ANN ARBOR POLICE DEPARTMENT TRAFFIC STOP DATA COLLECTION METHODS AND ANALYSIS STUDY

Report for the City of Ann Arbor

Written by: Dr. John C. Lamberth
Lamberth Consulting

February, 2004

LAMBERTH CONSULTING

Lamberth Consulting was formed in 2000 in an effort to provide racial profiling assessment, training, and communication services to universities, states, counties, cities, civil rights groups, litigators, and communities.

Dr. John Lamberth, CEO and founder of Lamberth Consulting, developed the nation's first racial profiling methodology in 1993. Since that time we have revised and adapted our methodology for highways, urban areas, suburban areas, and pedestrian populations. We have expanded our service offerings to include training solutions targeted towards law enforcement and community members, as well as communication planning services to help educate and inform all parties concerned about racial profiling issues.

Lamberth Consulting, LLC
120 Commons Court
Chadds Ford, PA 19317
phone 610.358.5700
fax 610.358.2890
on the web www.lamberthconsulting.com
email us at info@lamberthconsulting.com

LAMBERTH CONSULTING

PROJECT TEAM

Dr. John Lamberth
Dr. John Lamberth, CEO of Lamberth Consulting, has been working as a social psychologist for over 30 years, and has been consulting for over 20 years. His area of consulting expertise lies in statistics, surveying methodology, and social psychology. He has been qualified as an expert in these areas in State Courts in New Jersey, Pennsylvania, Iowa and Virginia and in Federal Courts in New Jersey, Connecticut and Virginia.

In 1993 Dr. Lamberth developed the first methodology used in the country to determine whether racial profiling was occurring. In *New Jersey v. Soto*, the Court relied upon his research methodology and statistical analysis in determining whether racial profiling occurred on the New Jersey Turnpike. In 1999, the Attorney General of New Jersey agreed that the New Jersey State Police were practicing racial profiling (Interim Report of the State Police Review Team Regarding Allegations of Racial Profiling @ www.state.nj.us/lps/intm_419.pdf). Since that time he has intensified his work in the area of racial profiling, assisting in litigation, and assisting police departments who are interested in voluntarily determining whether racial profiling is occurring in their jurisdictions.

Karl Lamberth
Karl Lamberth, President of Lamberth Consulting, has over ten years experience in the field of change management consulting to assist organizations and employees through the change process. From 1992 through 1999, he worked with Andersen Consulting in the State Government, Communications and Financial Services industry groups. In 1999 and 2000, Mr. Lamberth worked as the Director of Operations for CommerceOne's Global Services operation in the Southeast region.

Mr. Lamberth's primary areas of expertise are in instructional design, change management, project management, and communications planning. Mr. Lamberth has specific skills in training design and development, communications development and management, project planning, project control, quality assurance, risk management, and project reporting.

Jerry Clayton
Jerry Clayton, Vice President of Training, designed and developed Lamberth Consulting's training programs. Mr. Clayton currently oversees all training delivery, and regularly instructs Lamberth Consulting's training courses on Preventing Racial Profiling Practices, and Collecting Stop Data.

Mr. Clayton is a certified instructor in Interpersonal Communications, Cultural Diversity, Defensive Tactics, Aerosol Subject Restraint, Firearms, Special Weapons and Tactics, and Direct Supervision and Use of Force Management. He regularly instructs courses (Police, Corrections and D.N.R. academies) for the local community college and other law enforcement agencies. In addition, he frequently provides consulting and training services to the U.S. Department of Justice, National Institute of Corrections, in the areas of Field Training, Officer Program Development, and Training Program Development for agency training directors and coordinators.

Christine Oskison
Ms. Oskison is an Associate with Lamberth Consulting and provides data management and data analysis services for the organization. Ms. Oskison performs cross-tabulations and reports using data management software such as SPSS and Excel. Her data management

services are supervised by Dr. John Lamberth. Ms. Oskison holds an M.A. in Counseling from Eastern University, and a B.A. in Psychology from Dickinson College.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF MAPS

ACKNOWLEDGEMENTS

We would like to thank the Ann Arbor Police Department (AAPD) for their support and cooperation during the course of this study. From the beginning of the effort, we were able to call upon the resources of the AAPD for the components that are necessary to complete a study of this nature.

We worked closely with Deputy Chief Don Leach and Deputy Chief Greg Bazick. They provided us with information about police activity, special deployments, special circumstances within the City that influenced policing, and many other aspects of their work that would be necessary for us to understand when conducting this study. We thank them for their willingness to share their knowledge of this jurisdiction with us.

The successful identification of benchmark locations and of stop data that accurately reflects traffic in that location is essential to the successful completion of a study of racial profiling. The personnel of the AAPD who were assigned to this project worked and shared their insight and experience with us and helped to make the study run smoothly. Completing a project of this magnitude required significant cooperation from the AAPD, which we greatly appreciate.

We would also like to thank Chief Dan Oates and Susan Pollay for their review of the approach to conduct this study, and for their willingness and efforts to assist us in overcoming challenges associated with work of this nature.

EXECUTIVE SUMMARY

The past decade has seen increased awareness of the racial profiling issue among lawmakers, law enforcement agencies, and the communities in which they work. To address the issue, many agencies have begun collecting stop data, and analyzing the data to determine if racial profiling is occurring in their jurisdiction. Some collection and analysis efforts are due to threats of litigation or settlements. Others have been voluntary in nature, while still others have been legislatively mandated. One of the major issues in data analysis to date has been in determining the appropriate benchmark or standard to which the stop data are compared. The methodology employed in this study is one that has been employed in several studies across the country. This methodology employs what we believe to be the only appropriate benchmark for such an analysis: a measure of the driving population in the local area.

This study addressed the following 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?

Data on the transient population was collected at 9 locations throughout the city of Ann Arbor. These locations were selected due to the high number of stops at each, traffic patterns that were relatively representative of the jurisdiction, as well as accessibility for surveyors. Traffic surveys at each location were conducted by highly trained surveyors over a seven-week

period and on randomly selected days and times. These surveys provided the benchmark data to which stop data for that location was compared.

The results of this study are among the "best" that we have seen in our work around the country. They provide virtually no evidence for targeting of Blacks in Ann Arbor. At most locations the proportion of Black stops were very close to what one would expect based upon their presence in the transient population. In two locations, the proportion of Black stops was high enough to warrant a review of stops at that location by the department.

