Reporting Accurate Traffic Stop Data: Evidence-Based Best Practices

Missouri Sheriffs' Association

Summer Training Conference 2020

Richard R. Johnson, Ph.D.
Dolan Consulting Group
2840 Plaza Place, Suite 325, Raleigh, NC 27612
(919) 805-3020, info@dolanconsultinggroup.com

Richard R. Johnson, Ph.D.

Dr. Johnson is the Chief Academic Officer and director of research with the Dolan Consulting Group in Raleigh, North Carolina. Dr. Johnson retired from the University of Toledo as a full professor within the Department of Criminal Justice. As a professor, he taught a variety of criminal justice courses, including research methods courses, at the undergraduate and graduate level for 10 years. He has conducted extensive research on such topics as racial profiling, police use of force, the psychology of police-citizen interactions, effective responses to domestic violence, and police field supervision. He has authored 52 articles in peer-reviewed scientific journals and authored more than 100 articles in non-peer-reviewed magazines and reports.

He earned his bachelor’s degree in both criminal justice and public administration from the School of Public and Environmental Affairs (SPEA) at Indiana University. He earned his master’s degree in criminology at Indiana State University. He earned his doctorate in criminal justice, with a concentration on policing, from the University of Cincinnati. While at the University of Cincinnati, Dr. Johnson worked as a research assistant and project manager on racial profiling studies for the Pennsylvania State Police, Ohio State Highway Patrol, Cincinnati Police Department, Cleveland Police Department, Arizona Department of Public Safety, and the Nebraska State Patrol. Since earning his doctorate, he has privately assisted several law enforcement agencies with their biased-based studies and reports.

Before becoming an academic and researcher, Dr. Johnson served as a trooper with the Indiana State Police, and as a criminal investigator with the Kane County State’s Attorney Office in Illinois. He is also a proud military veteran having served as an infantry soldier and field medic in the U.S. Army and Army National Guard. He also served as a military police officer with the Air National Guard, including serving on active duty again during the war on terrorism.

What is Racial Profiling?

For decades the American public has been concerned about the practice that has been called "racial profiling." Racial profiling involves making judgements about one's moral character based on one's outward physical appearance, and then taking action against that individual based on those judgements. Regarding vehicle stops by the police, racial profiling refers to the practice of stopping drivers based on their demographic characteristics (such as their race, ethnicity, sex, or age), more than for law-violating behavior.

Based on a definition most Americans would support, here is a simple example of racial profiling in a traffic stop context. Imagine a patrol officer positioned in a parking lot, conducting traffic enforcement with a radar speed timing device. Over the course of a half-hour, the officer observes three vehicles exceeding the speed limit by more than 15 miles-per-hour. The first speeding vehicle is a minivan being driven by what appears to be a middle-class white woman. The officer, more motivated to uncover criminal activity than to simply issue speeding tickets, does not pursue the minivan on the assumption that this woman is unlikely to be engaged in criminal activity. The next speeding vehicle the officer observes is being driven by an elderly Asian man with white hair. The officer again ignores the traffic violation on the assumption that an elderly Asian man is unlikely to be engaged in anything more serious than speeding. The third speeding vehicle, however, is occupied by two African-American men who appear to be in their twenties. Despite ignoring the previous two vehicles that had committed the exact same traffic offense, the officer decides to pursue and stop the third speeding vehicle on the belief that young African-American men are more likely to be engaged in criminal activity.

The scenario described above is what has been termed "racial profiling" because the officer in this example used the demographic characteristics of the drivers as a "proxy" for criminal behavior. In other words, the officer judged the potential moral character of the three drivers based solely on such things as age, race, and sex. In survey after survey, most Americans have reported finding that sort of police behavior disturbing, offensive, and unethical.[1] The vast majority of Americans oppose racial profiling. As once articulated by a great American, Americans what to be judged "not by the color of their skin, but by the content of their character."[2]

Most law enforcement officers would agree. In fact, the International Association of Chiefs of Police (IACP) and the Commission on Accreditation for Law Enforcement Agencies (CALEA) have both issued formal statements against racial profiling or any other bias-based policing. These two organizations also offer model policies prohibiting the practice of bias-based policing. It is clear that the consensus both inside and outside of the law enforcement profession is that racial profiling is unjust and should not occur.

