias was present. The least amount of disparity involved white drivers, who were estimated to have been 0.5% more likely to have been stopped if no bias was present. No disparities were revealed regarding stops of Asian-American and Pacific Islander drivers, or those drivers included within the 'all other groups' category." (Dolan Consulting Group 2018, 86–87)

A comparison between the 2018 Dolan Consulting analysis and the current SMART analysis offers an excellent opportunity to illustrate the ways in which seemingly minor methodological and reporting decisions can shape the ultimate analysis and recommendations of such studies. The same data which serves as the basis for the Dolan report– 2017 AAPD traffic stops and reported collision data–is included in SMART's current analysis, although SMART's analysis includes data from 2018 and 2019 as well. However, the findings of the two reports differ in important ways. Similar to the 2004 Lamberth analysis for Ann Arbor (above) and the 2021 MSU report for the Michigan State Police (cf 2021 MSU/MSP Report below), SMART's analysis finds significant racial disparities in the rates African American motorists are stopped. In contrast to these three studies, the Dolan report finds only very few, and very small, disparities, leading them to conclude there is no evidence of biased policing in the data they considered. A further discussion of the methodological choices of the Dolan study is therefore warranted in order to understand why its conclusions are so dramatically different from other such studies, both previous and subsequent.

One key difference exists in the stated goal of the analysis. The Dolan Report repeatedly claims to offer an assessment of racial *bias* on the part of AAPD officers. However– like this report, the Lamberth Report, and the MSU/MSP report–their analysis is more accurately described as assessing evidence of racial *disparities*. Neither their methodological design nor data speak to the motivational or psychological factors guiding police decision making, only the frequency of certain outcomes of those decisions.

Another key distinction between the Dolan report and SMART's is the use of comparative benchmarks.The Dolan report conducts its analysis using two benchmarks. The first of these benchmarks is collision data, which is a common benchmark used in such analyses, such as in the current analysis and MSU/MSP report (below). The second benchmark against which certain traffic stops are compared consists of crime victim descriptions of criminal suspects. The authors make use of this benchmark in order to look for evidence of disparities within “criminal investigative” stops, which they distinguish from “traffic violation” stops. This methodological choice is highly unusual. We are unfamiliar with any other study making use of such a benchmark, nor do the report authors cite previous models for its use, nor do they refer to any peer reviewed literature that might support such a methodological choice. This lack of citational support is especially unusual because the existing literature is, in contrast, very clear that victim reports of offender race are notoriously problematic and hard to interpret (Beckett, Nyrop, and Pfingst 2006; Shah and Pease 2009; Xie and Baumer 2019; Johnson, Petersen, and Martinez 2022; Flexon et al. 2023) and therefore unlikely to serve as an objective baseline for such comparisons. One case in point is that, using this benchmark, the resulting expected rate which African Americans would be expected to be the target of traffic stops is significantly higher (40.5%, by Dolan’s calculations) than it would be by using their representation in collisions (11.9%) (Dolan Consulting Group 2018, 15–16). In other words, using Dolan’s crime victim reporting benchmark, one would expect to see African Americans represented at a rate of around 40% of all stops, while using the traffic collision benchmark one would expect that rate to be closer to 12%. Any assessment of whether African Americans are in actuality stopped at a higher or lower rate than would be expected is directly tied to which benchmark is used to set those expectations. For stops they classify as “criminal investigative stops”, that expectation is set very high using a benchmark with little to no methodological support in the existing literature.

Arguably more impactful on the conclusions of the Dolan report are the dual methodological decisions to report results only in disaggregated form and to use an unusually high standard of probability in order to report results as significant. While the decision to disaggregate data can be a useful tool to lend nuance to a comprehensive analysis, if such disaggregation occurs without a comprehensive or “overview” presentation of the data, it may obscure overarching trends and frequencies. This appears to be the case with the Dolan analysis, which never offers an assessment of AAPD traffic stops in aggregate. This methodological decision is even more confounding, in that the Units into which they disaggregate the data have no consistent practical, demographic, or administrative bearing on the conduct of AAPD traffic stops; they do not reflect distinct administrative entities or policing cohorts in the way they might in larger US cities. The decision by the Dolan report to disaggregate their data therefore lowers the overall number of cases assessed in any instance, making particular conclusions more difficult, without any apparent analytic benefit.