INTRODUCTION

Representatives from minority groups provide anecdotal evidence of racial profiling on the roadways spanning back decades, however, the specific measurement of the practice by law enforcement agencies was not formalized until 1994. During the criminal litigation case in New Jersey (*State v. Soto et al.*), a group of defendants alleged that New Jersey State troopers were targeting and stopping minorities on the highway, not because of their driving behavior, but because of the color of their skin. During the course of this case the race and ethnicity of the driving population was observed and recorded on portions of the New Jersey State Turnpike (Lamberth, 1994.) The driving population then was compared to the racial and ethnic make-up of the individuals stopped in New Jersey to determine whether a disproportionate percentage of minority drivers were being stopped relative to their presence on the roadway. This method was also used in Maryland (Lamberth, 1996), during the civil litigation case (*Wilkins v. Maryland State Police*) in which Robert Wilkins alleged that the rental car driven by his cousin on the Maryland State highway was stopped and searched by a drug-sniffing dog due to a "profile" prepared by the Maryland State Police which included Black males driving rental cars.

In the former case, the courts held for the defendants. The latter case was settled, and the issue of racial profiling began to develop greater national attention and exposure. It is important to note that the early work performed in this field, while groundbreaking, was limited due to the fact that it was conducted within the context of litigation. That is, the issue was reviewed in a confrontative forum between community and law enforcement participants. The work was completed slowly, and dialogue surrounding the science was combative. A dramatic shift resulting from state legislation and agency participation and leadership relative to this science

began to take place in the late 1990's. State legislatures have mandated data collection, and/or developed laws prohibiting racial profiling by law enforcement agencies. At the time of this report, 24 states have enacted legislation relative to this issue. An additional 10 states have legislation pending on the issue, and agencies in all but 2 states in the nation have undertaken data collection efforts due to mandate, decree, or of their own volition. Several significant events have occurred nationally which have influenced this shift in focus, and have helped to direct activities in this field.

In June 1999, the Department of Justice (DOJ) hosted a conference on "Strengthening Police-Community Relationships." The conference recognized that police are more effective when they have the trust and cooperation of the residents in their community. However, in many communities, especially minority communities, a lack of trust remains between law enforcement and local residents. This tension is exacerbated by allegations of police misconduct such as racial profiling.

The conference highlighted the need to identify proactive police practices to build trust, enhance police integrity and reduce police misconduct. Members at the conference determined that collecting data on traffic and pedestrian stops, analyzing this data, and providing the results for public review can help to shift debates on racial profiling from anecdotal reports to informed discussions. By being proactive about recognizing and addressing racial profiling, police communities can go a long way towards managing perceptions around racial profiling and strengthening police-community relationships.

In February 2000, the DOJ held a conference entitled "Traffic Stops and Data Collection: Analyzing and using the Data." In this session, more than 75 federal, state and local police

administrators, prosecutors, civil rights advocates, government officials as well as police labor leaders, researchers, and community leaders gathered to examine the collection, analysis, and use of data on traffic, pedestrian and other law enforcement stops. Collectively the participants reached several conclusions:

- Traffic stop data collection systems are needed to respond to the perceptions of racial profiling, to measure the reality, and to bridge the gap between minorities and police.
- Core data elements of traffic stop systems should include: date and time, location, race and ethnicity, gender, reasons for initiating the stop, actions taken by the officer, and duration of the encounter.
- Benchmarks for comparing data collected on stops are essential for conducting valid analyses. Without valid control groups, supportable statistical analyses are not possible.
- Data that is complete, accurate and truthful is critical.
- Analysis of data must be conducted by a capable and credible party.
- Publicizing traffic stop data can help to build trust between public law enforcement agencies and the public.

In August of 2001, the Police Executive Research Forum under a DOJ grant held a conference for leading researchers in the field to discuss issues relating to benchmarking for stop data collection and analysis. The conference was attended by social scientists, legal scholars and practitioners from several police departments. This conference was the first of its kind to bring leading scientists and researchers together to discuss the best methods for analyzing stop data.

In March of 2003, the SOROS Foundation provided support for a conference on racial profiling that was co-hosted by the Institute on Race and Justice at Northeastern University, the American Civil Liberties Union, the National Organization of Black Law Enforcement Executives, and Lamberth Consulting. The Conference, "Confronting Racial Profiling in the 21st Century: Implications for Racial Justice," featured 30 of the leading researchers in the country. The intent of the conference was to bring together researchers, law enforcement representatives and community representatives to collectively review the latest and most progressive methods for stop data collection and analysis. The conference also focused on post-stop activity, community engagement, and data auditing as primary subject topics.

In November, 2003, the Northwestern University Center for Public Safety and the Police Executive Research Forum held the Third National Symposium on Racial Profiling. The third day of that conference was given over to discussing issues of data collection and analysis. Specifically issues of risk management, benchmarking, post-stop activity, and related topics were discussed. Observational benchmarks, which were pioneered by Lamberth Consulting and were utilized in the Ann Arbor study, were cited as the most used and reliable of the strong benchmarks discussed.

From these conferences, a central and critical focus has become clear. To manage public perception about racial profiling and to strengthen community-policing relationships, the method used for collecting and analyzing stop data is critical. Two primary components must be in place to determine whether racial profiling is occurring: benchmarks and complete stop data.

The Right Benchmarks

"Benchmark data" refers to control data against which stop data can be compared to determine if any racial or ethnic group is being stopped at a disproportionate rate. The right benchmark can provide the racial and ethnic demographic for any given locality, whether it be an urban intersection or a state highway. Stop data can then be compared to the demographic, and a statistical analysis can be conducted which will help determine if some racial groups are being stopped more frequently than their demographic presence, which may indicate that profiling is occurring.

We believe that collecting the right benchmark, or understanding the true demographic of a locality, is essential to procuring valid results on profiling. If the assumed demographic is suspect, then the comparison to stop data may yield invalid results.

Today, the most experienced researchers in this field generally agree that the best method to measure roadway traffic is observational surveys, and many researchers have used observational surveys to attempt to validate other benchmark methods[1]. This means that the racial and ethnic mix of individuals traveling through a locality must be identified and recorded. A schedule must be developed to survey carefully chosen locations according to a randomly selected time schedule. If the right locations are surveyed according to the right schedule, then the demographic for a given locality may be assumed.

[1] Geoffrey P. Alpert, Michael R. Smith & Roger G. Dunham, "Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research." Paper presented at *Confronting Racial Profiling in the 21st Century: Implications for Racial Justice.* Boston, March, 2003. Amy Farrell, Jack McDevitt, Shea Cronin & Erica Pierce, "Developing a modified census benchmark for traffic stop data in Rhode Island." Jeff Rojek, Richard Rosenfeld and Scott Decker. "The influence of driver's race on traffic stops in Missouri." Paper presented at *Confronting Racial Profiling in the 21st Century: Implications for Racial Justice.* Boston, March, 2003.

Other benchmarks, such as census data on population demographics, have proven not to serve as reliable benchmarks. Census data measures static populations such as the demographic of households. Highway and pedestrian traffic represent transient populations. People work in different locations from which they live, and travel in different routes and different ways to get there. Additionally, tourism, business trips, and other populations not measured in census data, such as university populations, make the comparison suspect. For example, in *New Jersey v. Soto* (1996) and *Wilkins v. Maryland State Police* (1996), it was found that census data did not accurately predict highway transient traffic. For these reasons, we used direct observations of transient populations in this study.