Unfortunately, the law enforcement profession and the American public disagree about the prevalence of racial profiling and other forms of bias-based policing. A 2018 Gallup Poll survey

---

[1] Gabbidon, S. L., & Higgins, G. E. (2009). The role of race / ethnicity and race relations on public opinion related to the treatment of blacks by the police. *Police Quarterly, 12(1)*, 102-115; Weitzer, R., & Tuch, S. A. (2005). Racially-biased policing: determinants of citizen perceptions. *Social Forces, 83(3)*, 1009-1030.

[2] King, Martin L., Jr. "I Have a Dream." Speech. Lincoln Memorial, Washington, DC, 28 August, 1963.

found that 52% of Americans believe that their local police treat members of minority groups unfairly in such things as traffic stops. When broken down by the race / ethnicity of those surveyed, a third of Whites, half of Hispanics and Asians, and three-quarters of African-Americans believe that their local police treat members of minority groups unfairly.[3] Because the majority of public opinion has already been swayed to believe that racial profiling is widespread, the onus is now on the law enforcement profession to disprove this assumption.

What is Traffic Stop Data Analysis?

Traffic stop data analysis has been an attempt to use social scientific techniques to reveal evidence that racial profiling in traffic stops is occurring or not occurring. The premise behind this research is that if certain demographic groups are stopped disproportionately more often than they are found within the traffic violation population, then this suggests those groups are being stopped for reasons beyond simply violating traffic laws. For example, if about 45% of the speeders in a community on any given day are white males, one could assume that if the police are enforcing the traffic laws in an unbiased manner, around 45% of the police stops for speeding should involve a white male driver. What if, however, 75% of the police stops for speeding where of white males? One could then assume that officers are ignoring speeders of other races and sexes and primarily targeting white males for stops. This would suggest evidence that officers are considering the drivers' races and sexes when deciding to make stops for speeding.

Based on this premise, many social scientists have been examining data from traffic stops since the mid-1990s. These social scientists have identified a number of scientific complexities around this issue – it is not as straightforward an analysis as they had originally thought. The accurate and fair analysis of traffic stop data for patterns of disproportionate stops of any demographic groups has proven to be difficult and complex. It requires the careful gathering of stop data, the careful gathering of appropriate benchmark data, and correct analyses of these data.

IT IS NOT AS SIMPLE AS COLLECTING STOP DATA BY STATE, COUNTY, OR MUNICIPALITY AND COMPARING THESE STOPS TO CENSUS DATA! Please stop doing that.

Various social scientists, each with decades of social science research experience and many publications in peer-reviewed research journals, have identified many best practices to be followed when conducting traffic stop data analyses. Please see the publications listed below for more details on these many best practices:

Smith, M.R., Rojek, J.J., Petrocelli, M., & Withrow, B. (2017). Measuring disparities in police activities: a state of the art review. *Policing: An International Journal, 40(2)*, 166-183.

Tillyer, R., Engel, R.S. & Calnon-Cherkauskas, J. (2010), Best practices in vehicle stop data collection and analysis. *Policing: An International Journal, 33(1)*, 69-92.

Withrow, B. L. (2005). *Racial Profiling: From Rhetoric to Reason*. Upper Saddle River, NJ: Pearson-Prentice Hall.

---

[3] Jones, J. M. (2019). *Americans Less Satisfied with Treatment of Minority Groups*. Omaha, NE: Gallup Organization.

Withrow, B. L. (2010). *The Racial Profiling Controversy: What Every Police Leader Should Know*. New York, NY: Looseleaf Law Publishers.

Unfortunately, for whatever reason, the majority of traffic stop data analysis endeavors conducted by law enforcement agencies, state governments (including attorney general offices), and civil rights groups consistently ignore these best practices. For whatever reason, these organizations continue to insist on using analysis techniques that social scientists have already shown to be fatally flawed, lack an understanding of statistical analysis weaknesses, and often produce results that are not only incorrect, but usually biased against law enforcement officers and agencies. In this workshop we will reveal how the four most common traffic stop analysis errors are flawed, and how they can contribute to data analysis efforts that are unscientific and biased against the law enforcement profession.