A further hurdle for drawing significant results in the Dolan report is their choice of an unusually high p-value, which they set at $p < .001$, or 99.9% confidence. SMART is unaware of any other reputable traffic stop analysis that requires this level of confidence in order to report results as significant. While some analysts may distinguish results that meet or exceed that level of confidence (as does this analysis, the Lamberth report, and the 2021 MSU/MSP report), best practice in social science data reporting (which includes traffic stop data analysis) is to report results significant at a value of $p < 0.5$, or 95% confidence, a more lower “bar”. Taken together with their choice of benchmark and disaggregation strategy, their methodological decisions consistently make it more difficult to find and report evidence of disparities, ultimately leading to substantially different conclusions about racial disparities in AAPD traffic stops.

2021 Wolfe et al (MSU/MSP) Report

In 2021, researchers at Michigan State University (MSU) collaborated with the Michigan State Police external benchmark analysis of Michigan State Police (MSP) traffic stops conducted during 2020 (Wolfe, Carter, and Knode 2021). In addition to several other forms of analysis, the conducted a Collision Benchmark analysis similar to the one presented here. On the state level, their analysis found that African-American motorists are 18% more likely to be stopped than would be expected based on this benchmark, while Hispanic motorists (6% less likely) and Asian motorists (40% less likely) were stopped at a frequency that was *less* likely than would be expected. In addition to this statewide aggregate analysis, they also disaggregated their data according to MSP administrative district, of which Ann Arbor is located in District 1. They found that, in MSP District 1, African-American motorists are stopped 83% more often than would be expected. The odds ratios for Hispanic and Asian motorists in District 1 did not yield significant results.

Comparison Summary

In general, this survey of previous analysis can provide two insights: first, the analysis of traffic disparities presented here places contemporary disparities in Ann Arbor traffic stops as largely in line with previous such analyses, both in Ann Arbor and elsewhere. While the existence of such disparities is clear and evident, they largely mirror–and in some instances represent an improvement upon–analyses conducted at other times or other places. The notable exception to this consensus is the analysis conducted by the Dolan Group for the City of Ann Arbor in 2018; those differences can in large part be attributed to the unique methodological choices of that analysis, described above. The 2004 Lamberth report found that African Americans were 50% more likely to be stopped than expected, while the 2021 Wolfe report found statewide that African Americans are 18% more likely and, in District 1, 83% more likely to be stopped. These results should be compared with the estimation here that African American motorists as a whole are 30% more likely to be stopped than would be expected.

Conclusions & Recommendations

To recapitulate the above discussion, our analysis conclusively identified significant disparities across every dimension examined, with non-white motorists being Stopped and Searched more frequently and White motorists being Stopped and Searched less frequently than would be expected in every instance. These disparities were not uniform across racial categories nor across various Reasons for Contact. Some of the largest disparities identified in this analysis involve African-American male drivers for stops initiated for Equipment Violations (which occurred 2.41x more likely than would be expected) as well as for Searches after the initial stop (which occurred between 5.4x to 3.65x more often than would be expected). Despite these significant disparities, the above results place AAPD traffic stops generally in line, neither dramatically better or worse, than similar analyses.

Recommendations

The main charge for SMART was to conduct an analysis of potential disparities in AAPD traffic stops as well as to provide some tools through which policy makers, administrators and oversight practitioners might better understand them, not offer a comprehensive set of policy revisions for the practice of public safety. Nevertheless, based on this analysis and the experience of the project, SMART feels it important to offer a set of general recommendations for policy and administrative action that the above analysis may engender. These recommendations fall into two general categories: Data Management and Policy Development.

Data Management Recommendations

It is important to note that significant advancements in data management and transparency have already been achieved through the partnership between AAPD, ICPOC, and SMART that led to this analysis. At the initial stages of this collaboration, as each of these partners were in discussions about available data and the best mechanisms for its transmission, AAPD was able to successfully petition its main software vendor, CLEMIS, of its expectation to have access to the “age" data element so that it may be more easily reported. This has been completed by the vendor, and is now in the dataset for all regional police forces who make use of this service. Additionally, both Ann Arbor’s Transportation Commission and Independent Community Police Oversight Commission (ICPOC) passed resolutions recommending the regular release of AAPD traffic data, in large part modeled after the dataset developed in the course of this project. Both resolutions were subsequently approved by City Council and await implementation as a City Ordinance[8]. Finally, through a process independent from this particular collaboration, AAPD was able to work with the CLEMIS vendor to launch a new data transparency dashboard[9].

