Driver and Vehicle-Based Search Predictors: 2011-2016 Texas Highway Patrol Traffic Stop Data Analysis

Will Spillman

POLI 490H, Prof. Baumgartner
UNC-Chapel Hill, Fall 2019

Abstract

Racial bias in traffic stop searches is a well-established and pervasive national issue, but no comprehensive studies have been done that examine concurrent vehicle-based biases that could affect an officer's decision to search. A 2011-2016 traffic stop dataset from the Texas Highway Patrol, with descriptive variables concerning both the vehicle and the driver, provided an opportunity to conduct this kind of analysis. In this study, vehicle perception (luxury, average, or cheap) was classified by car make, and vehicle age was condensed into three categories (new, average, or old). This analysis reveals the predictive value of these variables while accounting for the driver's race and sex, and the results confirm and expand upon previous research. While racial bias against black male drivers is the largest single predictor of a traffic stop search, old cars draw a similar level of suspicion, and the effect of a vehicle's reputation, while negligible in total, can be significant for specific combinations of driver race and sex and vehicle age. In total, there were search rate disparities across all variables, and consideration of hit rates reveals that these differences were not the result of routine policing.

Keywords: perception, age, race, sex, search rate, hit rate

Introduction

While the obvious racism that characterized many mid-1900s U.S. police departments has faded, law enforcement interaction with minority groups remains a scrutinized issue. In recent decades, organizations and individuals dedicated to racial equality have become concerned with implicit racial bias in policing, a phenomenon that began as a suspicion and has since been confirmed by a large body of research. Political pressure has forced states and police departments across the country to release traffic stop data, and for this reason, most studies investigating police discrimination have relied on these records (Anderson 2019). Although traffic stop data contains plenty of information unrelated to motorists, these variables have been neglected due to the intense debate surrounding police racial bias. So far, no large-scale analyses have examined the potential collinearity of vehicle characteristics and driver demographics with respect to search rates. This paper will test the hypotheses that old cars will be searched more than new cars, cheap cars will be searched more than luxury cars, and the effect of racial bias on search rates will be diminished by consideration of vehicle-based variables.

Background

Since the mid-1990s, research concerning the idea of "driving while black" has exploded (Harris 1999). The issue has taken a prominent role in the national dialogue, with organized movements supported by millions of people protesting the racial injustice in both the criminal justice system and law enforcement. Part of the reason for the extensive amount of literature on race and policing is due to the fact that systematic discrimination against blacks is no longer direct and obvious. Rather, the concern is that subconscious perceptions of race and implicit biases could affect police officers' actions. According to the American Psychological Association, "one of the most well-demonstrated types of implicit bias is the unconscious association between black individuals and crime" (Weir 2016). Many experiments have investigated and found evidence that these unfounded perceptions are widespread, and the worry is that they could affect law enforcement officials just as much as average citizens (Weir 2016). Likewise, many data analyses have identified racial disparities in traffic stop search rates (Schafer 2006). The question is whether implicit biases, which have been shown to exist in the general population, drive these disparities or whether they are the result of omitted variables or routine police work. In a 2006 research paper, Joseph Shafer investigates this question and comes to the conclusion that "the view that profiles involving race are a matter of 'good' and expedient police work is not supported by this analysis" (Schafer 2006). This will be one of the main focuses of my work after I demonstrate that racial and gender disparities in search rates do exist in the Texas Highway Patrol Dataset. Research like this is necessary to heal the tensions that our country currently faces. If problems in policing are not exposed, there will be no change, and no change means no progress. In his book titled "Suspect Citizens: What 20 Million Traffic Stops Tells Us About Policing and Race," Frank Baumgartner demonstrates how research can lead to reforms. After North Carolina traffic stop data was analyzed, Fayetteville, N.C., acknowledged certain shortcomings and began to "emphasize traffic safety rather than regulatory enforcement" (Sides 2018). These changes were successful, and resulted in more favorable opinions of the police without leading to an increase in crime rate (Sides 2018). If more analysis can be done, more progress will occur. That is my primary motivation for this data analysis, and I hope to add to this extensive body of research surrounding race and policing, while also examining the effect of vehicle characteristics.