The four common traffic stop data analysis errors we will cover today are:
- Poor benchmarks
- Poor statistical analysis
- Aggregation bias
- Questionable data accuracy

Poor Benchmarks

A benchmark is a standard or point of reference against which things may be compared or assessed. In racial profiling research, a benchmark must be used for comparison before we can determine if "too many" drivers of a particular group are being stopped. In order for there to be a comparison to traffic stops made by officers in a jurisdiction, this benchmark must represent the sex and race / ethnicity proportions of those drivers committing traffic violations on the roadways within that jurisdiction.

For example, what if 35% of the stops for traffic violations performed by a police department involved African-American drivers. Is this a cause for concern? It is impossible to know unless we also have an appropriate benchmark measure of what percentage of the traffic law violators driving in the same jurisdiction are African-Americans. If we find that only 5% of the drivers committing traffic violations within the jurisdiction are African-Americans, then we have at least initial evidence to be concerned if 35% of stops involved members of that group despite being found on the roads so rarely. If, however, African-American drivers are found to make up 34% of all traffic violators driving within the community, then there would be no evidence for concern. Members of this groups would appear to be stopped in proportion to their representation among the traffic violating population within the jurisdiction.

The key problem in testing for racial profiling in traffic stops is estimating the "benchmark" against which to compare the race distribution of stopped drivers. For two decades now, ***it has been widely recognized that Census-based population data provide poor estimates of the population at risk of a traffic stop***. It is now scientific malpractice to continue to utilize Census-based population data as a benchmark for traffic stop data analysis. The studies below are some of the many sources that attest to the lack of validity of Census-based data to estimate the racial composition of driving populations.

Alpert, G. P., Smith, M. R., & Dunham, R. G. (2004). Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy, 6(1)*, 43-69.
Engel, R. S., & Calnon, J. M. (2004). Comparing benchmark methodologies for police-citizen contacts: traffic stop data collection for the Pennsylvania State Police. *Police Quarterly, 7(1)*, 97-125.
Engel, R. S., Calnon, J. M., Tillyer, R., Johnson, R. R., Liu, L., Wang, X. (2005). *Project on Police-Citizen Contacts: Year 2 Final Report*. Cincinnati, OH: University of Cincinnati.
Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association, 101(475)*, 878-887.
Herbert-Martinez, K. L., & Porter, B. E. (2006). Characterizing red light runners following implementation of a photo enforcement program. *Accident Analysis & Prevention, 38*, 862-870.
Lange, J. E., Johnson, M. B., & Voas, R. B. (2005). Testing the racial profiling hypothesis for seemingly disparate traffic stops on the New Jersey turnpike. *Justice Quarterly, 22(2)*, 194-223.
Meehan, A. J., & Ponder, M. (2002). How roadway composition matters in analyzing police data on racial profiling. *Police Quarterly, 5(3)*, 306-333.
Smith, W., Tomaskovic-Devey, D., Zingraff, M., Mason, H., Warren, P., & Wright, C. (2004). *The North Carolina Highway Traffic Study*. Washington, DC: National Institute of Justice.
Withrow, B. L. (2004). A comparison of commonly used benchmarks in racial profiling: a research note. *Justice Research and Policy, 6(1)*, 71-92.

In summary, these studies have revealed that:

- Census-based benchmarks (including registered driver counts) for a given geographical area never resemble the actual driving population.
- A large proportion of the drivers encountered on the roadways in any geographical location (Census block, zip code, city, etc.) reside outside of that location.
- As a result, Census-based data lack validity for use as a measure of the driving population at any location.

1. Alternative Benchmark Options

Roadway Observations
This usually involves employing researchers to observe the drivers on the roadways of the jurisdiction and record the apparent sex and race / ethnicity of the drivers observed. These observations sample the drivers found on the roadways for comparison to police stops in the same areas and during similar times of day. The limitations of using roadway observations as a benchmark are that these observations are costly and time consuming to collect, they tend to be limited to areas with clear lines of sight, are usually limited to a few specific roadway segments, are limited to days and times with good visibility (i.e., no night or poor weather observations), and the race determination of the drivers are estimates based on seeing the driver of a moving vehicle. Nevertheless, the validity of such methods has been tested and found more valid than

Census data for recording actual drivers on the roadways. The studies below are examples of traffic stop data analysis endeavors that have used roadway observations as their benchmark.