---

[8] Cf. Transportation Commission Traffic Stop Transparency Resolution of 3/16/22 (http://a2gov.legistar.com/LegislationDetail.aspx?ID=5524045&GUID=BE7D1529-7A1E-450E-ACAE-9C5768DE2DFA&Options=&Search=) and ICPOC Resolution To City Council Regarding Ann Arbor Traffic Transparency of 3/22/22 (http://a2gov.legistar.com/LegislationDetail.aspx?ID=5547235&GUID=341774CE-4912-41FE-BECD-C7CCF8B8DD20&Options=&Search=), both approved by City Council on 4/18/22

[9] Ann Arbor PD Transparncy Dashboard https://portal.arxcommunity.com/dashboards/community/mi-ci-annarbor-pd

In addition to these welcome developments, SMART offers specific recommendations for more consistent and robust data collection and publication practices, especially pertaining to post-stop outcomes, which would enable more nuance along this dimension in future analyses:

- SMART recommends implementing the data transparency resolutions developed by the Transportation Commission and ICPOC, and approved by City Council, as an official City Ordinance. This would ensure the consistent practice of data transparency necessary for policy makers, oversight practitioners, and general citizenry to assess and make recommendations about current public safety practice in Ann Arbor.
- Furthermore, SMART offers recommendations for the nature and format of the data that would be shared:
    - In order to be of utility for future research, this data should be made available in a downloadable file format such as .csv or .html.
    - AAPD and the City of Ann Arbor may want to reassess the racial categories available to officers in the CLEMIS system, potentially modifying them to more closely align with US census data.
    - Stop Outcomes should be reported in such a way that the frequency of cases in which multiple outcomes (Verbal Warnings, Citations, and Arrest) resulting from a single Stop are reported. This will enable more robust Outcomes Analysis, such as an analysis of Hit Rates, which may speak more directly to racially targeted police action.
    - Finally, in order to facilitate future analyses, it would be ideal to offer benchmark data, such as Collision data or Inter-Twilight time windows (for Veil of Darkness benchmarking), alongside Traffic Stop data. Such data can be difficult to locate for non-specialists, but are essential for calculating and assessing the types of Odds Ratios included in this report.
- Such regular and thorough data transparency could facilitate more nuanced future analyses, including: Internal Benchmarking, Veil of Darkness Benchmarking, Analysis of the Geo-Spatial distribution of AAPD interventions, and further cross-tabulation of frequencies by Age (in addition to race and Gender, as in this analysis).
- Finally, SMART would like to highlight for future researchers the utility of qualitative analyses in addition to the quantitative data described here. Such qualitative analyses might offer richer insight into the decision-making processes of AAPD officers, and their context.

Policy Development

SMART also offers some suggestions for next step policy developments based on the findings of this report.

- SMART recommends that police administrators, elected officials and oversight practitioners use this analysis to inform their priorities, taking into account especially the Reasons for Contact and post-contact Outcomes which exhibit the largest disparities.
- One potential solution is to examine the policy and procedural pathways that produce those disparities in order to assess whether there are alternative legislative remedies or procedural alternatives. For example, some communities have made use of traffic stop

data analyses such as this as an impetus to explore innovative pathways for policing regulatory or non-moving violations, such as would be grouped under Equipment Violations here (Susan Nembhard, and Kathryn L.S. Pettit 2023; City of Philadelphia Office of the Mayor 2021). This may be an especially attractive solution in Ann Arbor, as such violations represent some of the biggest disparities for African American motorists. For example, in a related but larger scale national study analyzing traffic stop disparities, Pierson et al conclude that "the downstream effects [of traffic stops] can be injurious even if individual stop decisions are not directly affected by the colour of one's skin. Similarly, enforcement of minor traffic violations, like broken tail lights—even if conducted uniformly and without animus—can place heavy burdens on black and Hispanic drivers without improving public safety" (Pierson et al. 2020).

- SMART also urges policy makers to consider how current and future initiatives may contribute to existing disparities, especially in their enforcement. This is especially important as the City pursues new transportation goals[10].

[10] City of Ann Arbor "Moving Together Toward Vision Zero Comprehensive Transportation Plan" (June 2021) https://www.a2gov.org/departments/engineering/Documents/Ann%20Arbor%20Moving%20Together_Final%20Plan_June%202021.pdf

Appendix

Outcome of Contact (Outcome of Stops)