Database Re-Configuration and Variable Creation

The database that I analyzed for this paper contained more than 12 million observations of Texas Highway Patrol traffic stops from the years 2011-2016. This enormous sample size provided several benefits. First, it allowed me to study the search and hit rates of several specific groups. For example, only 14,287 observations fit the description of "Black Male – Old Luxury Car," which represented 0.001% of all stops. However, despite the scarcity of this group relative to the total number of stops, the sample size was still large enough to provide meaningful results. Second, it ensured the accuracy of the mean search and hit rate values for the entire dataset (about 20% and 40%, respectively) and essentially pushed the standard error of regression coefficient estimates to zero. The only downside of the dataset's immensity was the increased number of spelling mistakes in the "vehiclemake" column, but the process of correcting these errors was more tedious than time-consuming.

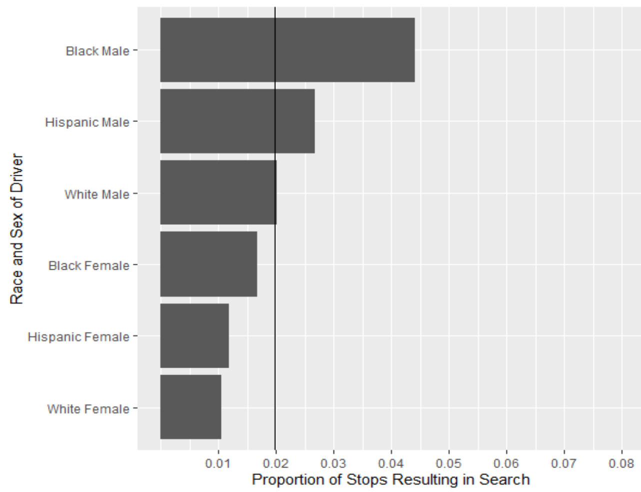
Once the spelling mistakes were fixed, I re-configured the dataset to eliminate irrelevant observations. First, I removed all observations in which the driver's race was not either white, black, or Hispanic. This is because the other races were not represented at a high enough rate to support subclassification. Second, I dropped all observations that had unclear or uncommon vehicle makes, which left 31 distinct car brands in the database that would later be classified into a "vehicle perception" variable. Lastly, I cut the few observations that had missing "vehicleage" and "searchoccur" values for simplification purposes.

To produce understandable results, I needed to create uncomplicated categorical variables that summarized the "vehiclemake and "vehicleage" columns, both of which had over 20 possible values. To simplify "vehiclemake," I created a variable called "vehicle perception" that had three possible values (luxury, average, or cheap). To determine the vehicle perception value of a certain make, I relied on my own judgement. The majority of makes were categorized as average, while the luxury and cheap categories contained seven and five distinct makes, respectively. The seven luxury brands were Porsche, Audi, BMW, Jaguar, Lincoln, Mercedes-Benz, Cadillac, and Volkswagen, while the five cheap brands were Hyundai, Volvo, Mitsubishi, Mazda, and Saturn.

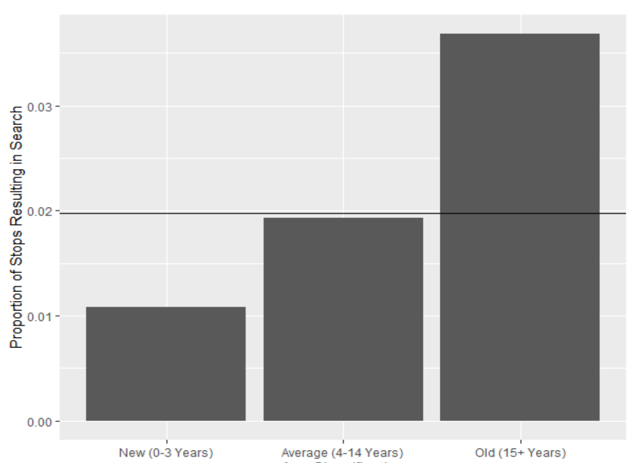
In the original dataset, the “vehicleage” variable had possible values of 0, 1, 2, etc… into the upper-20s. Because the total number of observations decreased significantly for “vehicleage” values greater than or equal to 15, I re-coded the data to make the maximum vehicle age “15+”. Then, I created variable called “age classification” with three possible values (new, average, or old). Cars between the vehicle ages of 0-3, 4-14, and 15+ were considered new, average, and old, respectively. Once again, the categorization of the age classification variable was influenced by my own experience, as these definitions are naturally arbitrary. With the dataset finalized, I started the process of collecting results that assessed the predictive value of driver and vehicle-based variables on search and hit rates.