Engel, R. S., Calnon, J. M., Tillyer, R., Johnson, R. R., Liu, L., Wang, X. (2005). *Project on Police-Citizen Contacts: Year 2 Final Report*. Cincinnati, OH: University of Cincinnati.

Meehan, A. J., & Ponder, M. (2002). How roadway composition matters in analyzing police data on racial profiling. *Police Quarterly, 5(3)*, 306-333.

Smith, W., Tomaskovic-Devey, D., Zingraff, M., Mason, H., Warren, P., & Wright, C. (2004). *The North Carolina Highway Traffic Study*. Washington, DC: National Institute of Justice.

Automated Observations

Automated observations are similar to roadway observation except that the observations of the drivers are initially made and recorded by some mechanical device, such as an automated license plate reader or automated speed or traffic signal camera. The automated device photographs the driver and / or license plate of vehicles operating on the roadways within the jurisdiction, then researchers later determine the sex and race of either the driver (if the driver is photographed), or the vehicle's registered owner (if only the license plate is recorded). The limitations of this method include being costly and time consuming (still need observers to look at the photos and determine driver race), they are limited to areas with clear lines of sight and usually limited to a few specific roadway segments, and they are limited by night and weather conditions. Nevertheless, these methods have been tested and proven more valid than Census data. The studies below are examples of traffic stop data analysis endeavors that have used automated observations as their benchmark.

Lange, J. E., Blackman, K. O., & Johnson, M. B. (2001). *Speed Violation Survey of the New Jersey Turnpike: Final Report*. Trenton, NJ: New Jersey Attorney General's Office, Public Services Research Institute.

Lange, J. E., Johnson, M. B., & Voas, R. B. (2005). Testing the racial profiling hypothesis for seemingly disparate traffic stops on the New Jersey turnpike. *Justice Quarterly, 22(2)*, 194-223.

Herbert-Martinez, K. L., & Porter, B. E. (2006). Characterizing red light runners following implementation of a photo enforcement program. *Accident Analysis & Prevention, 38*, 862-870.

"Veil of Darkness" Method

This method uses the race distribution of police stops made during hours of darkness (when it is assumed that officers generally cannot see the driver's race prior to stop) as the benchmark for comparison to stops made by officers during daylight hours (when it is assumed that officers generally can see the race of the drivers prior to stop). The limitation here is that this method assumes no differences, daytime to nighttime, in vehicle stop behavior by the police, assumes that officers really can tell the races of drivers (pre-stop) in daylight and not at night, and assumes no racial differences in traffic patterns from daytime to nighttime. The studies below are examples of traffic stop data analysis endeavors that have used the veil of darkness method.

Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association, 101(475)*, 878-887.
Lundman, R. J., & Kowalski, B. R. (2009). Speeding while black? Assessing the generalizability of Lange et al.'s (2001, 2005) New Jersey Turnpike Speeding Survey findings. *Justice Quarterly, 26(3)*, 504-527.
Worden, R. E., McLean, S. J., & Wheeler, A. P. (2012). Testing for racial profiling with the veil-of-darkness method. *Police Quarterly, 15(1)*, 92-111.

Crash Driver Data

This benchmark method uses the drivers involved in crashes within the jurisdiction as a sample of the drivers (especially poor drivers) on the roadway. When officers take crash reports, the sex and race / ethnicity of all drivers involved in the crash are recorded and become a measure of who is driving within the jurisdiction. The limitations of this method are that it usually requires collecting driver race and sex data in-house as most states no longer record such data on crashes, and another limitation is that crashes tend to cluster in certain areas so some areas of the jurisdiction may still lack a benchmark measure. The benefit of this method it that it is citizen-initiated, crashes occur all over jurisdiction, in all weather, any time of the day. In this method the race and sex of the benchmark drivers are determined with greater accuracy, and this method is much cheaper and more efficient than other benchmark methods. The studies below have utilized the crash driver benchmark method.

Alpert, G. P., Smith, M. R., & Dunham, R. G. (2004). Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy, 6(1)*, 43-69.
Dolan Consulting Group (2018). *Evaluating Fairness in Traffic Stops by the Ann Arbor Police Department, Final Report*. Raleigh, NC: Dolan Consulting Group.
Withrow, B. L., & Williams, H. (2015). Proposing a benchmark based on vehicle collision data in racial profiling research. *Criminal Justice Review, 40(4)*, 449-469.