The AAPD data shared with SMART for this analysis records two dimensions of information: whether a Search was conducted (as well as what type of Search, if one was conducted), and the final outcome of the stop (Verbal Warning, Citation, Assist, or Arrest). The data shared with SMART contains only one possible outcome for each stop. However, other analyses have included data that had the potential to record more than one such outcome–for example, cases in which there was a citation for one offense and an arrest for another (cf. Wolfe, Carter, and Knode 2021, 18). This serves as one potential limitation of any analysis of this, potentially causing misleading results in several instances. Further complicating any analysis of Outcomes in the data provided to SMART and discussed in this Report, it is unclear how the singular reported outcome is chosen. For example, if Outcomes are reported according to an escalating scale (from whether simply a Verbal warning was issued, to whether at least one Citation results, to whether the stop resulted in Arrest), it may be that results are skewed so as to make it appear those with the harshest penalties (Arrests) received fewer Citations and Verbal Warnings than they may have received, or conversely, that those groups not receiving the harshest penalties received higher levels of Citations and Verbal Warnings than is the actual case. However, in discussion with AAPD leadership, there does not seem to be a standard policy or practice for which outcome is recorded in such cases, making it extremely difficult to draw broadly valid conclusions from existing AAPD Outcomes data (for a further discussion of this challenge see Discussion below).

For those reasons, Outcome of Contact was not included in the Descriptive section of this report nor in the Analysis. However, we have included the basic descriptive statistics here, as reported by AAPD to SMART under the limitations described above. Possible outcomes for traffic stops in this data set include: Arrest, Citation, Verbal Warning or Assist. Verbal warnings were the most common outcome of traffic stops, accounting for 55.38% of the total stops. Citations were also common, at 43.03% of stops. Assists and Arrests accounted for just under 2% of total stop outcomes. See Table 4 Outcome of Contact.

Table 4 Outcome of Contact

Outcome of Contact	n	%
Arrest	446	1.3%
Assist	102	0.3%
Citation	14,907	43.0%
Verbal Warn	19,176	55.4%
Grand Total	**34,631**	**100.0%**

Works Cited

Alpert, Geoffrey P., Michael R. Smith, and Roger G. Dunham. 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.

Barnum, Chris, Jared Miller, and Grant Miller. 2015. "An Evaluation of an Observational Benchmark Used in Assessing Disproportionality in Police Traffic Stops: A Research Note." *Policing: A Journal of Policy and Practice* 9 (4): 405–17.

Baumgartner, Frank R., Leah Christiani, Derek A. Epp, Kevin Roach, and Kelsey Shoub. 2017. "Racial Disparities in Traffic Stop Outcomes." *Duke FL & Soc. Change* 9: 21.

Baumgartner, Frank R., Derek A. Epp, and Kelsey Shoub. 2018. *Suspect Citizens: What 20 Million Traffic Stops Tell Us about Policing and Race*. United Kingdom ; New York, NY: Cambridge University Press.

Beckett, Katherine, Kris Nyrop, and Lori Pfingst. 2006. "RACE, DRUGS, AND POLICING: UNDERSTANDING DISPARITIES IN DRUG DELIVERY ARRESTS*." *Criminology* 44 (1): 105–37. https://doi.org/10.1111/j.1745-9125.2006.00044.x.

City of Philadelphia Office of the Mayor. 2021. *Executive Order 6-21: Implementation of Driving Equality Policy*. https://www.phila.gov/media/20211109145453/executive-order-2021-06.pdf.

Dolan Consulting Group, LLC. 2018. "Evaluating Fairness in Traffic Stops by the Ann Arbor Police Department: Final Report."

Engel, Robin Shepard, and Jennifer M. Calnon. 2004. "Comparing Benchmark Methodologies for Police-Citizen Contacts: Traffic Stop Data Collection for the Pennsylvania State Police." *Police Quarterly* 7 (1): 97–125.

Epp, Charles R., Steven Maynard-Moody, and Donald P. Haider-Markel. 2014. *Pulled over: How Police Stops Define Race and Citizenship*. The Chicago Series in Law and Society. Chicago ; London: The University of Chicago Press.

Flexon, Jamie L., Lin Liu, Richard G. Greenleaf, and Nerissa James. 2023. "Income and Calling the Police: Examining a Nuanced Relationship Toward Theoretical Refinement." *Victims & Offenders*, February, 1–22. https://doi.org/10.1080/15564886.2023.2175095.

Fridell, Lorie A. 2005. "Understanding Race Data from Vehicle Stops: A Stakeholder's Guide." Washington, D.C: Police Executive Research Forum.

Grogger, Jeffrey, and Greg Ridgeway. 2006. "Testing for Racial Profiling in Traffic Stops From Behind a Veil of Darkness." *Journal of the American Statistical Association* 101 (475): 878–87. https://doi.org/10.1198/016214506000000168.

Johnson, Oshea D., Nick Petersen, and Brandon P. Martinez. 2022. "Punishing Neighborhood 'Outsiders': Neighborhood Punishment Rates and the Spatial Mis(Match) Between Defendants' Residence and Arrest Locations." *Crime & Delinquency*, September, 001112872211177. https://doi.org/10.1177/00111287221117757.