Complete Stop Data

The second set of critical data is the police stop data. For the purposes of this report, we make a distinction between stop data, and ticket data. Stop data refer to all police stops (traffic or pedestrian) that do not result in the subject of the stop receiving a ticket. Ticket data refer to police stops that result in the subject of the stop receiving a ticket.

Ticket data may be compared to benchmark data to determine if racial profiling is occurring. However, the majority of police stops are not ticketed. For example, approximately 63% of all police stops in New Jersey (*New Jersey v. Soto*) were not ticketed, and approximately 75% of all stops in Arizona (*Arizona v. Folkes*) were not ticketed. Analyzing these data are important, perhaps more so than ticket data alone, and should be performed if at all possible.

The content of the stop and ticket data is equally important. In addition to race and ethnicity, the time of the stop and specific location are crucial so that valid comparisons against transient demographics can be conducted. On highways, this means that mile marker and traffic

direction must be known to conduct valid comparisons. In urban areas, street name and nearest cross streets, or equally specific location data, must be known to conduct valid comparisons. Generalizations are not enough. Transient populations vary according to time of day and specific location. For example, the transient population in an urban area may differ significantly from one street corner to the next, depending upon the businesses, homes and university locations, and the time of day.

Data Analysis Considerations

We should note that the question of how to perform data analysis is not simple, nor have all researchers historically agreed on the best methods to conduct the analysis. This makes sense given the relative youth of this discipline, and the burgeoning nature of the issue. As mentioned previously, most researchers today agree that the best method for determining transient populations is observational surveys. We feel it is important, however, to discuss some points of current interest and review in the academic community relative to conducting this type of analysis.

Violators

One question facing those attempting to analyze traffic stop data involves the selection of the most appropriate benchmark to use for comparison. A number of measures have been used in the research to date and an open question remains as to whether using estimates of the population violating traffic laws is an improvement over estimates of drivers operating on a community's roadways. Courts (beginning with the *Soto* and *Wilkins* decisions) have said violators, but then quickly changed their focus when it became obvious that the two were virtually synonymous.

Court decisions uniformly support the notion that any motorist violating a traffic law is subject to being stopped by police and are the appropriate group to benchmark. However, to date, empirical evidence supports the contention that traffic and violators are synonymous, and in *Soto* the Court essentially used traffic and violators interchangeably.

The first scientific measurement of the appropriate comparison number for traffic stops determined both the proportion of Black motorists in the traffic stream, and those violating at least one traffic law (*New Jersey v. Soto, et al.*). The evidence in that case subsequently has determined that the two are virtually synonymous. First, in *Soto* and in *Wilkins v. Maryland State Police,* virtually every motorist was speeding (98.3% in *Soto* and 93.3% in *Wilkins*). More recently, Lamberth (2003)[2] reported a study in which police officers were given 5 minutes to determine whether randomly selected cars were violating some traffic law. The study concluded that fully 94% of the drivers were violating some law and it took a mean of 28 seconds for the officers to spot the violation.

For the reasons stated above, we have used the traffic estimates as our benchmarks in Ann Arbor. However, we should note that direct research measuring differences between racial or ethnic groups and driving behavior is very limited. While empirical evidence suggests that traffic violators and traffic motorists are virtually identical, a question remains as to whether one racial or ethnic group is more likely to violate egregiously than another. That is, it is theoretically possible, while perhaps not intuitive, that one racial or ethnic group is more likely to speed excessively, or drive vehicles with severe vehicle codes violations, or run traffic lights more often, etc. To date, empirical evidence is scant on these issues and mixed. We feel that

[2] Lamberth, John, "Measuring the racial/ethnic make up of traffic: The how, what and why." Paper presented at *Confronting Racial Profiling in the 21st Century: Implications for Racial Justice.* Boston, March, 2003.

one important area of future research in this field is a focused review of driving behavior among different racial and ethnic groups.

Comparing Transient to Demographic Data

As we have said previously, transient and demographic data are often quite different. The racial composition of this transient population may or may not mirror the population of the city or county. For example, Washtenaw County has a population that is 11.4% Black of those 18 and above. Ann Arbor has a Black population of 8.0% of those 18 and over. If we used the county percentages as the benchmark to which to compare the stops of the Ann Arbor Police, then we would underestimate the percentage of Blacks in the driving population at 4 locations and overestimate it at 5. If we used Ann Arbor City percentages, we would underestimate the Blacks in the driving population at 7 of the 9 locations, overestimate it at 1 and be correct at 1. For example, the transient population in Ann Arbor ranged from a low of 7.7% Black motorists at Miller and Newport (which is 3.8% lower than the census data in Ann Arbor), to a high of 12.9% Black motorists at S. University and State (which is 61.3% higher than the same census data).

However, the story is even more complex than this. If we compare census data to transient data for other cities, the differences are even more disparate. In Kansas City, Kansas, the Black population is 27.9% Black of those 18 and over and 14.7% Hispanic. There were 8 locations in that city that were benchmarked for traffic purposes. The Black driving population ranged from 8.9% at the lowest measured location to 86.9% at the highest measured location. The Hispanic traffic ranged from 2.3% at the lowest measured location to 40.7% at the highest. Clearly using census data for that city would have overestimated the Black and Hispanic traffic

at some locations and underestimated it at others. There are smaller geographic census enumerations of population to which our benchmark locations can be compared, i.e., census tracts, which average 4,000 residents. These can also be compared to the locations that we benchmarked. For illustrative purposes, we provide both traffic demographics and census demographics for each benchmark location in Kansas City and Ann Arbor.

Table 1: Kansas City - Black and Latino Benchmark vs. Census Tract[3]

Location	Benchmark % Black Traffic	Black Census	Comp. Disparity	Benchmark % Latino Traffic	Latino Census	Comp. Disparity
13th & Quindaro	87.6	90.2	+ 2.9%	2.3	5.9	+61.0%
18th & Parallel	84.9	81.9	- 3.7%	2.5	6.5	+61.5%
38th & State	51.6	41.6	-24.0%	9.2	21.0	+56.2%
59th & Leavenworth	38.7	29.0	-33.4%	3.0	4.9	+38.8%
78th & State	30.2	39.0	-22.6%	5.4	7.8	+30.8%
Metropolitan & Woodland	11.2	17.7	+36.7%	29.6	39.4	+24.9%
10th & Kansas	9.0	1.2	-650.0%	40.1	50.6	+20.8%
43rd & Rainbow	10.0	12.5	+20.0%	8.1	21.7	+62.7%

Census data overestimates traffic in 10 of the 16 comparisons and underestimates it in 6.