REMEMBER: CENSUS-BASED DATA HAVE LONG AGO BEEN DEMONSTRATED TO BE A BIASED BENCHMARK THAT LACKS VALIDITY FOR USE AS A BENCHMARK TO ESTIMATE DRIVING POPULATIONS. STOP USING IT!

NOTES:

2. Poor Statistical Analysis

Most faulty traffic-stop data analysis endeavors have relied upon the use of the odds-ratio, or disproportionality index, as a statistical analysis technique. This analysis technique is often misinterpreted by laypersons who do not understand the statistical / mathematical principle of the law of large numbers. When dealing with small numbers, a small unit change can result in a large proportional change. For example, if there are two people (one male and one female), even though only one individual is female, she makes up 50% of the sample. If only one more person is added to the group, and that person is also female, the number of females in the sample increased by 100%! As traffic stop data analyses are concerned with the representation of minority groups within the stops, by the very nature of being a **minority** (i.e., fewer in number than the majority), these groups will represent smaller proportions. Therefore, the use of the disproportionality index inflates the impact of minor changes, such as one or two more traffic stops than was expected.

Disproportionality Index (DI)

$$\textbf{DI} = \frac{\textbf{Percentage of race actually encountered (actual stops made)}}{\textbf{Percentage of race expected if no bias (benchmark)}}$$

EXAMPLE 1:
35% of the stops for speeding were Hispanic drivers, and 29% of the speeders in the community are Hispanic. What is the DI?

$$\textbf{DI} = \frac{\textbf{Percentage of race actually encountered (35\%)}}{\textbf{Percentage of race expected if no bias (29\%)}}$$

35% ÷ 29% = 1.21
The Disproportionality Index is 1.21

EXAMPLE 2:
55% of the stops for speeding were white drivers, and 65% of the speeders in the community are white. What is the DI?

$$\textbf{DI} = \frac{\textbf{Percentage of race actually encountered (55\%)}}{\textbf{Percentage of race expected if no bias (65\%)}}$$

55% ÷ 65% = 0.85

The Disproportionality Index is 0.85

To interpret the Disproportionality Index, do the following: (DI – 1.00) x 100 = % more or less likely

DI 1.00 = No disproportionate outcome (0% more or less likely)

DI >1.00 = More likely than expected

DI <1.00 = Less likely than expected

Note that the DI is unstable with small numbers. If 60% of persons stopped were white, and 55% of drivers at risk of stop were white, how many percentage points is 60% from 55%? 60% - 55% = 5 percentage points. Now, what is the DI for this difference?
What is the DI? 60% ÷ 55% = **1.09** or white drivers were 9% more likely to be stopped than was expected, even though the difference was only 5 percentage points.

Now, what is 6% of persons stopped were Asian, and 1% of drivers at risk were Asian? How many percentage points is 6% from 1%? 6% - 1% = 5 percentage points, just like the example above. But what is the DI for this difference?
What is the DI? 6% ÷ 1% = **6.00** or Asian drivers were **500%** more likely to be stopped than was expected, even though the percentage point difference was still only 5 percentage points!

WHEN DEALING WITH SMALL NUMBERS OR PERCENTAGES, A ONE UNIT CHANGE CAUSES A MUCH LARGER PROPORTIONAL CHANGE.

Alternative to the DI
Instead of using the DI, simply report the estimate the number of actual human beings impacted by any stop disparities. You know how many stops were actually made, and your benchmark (if a proper one is used) tells you what percentage of those stops should have included which groups. For example, if 75% of the drivers involved in crashes within your jurisdiction were white, and your agency made 1000 traffic stops, you could estimate that about 750 of the drivers actually stopped should have white. If, in fact, 800 white drivers were stopped, you would report that out of 1000 traffic stops, your officers stopped 50 more white drivers than would have been expected based on drivers involved in crashes (the benchmark).