Results

I generated three initial graphs to answer two basic questions. First, are there search rate disparities between different driver race/sex combinations? Second, are there search rate disparities between vehicles with different age classification and vehicle perception values? Figure 1 provided an answer to the first question:



The vertical line that intercepts the x-axis at about 0.02 represents the average search rate across all observations. From the years 2011-2016, the Texas Highway Patrol searched ~2% of all drivers after a traffic stop. According to this figure, white males were searched at the average rate, black males were searched at ~2.25 times the average rate, and Hispanic males were searched at ~1.25 times the average rate. On top of this racial disparity, there is a clear gender disparity, as male drivers of all races were searched more than female drivers. Next, I examined possible differences in search rates with respect to vehicle-based variables, starting with age classification. Figure 2 presents this relationship:



There is a clear variation in search rates for cars of different age classification categories, with the trend following what I expected in my hypothesis. New cars were searched at half the rate of average-aged cars, while old cars were searched ~1.75 times more than average-aged cars. Once I discovered this vehicle age search rate disparity, a new question surfaced: is the age classification variable correlated to the race/sex variable? The results of two separate regression models helped resolve this uncertainty. The first regression model presents the exact same findings as shown in Figure 1, but in a different format:

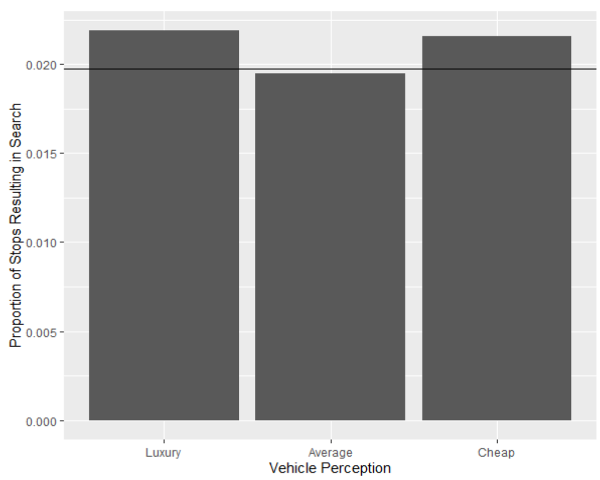
	Coefficient Estimate
(Intercept)	0.02
White Female Driver	-0.01
Black Male Driver	0.024
Black Female Driver	-0.003
Hispanic Male Driver	0.007
Hispanic Female Driver	-0.008

In this table, the coefficient estimate of the intercept represents the proportion of stops that resulted in searches for white male drivers. The coefficient estimate for a particular race/sex combination represents the predicted change in the proportion of stops that would result in a search if the race/sex of the driver switched from white male to that combination. If the coefficient estimates change once I add the age classification variable to the regression model, it would indicate that part of the racial disparity in search rates can be explained by the vehicle's age. The second regression model shows almost no relationship between these two independent variables:

	Coefficient Estimate
(Intercept)	0.02
White Female Driver	-0.008
Black Male Driver	0.023
Black Female Driver	-0.002
Hispanic Male Driver	0.005
Hispanic Female Driver	-0.008
New Car (0-3 Years)	-0.008
Old Car (15+ Years)	0.016

This more inclusive regression model contradicts my hypothesis. According to the table, the coefficient estimate for old cars is 0.016, and the estimate for new cars is -0.008. I would have expected the appearance of these significant predictors to either increase or decrease the coefficient estimate for black male drivers, however, it remained stable at 0.023. Similarly, the other race/sex combinations did not experience significant changes in their coefficient estimates. While I was correct that old cars were searched more than new cars, this trend existed independent of the observed race/sex disparity. After this second regression, age classification emerged as a clear predictor of search rates alongside certain race/sex combinations.