Lamberth, John. 1994. "Revised Statistical Analysis of the Incidence of Police Stops and Arrests of Black Drivers/Travelers on the New Jersey Turnpike between Exits or Interchanges 1 and 3 from the Years 1988 through 1991."

Lamberth, John C. and Lamberth Consulting. 2004. "Ann Arbor Police Department Traffic Stop Data Collection Methods and Analysis Study." Ann Arbor, MI.

McLean, Kyle, and Jeff Rojek. 2016. "Traffic Stops, Race, and Measurement." *The Handbook of Measurement Issues in Criminology and Criminal Justice*, 452–72. https://doi.org/10.1002/9781118868799.ch20.

Neil, Roland, and Christopher Winship. 2019. "Methodological Challenges and Opportunities in Testing for Racial Discrimination in Policing." *Annual Review of Criminology*. Annual Reviews Inc. https://doi.org/10.1146/annurev-criminol-011518-024731.

Pierson, Emma, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, et al. 2020. “A Large-Scale Analysis of Racial Disparities in Police Stops across the United States.” *Nature Human Behaviour* 4 (7): 736–45. https://doi.org/10.1038/s41562-020-0858-1.

Ridgeway, Greg, and John MacDonald. 2010. “Chapter 7. Methods for Assessing Racially Biased Policing:” In *Chapter 7. Methods for Assessing Racially Biased Policing*, 180–204. New York University Press. https://doi.org/10.18574/nyu/9780814776155.003.0007.

Ross, Matthew B., Jesse J. Kalinowski, and Kenneth Barone. 2020. “Testing for Disparities in Traffic Stops: Best Practices from the Connecticut Model.” *Criminology and Public Policy* 19 (4): 1289–1303. https://doi.org/10.1111/1745-9133.12528.

Shah, Rabindra, and Ken Pease. 2009. “Crime, Race and Reporting to the Police.” *The Howard Journal of Criminal Justice* 31 (3): 192–99. https://doi.org/10.1111/j.1468-2311.1992.tb00741.x.

Smith, Michael R., Robert Tillyer, Caleb Lloyd, and Matt Petrocelli. 2021. “Benchmarking Disparities in Police Stops: A Comparative Application of 2nd and 3rd Generation Techniques.” *Justice Quarterly* 38 (3): 513–36. https://doi.org/10.1080/07418825.2019.1660395.

Stephanie Seguino, Nancy Brooks, and Pat Autilio. 2021. “Trends in Racial Disparities in Vermont Traffic Stops, 2014-19.” https://www.uvm.edu/sites/default/files/Department-of-Economics/seguino%20studies/Seguino_Brooks_Autilio_All_Vermont_revised_12.21.pdf.

Susan Nembhard, and Kathryn L.S. Pettit. 2023. “Using Traffic Stop Data for Change in Mecklenburg County, North Carolina.” *Urban Institute* (blog). March 6, 2023. https://www.urban.org/catalyst-grant-program-insights/using-traffic-stop-data-change-mecklenburg-county-north-carolina.

Taniguchi, Travis A., Joshua A. Hendrix, Alison Levin-Rector, Brian P. Aagaard, Kevin J. Strom, and Stephanie A. Zimmer. 2017. “Extending the Veil of Darkness Approach: An Examination of Racial Disproportionality in Traffic Stops in Durham, NC.” *Police Quarterly* 20 (4): 420–48. https://doi.org/10.1177/1098611117721665.

Tillyer, Rob, Robin S. Engel, and Jennifer Calnon Cherkauskas. 2010. “Best Practices in Vehicle Stop Data Collection and Analysis.” *Policing* 33 (1): 69–92. https://doi.org/10.1108/13639511011020601.

Warren, Patricia, Donald Tomaskovic-Devey, William Smith, Matthew Zingraff, and Marcinda Mason. 2006. “Driving While Black: Bias Processes and Racial Disparity in Police Stops*.” *Criminology* 44 (3): 709–38. https://doi.org/10.1111/j.1745-9125.2006.00061.x.

Wolfe, Scott, Travis Carter, and Jedidiah Knode. 2021. “Michigan State Police Traffic Stop External Benchmarking: A Final Report on Racial and Ethnic Disparities.” Michigan State University School of Criminal Justice.

Xie, Min, and Eric P. Baumer. 2019. “Crime Victims’ Decisions to Call the Police: Past Research and New Directions.” *Annual Review of Criminology* 2 (1): 217–40. https://doi.org/10.1146/annurev-criminol-011518-024748.