Is this evidence of systematic bias against white drivers? That depends on the total number of stops and the size of your agency. Out of the 52 weeks in a year, 50 more stops of white drivers suggests less than 1 extra stop of a white occurred (on average) each week. If there are 30 officers employed by your agency, that means there was 1.67 more white stops per officer than expected for the entire year. Broken down by week, that means there were 0.032 extra stops of white drivers per officer per week. Broken down by day, this means there were 0.0046 extra stops of white drivers per officer per day. In other words, based on this average, each officer stopped ***one*** more driver than expected, every ***213 days***. This lacks evidence of officers systematically targeting white drivers for stops.

NOTES:

3. Aggregation Bias

In statistics, aggregation bias is the difference between statistical outcomes for the group and statistical outcomes for the individuals that make up that group. Aggregation bias leads to an "ecological fallacy" — the conclusion that what is true for the group must be true for the sub-group or individual. It's called aggregation bias because you're using aggregated data and extrapolating it inappropriately. For example, you might have data showing that inner city students tend to perform poorly on standardized tests. That doesn't mean any one individual, or one particular school within that school district, will perform poorly. Likewise, you might show that one particular state has a lower than average per-capita income. You cannot say for sure that every county in that state has a lower than average income.

Analyzing traffic stop data at the state, county, or even city / town level is analyzing aggregate data – lumping together stops from various areas and times of day that may have very different racial differences in traffic patterns. For example, what if 75% of the drivers in the east side of town are African-Americans, and only 50% of drivers in the west side of town are African-Americans? The aggregate (average) for the town is 52.5% of drivers are African-American if this is not broken down by sides of town.

Now, what if 110 traffic stops occur in the east side of town and 75% of the drivers are African-American (as they should be)? This means 83 African-American's stopped (75% of the east side stops). Now what if 90 stops occurred on the west side, and 50% of the drivers were African-American (as they should be)? That means 45 west side stops were of African-American drivers (50%). But if the state reports the city's data in aggregate, out of 200 total traffic stops, 128 stops (64%) involved African-American drivers, even though the aggregate benchmark was 52.5% African-American drivers. Even though both sides of town had stops that perfectly matched the benchmark for their sides of town, because there were a few more stops made on one side of town, aggregating the data produced an untrue aggregate outcome. False evidence of disproportionate stops gets reported every time we aggregate our data.

The ability to disaggregate in traffic stop data analysis is limited by the benchmark being used. You need ***at least*** 100 benchmark observations (i.e., crash drivers, drivers observed, etc.) for an accurate benchmark. It is preferable to have hundreds or even thousands of benchmark observations to improve validity. If you break down your agency's stops by shift or district and end up with fewer than 100 benchmark observations, by necessity you need to aggregate up to the next higher level. As a result, small agencies with a dozen or fewer officers likely will not be able to disaggregate at all and will have no option but to examine their data at the agency level. Large cities with hundreds of officers, on the other hand, can usually break down their data and analyze stops by precincts and shifts.

NOTES:

4. Data Accuracy

Data quality matters, especially with regards to the most important variable involved in traffic stop data analysis – the driver's race. You need good data because garbage in = garbage out. Data needs to be valid and reliable. Validity refers to how accurately a method measures what it is intended to measure. If research has high validity, that means it produces results that correspond to real properties, characteristics, and variations in the real world. In other words, the races / ethnicities of the drivers being stopped are being recorded correctly. Reliability refers to how accurately a method measures what it is intended to measure every time. In other words, the races / ethnicities of drivers are being recorded in the same way (to the same standard) every time.

Lack of data or missing data is a serious problem. Officer and union buy-in will help the agency, but incomplete data suggests deceit to the public and only confirms their suspicion that racial profiling is happening. However, in order to ensure that the race data being collected is both valid (accurate) and reliable (measured the same way each time), the race / ethnicity categories need to be accurate and consistent across databases.

All federal government agencies utilize the U.S. Census Race / Ethnicity Categories implemented by the Clinton Administration in 1996 were developed by a panel of sociologists, anthropologists, and scientists to categories humans based on the most scientific principles. Race refers to the genetic features (such as skin pigment, hair color and texture, eye color, nose structure, and many other biological features) we inherited from our ancestors based on the climate in which those ancestors inherited. Race is biological – what you are examining when you have an ancestral DNA test – but it is important to note that almost ALL humans today are multiracial because of human group sexual interaction over time. Nevertheless, a majority of people in the world still display the physical characteristics of at least one predominant ancient racial group. Ethnicity, on the other hand, refers to culture, which is often closely tied to language. Race is sometimes hard to determine based on visual cues, and ethnicity is even harder to determine without questioning the individual.