Table 2: Ann Arbor - Black and Latino Benchmark vs. Census Tract[3]

Location	Benchmark % Black Traffic	Black Census	Comp. Disparity	Benchmark % Latino Traffic	Latino Census	Comp. Disparity
S. University & State	12.9	9.2	-40.2	0.4	3.6	+88.9
S. University & Washtenaw	11.6	2.6	-346.1	0.5	3.7	+86.5
Fourth & Huron	8.5	6.8*	-25.0	0.5	4.1*	+87.8
Hubbard & Huron Parkway	10.7	12.1	+11.6	0.6	4.4	+86.4

[3] Note: The comparative disparity is arrived at by subtracting the traffic percentage from the census percentage and dividing by the census percentage.

Stadium & Washtenaw	11.7	4.5	-160.0	0.6	2.1*	+71.4
Stadium & Main	10.1	6.0	-68.3	1.6	3.6*	+55.6
Eisenhower & State	12.3	10.2	-20.6	1.2	4.3	+72.1
Miller & Newport	7.7	8.9	+13.5	1.0	2.3*	+56.5
Seventh & Pauline	8.0	5.6	-42.9	1.3	3.3*	+60.6

* Indicates adjustment to the total percentage on the basis of jurisdiction wide proportion of minorities 16 and up, as census tracts with less than 100 minorities are not reported by age and race/ethnicity. Each census total has 1 or more census tracts that are not reported by age and race/ethnicity.

Census overestimates the traffic at 11 locations and underestimates it at 7.

The comparisons are of the intersection benchmarked and census demographics for that census tract, or, if the benchmark location abut more than one census tract, the average population for those tracts. Suffice it to say that there are large discrepancies between the traffic and residents of a large majority of the 34 comparisons of traffic and census. Eleven of these underestimate the minority population in the traffic and 23 overestimate it. The fact that these differences are not consistently an over, or under representation of minority motorists in the transient population have so far made it impossible to develop an algorithm to accurately utilize census data to estimate traffic data. This discrepancy between the transient population and census data and among different locations in the city is fundamental to understanding racial profiling and assessing whether it is occurring or not. It is this precision of measurement, accurately identifying the "transient" population at specific locations, that the methodology used in the present study allows.

METHODOLOGY: OVERVIEW

Stop Data

The AAPD started collecting stop data at the direction of an Ann Arbor City Council Resolution, passed in February of 2000. However, as there were no agreed upon benchmarks and no agreement on how to analyze those data, they wisely did not publish them. When Lamberth Consulting began working with the AAPD, we were told that data could be delineated to street and cross street by connecting the stop data being collected for the study to the Computer Assisted Dispatch (CAD) locations. It was discovered in late 2001 that the unknown rate from this matching process was 27% and then 11.67%. Lamberth Consulting, in consultation with the AAPD, deemed that unacceptable. In a meeting with city and department officials in early 2002, the decision was made to add a CAD identifier to the stop data that were being collected. Data were then collected between April and October, 2002. When AAPD forwarded these data to Lamberth Consulting, it was discovered that there appeared to be between 18% and 25% unreported traffic stops. After inspecting these data, it appeared that there were at least four possible explanations. One was that officers were not turning in all of their stops, but even when additional stop forms were added to the database, there still appeared to be unreported traffic stops. Another explanation was that there was a technical problem with the input of data to the CAD system and that officers were inadvertently entering the same stop more than once. This was a function of the way data are entered into database tables in the CAD system when an officer enters activity from a mobile computer versus when the same information is entered by a dispatcher from a networked computer in the dispatch center. Additionally it was determined that the Scantron machine was not reading significant numbers of Scantron reports. Finally, faulty query syntax when analyzing data was discovered that led to a

further misrepresentation of stops at various locations. As it was impossible to determine to what extent these problems were occurring, it was decided to proceed with further data collection, a review of procedures put in place by the department. Additionally, an analysis of one month of the new data was analyzed. Data from January, 2003 through September, 2003 was provided for data analysis purposes. After careful analysis of the new dataset and additions to the officer and supervisors protocol concerning recording of stops, these data were determined to be sufficient both in quantity (at seven of the benchmarked locations) and quality for the data analysis needed for this study.

Whereas the earlier databases had seemingly unreported traffic stops of between 18 and 25 percent, the new dataset has an apparent unreported rate of 3.2% or less. This is under even the most stringent standards that have been established in both litigation and reviewing the best department's reporting percentages.

The AAPD instituted additional procedures for assuring officer compliance with reporting requirements. In January of 2003 the AAPD developed a new process to ensure 100% compliance of officers submitting completed Scantron forms for each traffic stop that they conducted. The process, which is still being utilized, consists of the following seven steps:

1. Each traffic stop is captured on the AAPD CAD system.
2. Officers submit a Scantron form for each traffic stop that they make at the end of their shifts to their individual shift command officers.
3. The following day the command officers run the CAD report listing all of the traffic stops that the individual officers on their shift conducted the previous day.

4. The command officers then match the submitted Scrantron forms to the CAD report.

5. The command officer certifies that there is a Scantron form for each listed traffic stop on the CAD report by signing the CAD report.

6. The CAD report cannot be submitted until all Scrantron forms have been submitted for a shift. The Scantron forms are physically attached to the CAD report. The Scantron forms and the CAD report are submitted to data processing.

7. A data processing supervisor monitors submission of the forms to ensure that a CAD report is submitted for each working shift of police officers.

Lamberth Consulting provided the "Collecting Stop Data Training: Training for Trainers" course which occurred during four training sessions conducted on September 6, 7 and October 2 of 2001. One hundred twenty eight officers attended, which at the time of the training equaled approximately 70% of the AAPD staff. The course was designed to accomplish two tasks: to provide training to as many of the current AAPD staff as possible and to assist the agency in developing the internal capacity to deliver the training to new and existing employees by providing staff instructor's the knowledge and skills necessary to deliver the course in the future. The course covered the following topic areas:

Racial Profiling Overview - Module Objectives:

- Provide definitions of "racial profiling" and "behavioral profiling."
- Create an awareness of at least 3 impacts that racial profiling may have on law enforcement and the community.

- Create an awareness of at least 3 risk factors for law enforcement agencies that participate in racial profiling when on patrol or working in the field.
- Discuss why racial profiling is an ineffective method for identifying and deterring criminal activity.
- Define discretion, police officers' use of discretion and the negative and positive impacts of the application of discretion.

Data Collection Programs - Module Objectives:

- Provide definitions of the following terms: stop data, stop data collection programs, benchmark data, and benchmark surveys.
- List at least 4 critical elements of a comprehensive data collection program.
- Describe the officer's critical role within the stop data collection program.
- Discuss the completion of the Electronic Data Sheets (EDS) to collect stop data.