U.S. Census Bureau Race Categories:

White. A person having origins in any of the original peoples of Europe, the Middle East, or North Africa. It includes people who indicate their race as "White" or report entries such as Irish, German, Italian, Lebanese, Arab, Moroccan, or Caucasian.

Black or African American. A person having origins in any of the Black racial groups of Africa. It includes people who indicate their race as "Black or African American," or report entries such as African American, Kenyan, Nigerian, or Haitian.

American Indian and Alaska Native. A person having origins in any of the original peoples of North and South America (including Central America) and who maintains tribal affiliation or community attachment. This category includes people who indicate their race as "American Indian or Alaska Native" or report entries such as Navajo, Blackfeet, Inupiat, Yup'ik, or Central American Indian groups or South American Indian groups.

Asian. A person having origins in any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. This includes people who reported detailed Asian responses such as: "Asian Indian," "Chinese," "Filipino," "Korean," "Japanese," "Vietnamese," and "Other Asian" or provide other detailed Asian responses.

Native Hawaiian and Other Pacific Islander. A person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands. It includes people who reported their race as "Fijian," "Guamanian or Chamorro," "Marshallese," "Native Hawaiian," "Samoan," "Tongan," and "Other Pacific Islander" or provide other detailed Pacific Islander responses.

Two or More Races. People may choose to provide two or more races either by checking two or more race response check boxes, by providing multiple responses, or by some combination of check boxes and other responses. For data product purposes, "Two or More Races" refers to combinations of two or more of the following race categories: "White," "Black or African American," American Indian or Alaska Native," "Asian," Native Hawaiian or Other Pacific Islander," or "Some Other Race"

Ethnicity:
Hispanic

Arabic
And over 400 other categories

Race and ethnicity identification accuracy is crucially important to traffic stop data analysis. How accurately is race being measured by your officers? Are you using race / ethnicity categories based on the recommendations of scientists, or do you use categories created by laypersons lacking experience or knowledge in this area? Do your officers record race and ethnicity differently? Race and ethnicity are not the same thing. How do your officers know how to categorize individuals by race on traffic stops? Do all of your officers use the same standards?

Determining race and ethnicity on physical appearance can be extremely difficult. Improper identification may falsely increase minority representation in certain categories during traffic stops. What if one member of your department, due to lack of training and / or life experience, inaccurately labels a driver of Indian, Pakistani, or Bangladeshi ancestry as being Hispanic? In such a case, your agency fails to record a stop of an Asian driver, but also records a stop of a Hispanic driver **that never actually happened**! What if this officer makes this error about once a month? At the end of the year your agency will underreport 12 stops of Asian drivers, but also report 12 stops of Hispanic drivers **that never actually happened**!

Despite the fact that race and ethnicity identification is extremely difficult at times and has questionable validity or reliability, this is the data that is being used as "proof" that your agency is, or is not, engaging in racial profiling. This is a serious problem.

NOTES:

Conclusion

The purpose of traffic stop data analysis is to examine for evidence of biased policing against certain demographic groups. Many social scientists have approached such analyses in an unbiased, scientific manner and have revealed there are many difficulties to examining such data correctly. Accurate benchmark measures are difficult to obtain, and likely none are perfectly accurate. Nevertheless, Census-based benchmarks have been repeatedly proven to not only be invalid measures of driving populations, but also consistently biased against the police in suburban and rural areas. Beyond the use of proper benchmarks, the way the data is analyzed also is extremely important as using the disproportionality index and reporting aggregate data consistently result in untrue results or results laypersons misinterpret. Finally, the validity of the race data upon which these data are based – and officers' reputations depend – is questionable from the start.

Perhaps one reason so many Americans today are convinced that the practice of racial profiling by the police is widespread is because so many government agencies, civil rights organizations, and activist groups have ignored the documented best practices for traffic stop data collection. Despite all of the expert evidence, published in government reports and peer-reviewed scientific journals, these entities refuse to follow best practices and insist on using methods that are guaranteed to suggest police bias when no such bias actually exists.