Benchmarking

Benchmarking locations were chosen in Ann Arbor after an initial conference with representatives of the AAPD. This selection occurred on July 25, 2001 and was done by the Project Director and the Project Manager for Lamberth Consulting. Every possible benchmark location was inspected and information relating to the location was discussed, which included criteria such as:

- Traffic patterns (nearby jurisdictions, organizations, entertainment, etc. that might influence or impact traffic driving patterns),

- Traffic density (the number of cars traveling in each direction within a specified timeframe),
- Sight lines for surveyors (surveyor positioning, distance to traffic, and any obstacles that might impede sight),
- Lighting or lack thereof (required for night time surveying),
- Surveyor safety,
- Police activity, and
- Type of vehicles stopped by police.

Surveyor Training

Teams of surveyors were hired and trained to visually identify and manually record the race and ethnicity of individuals who comprise the transient populations. The training session was held on October 9, 2001 at the AAPD and several locations in the city. Survey training is critical to ensure that surveyors understand the surveying process, surveyor positioning, daytime and nighttime surveying guidelines, data recording procedures, quality assurance reviews such as the assessment of inter-rater reliability, and the data cataloguing steps required for this work. During this session, survey team leaders also were trained on survey management tasks such as status reporting, interacting with police department personnel, and supervising surveyors. The survey training consisted of:

- A high-level overview of the purpose of the Ann Arbor study. The intent of this portion of the training was to provide surveyors with a basic understanding of the importance of the study and the critical role that they would play in the study.

- An explanation of the survey method, schedule, and roles were discussed. The survey procedures were diagrammed and reviewed. The intent of this portion of the training was to provide surveyors with a basic understanding of how the survey would be conducted.

- Hands-on practice in the field in which surveyors practiced on-location, using the actual data sheets developed for the survey. During this portion of the training, guidance was provided on data capture, review, and feedback to surveyors on the methods and tips for positioning, and data recording. Surveyor data sheets were reviewed, and feedback was provided on performance. The intent of this portion of the training was to provide surveyors a chance to practice in a "consequence-free" environment before conducting the actual survey. Inter-rater reliability coefficients were computed to ensure that surveyors were trained to criterion.[4]

Observation Survey Types

Stationary surveys were conducted for this project, during which surveyors stood at street corners to record the race or ethnicity of individuals. Drivers' race or ethnicity was categorized as Asian, Black, Hispanic, White, Middle Eastern, Other, or Unknown. The surveyors recorded populations at predetermined times and predetermined locations.

Each survey team was comprised of two individuals, one team leader and one surveyor. The team leader was responsible for supervising the team, keeping track of survey times, interacting with police liaisons, and organizing and collecting the data sheets. The team leader also acted as a surveyor. Each surveyor was responsible for capturing data for traffic moving in

[4] A minimum inter-rater reliability coefficient (i.e., the percent of agreement between two surveyors observing the same car at the same time) of .80 was used as this criterion. This is a commonly accepted standard in social science research.

one direction (North, South, East, or West). Surveyors captured data for one lane at a time and alternated lanes.

Benchmarking quality assurance activities were conducted by the project manager and the survey team leads throughout the duration of the surveys. Quality assurance was conducted to ensure that surveying was conducted on schedule and surveying was conducted properly. Measures were also taken to gauge surveyor reliability, or extent to which surveyors perceptions of race were consistent. These activities consisted of:

- Conducting inter-rater reliability tests to measure the extent to which surveyors perceived race the same. These tests were conducted by both survey teams, and were conducted at several locations. The quality assurance manager administered and reviewed the first series of tests, while survey team leads conducted and reviewed later tests.
- Contacting police liaisons from each department to provide them with the survey schedule, and answer any questions they might have about the benchmarking activities.
- Conducting pre-survey reviews for each location to determine positioning, scheduling, materials and preparation reviews, and contingency planning.
- Conducting on-going status meetings to review survey progress, discuss issues, and review surveyor performance.
- Conducting post-survey reviews to ensure timing, survey scheduling and review data cataloguing and data entry schedules.
- Spot-checking survey teams during the survey periods to observe team location and survey timing during the process. During this time, team leads provided status and feedback about the survey sessions.

- Conducting periodic reviews of data captured to ensure that the data sheets were properly catalogued and filed.

- Conducting data entry reviews to ensure that data entered matched the data recorded.

Establishing the Context

Following data collection and preliminary analysis, it is necessary to determine enforcement contexts that may explain any disparities that are noted. This endeavor was carried out on January 5, 2004 in a conference call between representatives of the City, the AAPD and Lamberth Consulting. Preliminary results were shared with the department and they were asked to consider the context of their department during the time that data collection was occurring. Were there special circumstances that would help us understand any disparities that occurred, or were there any general conditions in their area among the motoring public that would allow us to understand any disparities? This exercise is essential if we are to reach a conclusion about the meaning of disparities between benchmark percentages of minority motorists and stop percentages of minority motorists. It is important to evaluate these disparities in the context of effective and appropriate policing. For example, while stopping of minority motorists for stereotypic reasons is considered to be racial profiling, there are circumstances when a specific minority can be stopped legally by a police department.

Indeed, there are times when minority motorists fitting the description of a known perpetrator of a specific crime should be stopped at a higher rate than their presence in the motoring public would suggest. That is, if there is a crime in which the suspects are identified as a minority, police will appropriately pay closer attention to that minority during the time that those suspects are being sought. This probably will result in higher stop rates for that minority during that time period. These data will need to be evaluated with that context in mind.

Representatives of the AAPD informed us that there were no special circumstances that would affect the results of the survey at the seven locations benchmarked.

RESULTS

Overall surveyors categorized 20,391 car drivers for race/ethnicity. Of these 19,560 or 95.9% were race/ethnicity identified. This is a high rate of racial identifications, particularly taking into consideration that some of the surveying was completed during dusk and night hours. This rate, in part, may be attributed to the excellent lighting present in Ann Arbor, which significantly aided nighttime surveying.

With regard to gender, 20,391 motorists were categorized for gender with 19,651 (96.4%) being gender identified. Again, these rates are high and more than meet scientific standards for surveying. The classification percentages serve as an internal validity check on the surveyors and indicate that they were being diligent and observing the survey protocol.

Inter-rater Reliability

One of the scientific standards for assuring that different raters are making the same determinations with regard to race is a technique called inter-rater reliability (IRR). This involves two surveyors determining the race of drivers of the exact same cars. Several inter-rater reliability tests were run. The first was conducted on October 18, 2001 at approximately 9:30 a.m. The location was Hubbard & Huron. The race IRR was .88 and the IRR for gender was .93. The second was conducted on October 18, 2001 at approximately 11:00 a.m. The location was S. University & Washtenaw, specifically targeting southbound Washtenaw traffic. The result for race was an IRR of .89 and for gender the IRR was .93. The next event occurred on October 26, 2001 at approximately 6:45pm. The location was Seventh & Pauline, specifically targeting eastbound Pauline traffic. The IRR for race was .86 and for gender was .92. The next event occurred on October 27, 2001 at approximately 7:00pm. The location was Eisenhower &

State, specifically targeting northbound State Street. The IRR was .79 for race and .88 for gender. The final event occurred on October 27, 2001 at approximately 7:30p. The location was S. University and Washtenaw, specifically targeting southbound S. University. The IRR for race was .85 and 100% for gender.

For all inter-rater reliability studies of race, in daylight, inter-rater reliability was .88. In dusk or dark conditions, the inter-rater reliability was .84. There is a consistency in these inter-rater reliability tests that has been commented on before (*New Jersey v. Soto*) and also makes common sense. Under poorer lighting conditions, it is more difficult to determine the race/ethnicity of motorists. These measurement errors are a normal part of scientific observation and are considered when statistical analyses are computed.

Analyses

The major analysis that we will report is the odds-ratio of being stopped if the motorist is Black versus if they are not Black. Exact equality in that analysis is when the odds ratio is 1. The odds ratio is best understood by filling in the ratio in the following sentence: If you are Black you are ___ times as likely to be stopped as if you are not Black. If a perfect world of no racial profiling all of the ratios would be 1. This would mean that Blacks are no more likely to be stopped that non-minorities. More realistically, we would expect some of the ratios over 1 and some less than 1. However, we know that there are errors of measurement in the benchmarks and errors of measurement in the stop data. Therefore, we have taken the position that odds ratios between 1 and 1.5 are benign, and that odds ratios of 1.5 to 2.0 suggest that in the absence of other explanations, targeting of Blacks may be occurring. Benchmarks of over 2 should be seriously considered by the AAPD.

Each of the comparisons between the benchmark percentage of Blacks and the stop percentage of Blacks was analyzed using the chi-square (χ^2) analysis and these analyses are presented in Appendix A. This analysis determines whether the observed differences are real or the result of chance. By convention, statisticians use the .05 level of probability to determine statistical significance. That is, if the observed result would occur 5 or less times out of 100, then it is treated as a real result, not a chance finding. As probabilities decrease, we become more confident that the result is real, so probabilities are reported as significant if they are .05 or less.

Lamberth Consulting in consultation with the AAPD identified several possible benchmark locations in the city on the basis of police activity. Each of these locations was considered and 9 were selected on the basis of geographic considerations, road construction and surveyor accessibility. The locations were spread throughout the city, encompassing all four police patrol districts and represent police activity in all areas of the city. The locations were:

- S. University & State
- S. University & Washtenaw
- Fourth & Huron
- Hubbard & Huron Parkway
- Stadium & Washtenaw
- Stadium & Main
- Eisenhower & State
- Miller & Newport

- Seventh & Pauline

When the stop data were provided by the AAPD, the Seventh & Pauline and Miller & Newport locations did not have enough stops for meaningful analysis (39 and 70 respectively). For completeness, we will provide benchmark data for those two locations in Table 3, but will not provide any other comparisons due to the low number of stops at those locations.

The AAPD had an existing program of data collection that has been ongoing since July of 2000**.** The ongoing data collection program merely identifies stops as being within a specific geographic area that is too large to be useful in locating the stops within an appropriate perimeter for the 9 benchmark locations. Perimeters were drawn around the benchmark location in consultation with the AAPD.

Race

Two of the locations, Seventh & Pauline and Miller & Newport did not have enough stops in the database to allow reliable measurement. We simply report the benchmarks for these locations.

Table 3 presents the data for race of drivers benchmarked and race of drivers stopped by the AAPD at the 7 locations benchmarked.

Table 3: Race Analysis

Location	Bench Mark N	Bench Mark % Black	Stop N	Stop % Black	Diff	Odds Ratio
S. University & State	1647	12.9%	333	15.6	2.7	1.2
S. University & Washtenaw	1715	11.6%	1389	16.8	5.2	1.5
Fourth & Huron	2417	8.5%	1341	11.0	2.5	1.3
Hubbard & Huron Parkway	1402	10.7%	460	16.1	5.4	1.6
Stadium & Washtenaw	2569	11.7%	1891	18.7	7.0	1.7
Stadium & Main	3263	10.1%	1697	13.7	3.6	1.4
Eisenhower & State	3566	12.3%	1262	13.9	1.6	1.2
Miller & Newport	1684	7.9%	N/A	N/A	N/A	N/A
Seventh & Pauline	1169	8.0%	N/A	N/A	N/A	N/A

The first column in Table 3 refers to the location of the stops. The second column refers to the number of motorists (N) recorded in the benchmark. The next column refers to the percentage Blacks in benchmark data. The next column refers to the number (N) of stops in the existing stop data. The next refers to the percentage of Black stops. The next refers to the percent difference, and the final column refers to the odds ratio of being stopped if you are Black.

In these data, we see that 5 of the 7 odds ratios are in the benign range. Two are above 1.5 and those only moderately so. If these data are collapsed and an odds ratio is computed (considering that 1295 [15.5%] Black motorists were stopped and, based upon the benchmark data, one would expect that 914 [10.9%] Black motorists would be stopped), the overall odds ratio is 1.5, falling in the benign area. This indicates that there is no profiling in the stops of Black motorists going on overall in the benchmark areas selected for study. There are, of course

two locations, Hubbard & Huron Parkway and Stadium & Washtenaw that are above the 1.5 benign ranges, and the AAPD might wish to consider these areas further.

While we recommend that the AAPD remain vigilant about all areas of the City, we do not want this recommendation to be misconstrued. Overall, the results of the stop data in Ann Arbor are among the best that we have seen to date in the jurisdictions where we have assessed racial profiling. To give some perspective to the situation, there have been overall odds ratios as high as 4.85 in other jurisdictions and several that we have seen in the mid 3 range. To have an overall odds ratio of 1.5 means that overall there is no evidence of profiling. That there will be variation about that 1.5 average is inevitable, and we will continue to recommend that the areas that are higher than expected be scrutinized.

Ethnicity

The AAPD makes very few stops of Hispanics (3% of their stops at the benchmark locations) and Middle Easterners (< 5% of the stops at benchmark locations). The stops of Hispanics at the 7 locations benchmarked ranged from 8 at Hubbard & Huron to 74 at Stadium & Main. The stops of Middle Easterners ranged from 0 to 78. With this small number of Hispanic and Middle Easterner stops, any analysis that we would provide would be highly unreliable. Under these circumstances, providing an analysis would be inappropriate.

CONCLUSIONS

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. While there is no evidence overall that the AAPD is targeting Black motorists for stops, two of the odds ratios, at Hubbard & Huron Parkway and Stadium & Washtenaw are slightly higher than 1.5. These slight elevations should be monitored by the department to determine whether these locations are simply idiosyncratic in these data, whether they are locations that continue to show high activity and odds ratios, or both. There is inevitably variation around the average of stops for Black motorists for the department as a whole and some of the elevation at these two locations may be the result of chance.

RECOMMENDATIONS

Diligent management of racial profiling issues includes continued measurement and maintenance of programs. We have provided the following list of recommendations for the AAPD to consider as they continue to address this important issue.

1. We recommend that the AAPD continue to monitor those two locations where the Odds Ratios were slightly high, i.e., at Hubbard & Huron Parkway and Stadium & Washtenaw.

2. We recommend that the AAPD continue their racial profiling data collection program until they have enough stops of Hispanics and Middle Easterners to analyze adequately.

3. The science of analyzing racial profiling activity has advanced since the inception of this study to include an emphasis on post-stop activity. While out of scope for this study, the AAPD should consider the collection and data analysis of post-stop activity.

4. We recommend that the department continue the procedures instituted in January of 2003 to assure recording of racial profiling stop data. This allows for a more thorough audit of police stops.

5. We recommend that the AAPD consider providing officers training that is specifically designed to target racial profiling, including information that informs officers about when they can and cannot use race, behavioral profiling, what the hit rates for searching Black and Hispanic motorists are and other subjects that are specifically targeted to helping officers deal with the proper and improper use of race in policing.

6. We recommend that the AAPD continue to work with community groups to achieve a better understanding of the needs of the department by the community and of the needs of the community by the department.

APPENDIX A – CHI-SQUARES

Whereas odds ratios provide a good analysis of the probabilities of being stopped for each racial/ethnic group, the chi-square analysis takes into consideration sample sizes (number of stops of each group) to determine the likelihood of observed differences due to chance. By convention, statisticians use the .05 level of probability to determine the statistical significance of an analysis. That is, if the observed result would occur five or fewer times out of a hundred, then it is treated as a real result, not a chance finding. As probabilities decrease, we become more confident that the result is real, so probabilities normally are reported as statistically significant if they are .05 or less.

Unlike odds ratios, the chi-square statistic is sensitive to sample size. When conducting chi-square analyses on large samples, as in this case, small observed differences might reach statistical significance due to the size of the sample. That is, differences might be statistically significant but not meaningful. Thus, it is important to consider the results of the chi-square analyses and odds ratios together to consider whether statistically significant differences are in fact meaningful differences.

Table A.1 provides the χ^2 and probabilities for Black motorists for each of the 9 benchmark locations. When there are small numbers of one group or another, as sometimes occurs in these data, the chi-square is not an appropriate test. Fisher's Exact Test (FET) is used under those circumstances. It does not give a chi-square value and thus is listed in the tables as FET with the associated probability.

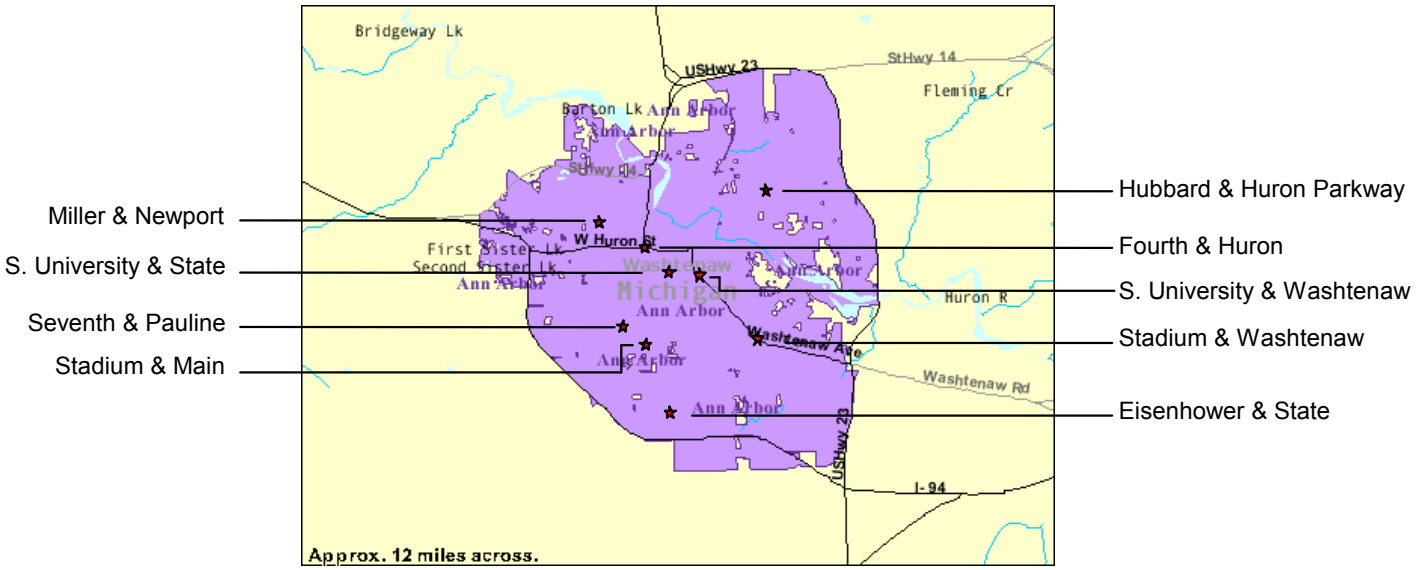
Table A.1: χ2 Analysis

Location	Chi-square	Probability
S. University & State	1.8	NS
S. University & Washtenaw	17.1	<.001
Fourth & Huron	6.6	<.03
Hubbard & Huron Parkway	9.5	<.01
Stadium & Washtenaw	42.0	<.001
Stadium & Main	14.5	<.001
Eisenhower & State	2.2	NS
Miller & Newport	N/A	N/A
Seventh & Pauline	N/A	N/A

APPENDIX B – MAPS

The following maps have been provided to identify the benchmark locations used in Ann Arbor to conduct the study. These maps are provided for reference purposes only.

Map 1: City of Ann Arbor

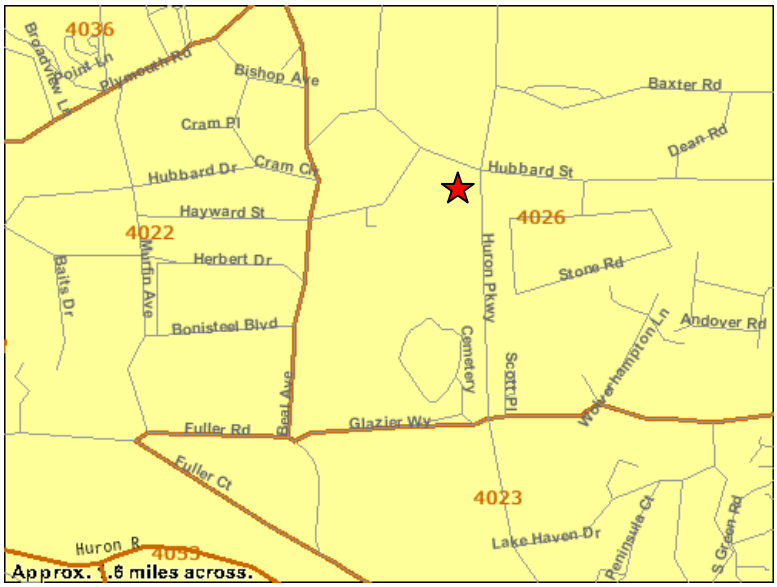


Map 2: S. University & State

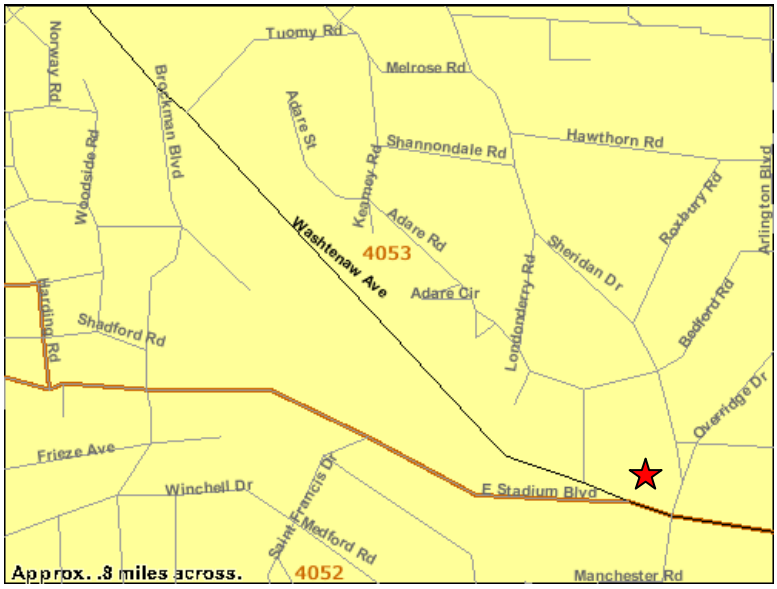
Map 3: S. University & Washtenaw

Map 4: Fourth & Huron

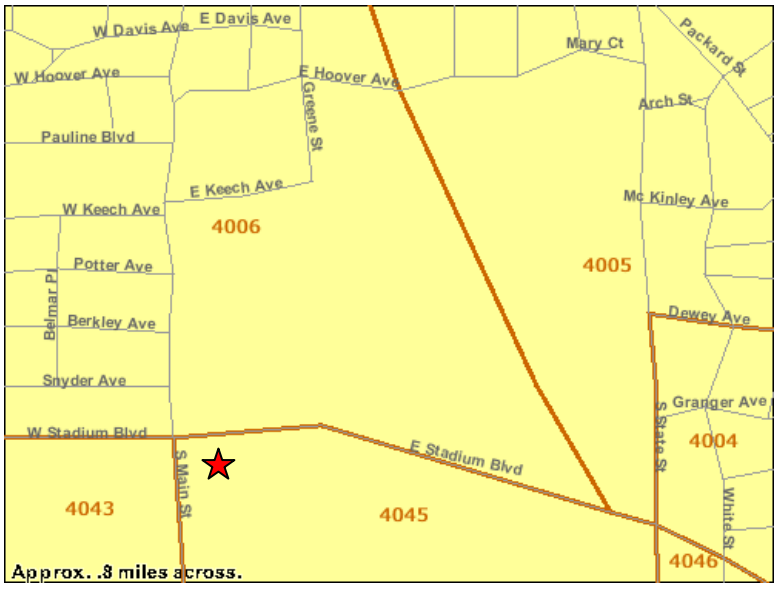
Map 5: Hubbard & Huron Parkway



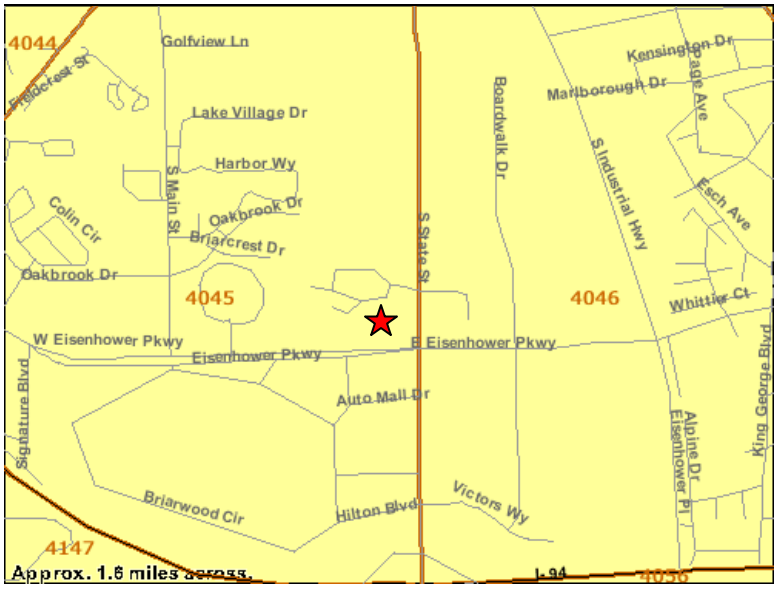
Map 6: Stadium & Washtenaw



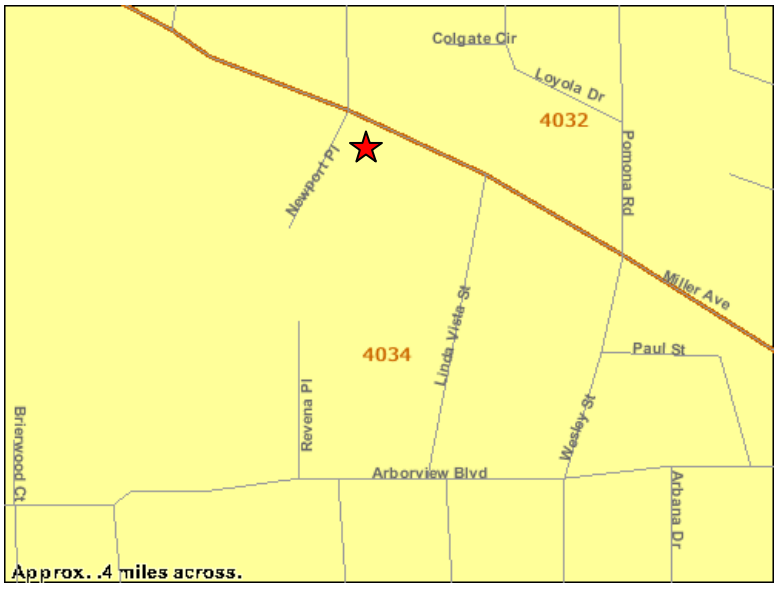
Map 7: Stadium & Main



Map 8: Eisenhower & State



Map 9: Miller & Newport



Map 10: Seventh & Pauline