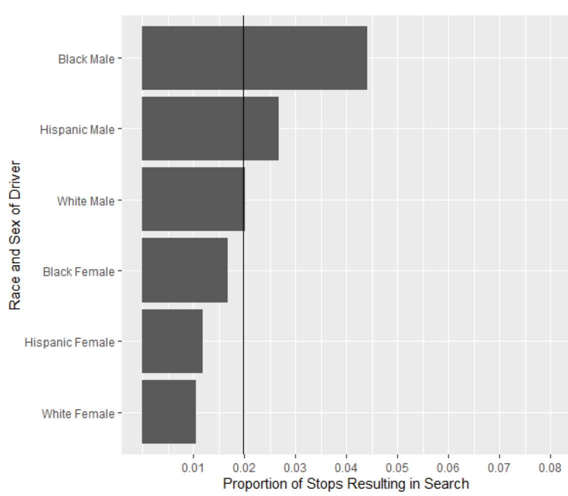
Finally, I sought to determine if there were search rate disparities between vehicles with different public perceptions. Figure 3 plots vehicle perception against the proportion of stops resulting in a search:

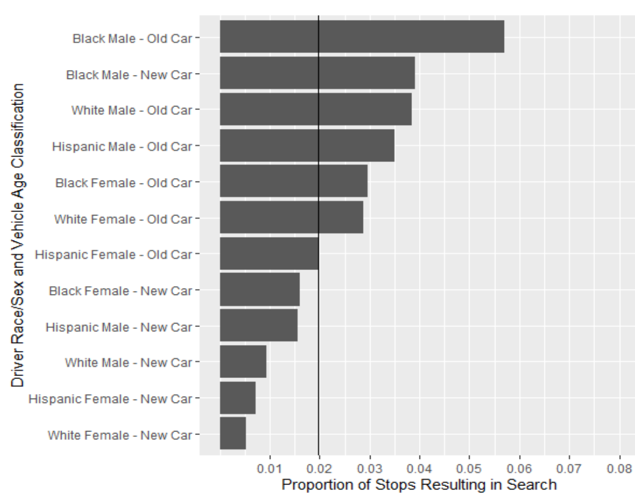


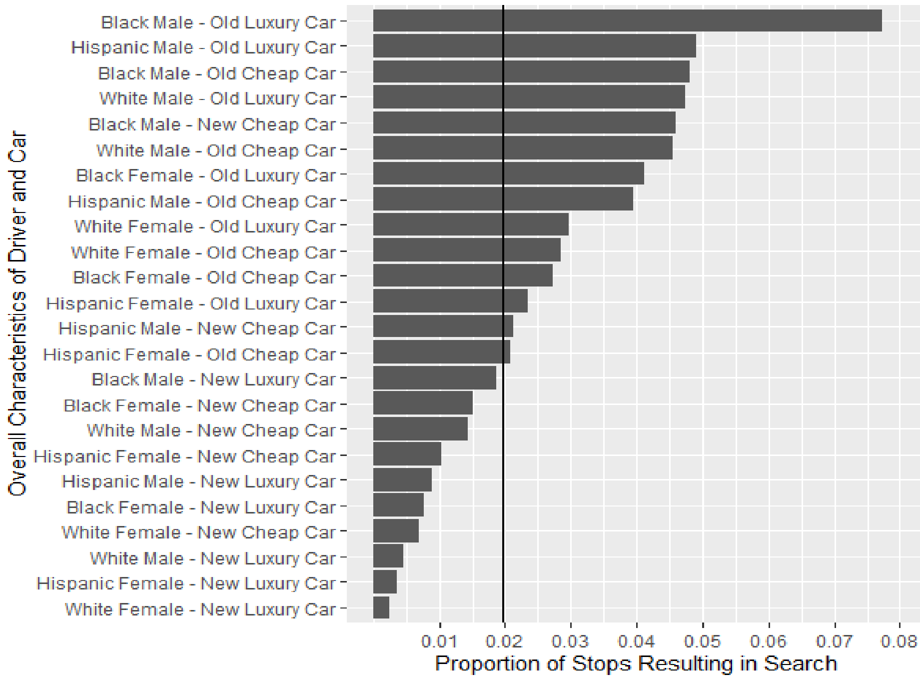
Although Figure 3 shows that search rates for luxury and cheap cars are higher than the overall average rate, this disparity appears insignificant. The y-axis breaks in this figure are in increments of 0.005, which means that luxury and cheap cars were only searched ~1.1 times more than the average. However, it is possible that the “vehicle perception” variable is associated with vehicle age or driver race/sex. This could be the reason why there is no observable disparity in Figure 3. In the third and final regression model, I sought to determine the overall predictive value of vehicle perception:

	Coefficient Estimate
(Intercept)	0.019
White Female Driver	-0.008
Black Male Driver	0.023
Black Female Driver	-0.003
Hispanic Male Driver	0.005
Hispanic Female Driver	-0.008
New Car (0-3 Years)	-0.008
Old Car (15+ Years)	0.016
Luxury Car	0.001
Cheap Car	0.004

Although it was unlikely that vehicle age or driver race/sex masked large search rate disparities between vehicles with different reputations, this table confirms that vehicle perception by itself does not significantly influence an officer's decision to search after a traffic stop. However, this regression model only considers the overall effect of the variables. What if vehicle perception, when combined with a certain driver demographic and vehicle age, had a pronounced effect? In other words, could the specification of whether a car is luxury or cheap complete an existing stereotype that includes a driver's race/sex and vehicle age? To test whether this is the case, I generated a progression of three figures, beginning with Figure 1, that shows the incremental effects on search rates of both vehicle-based variables:

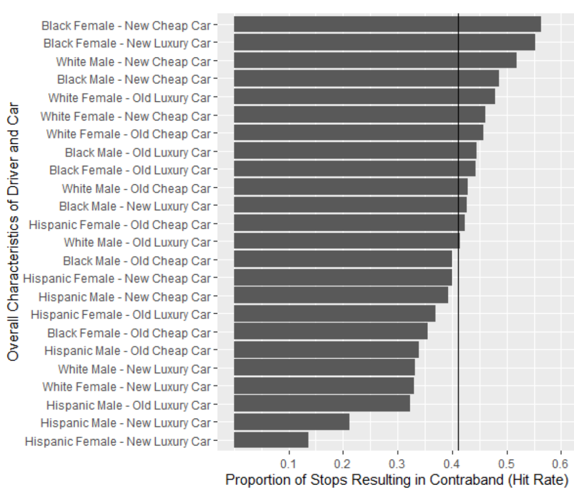




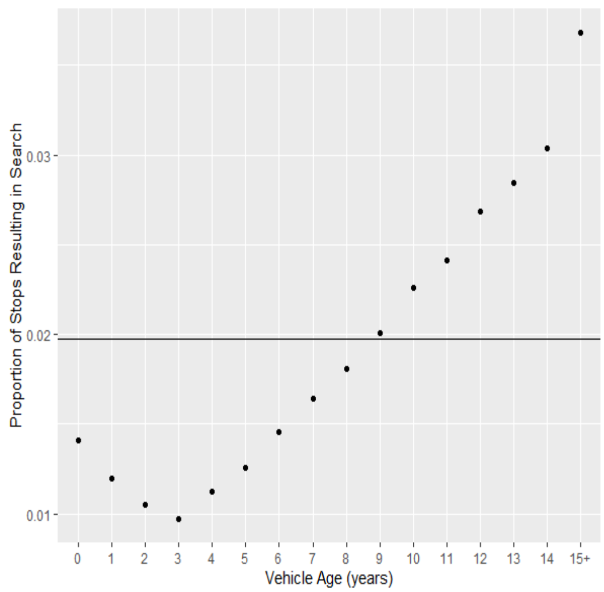


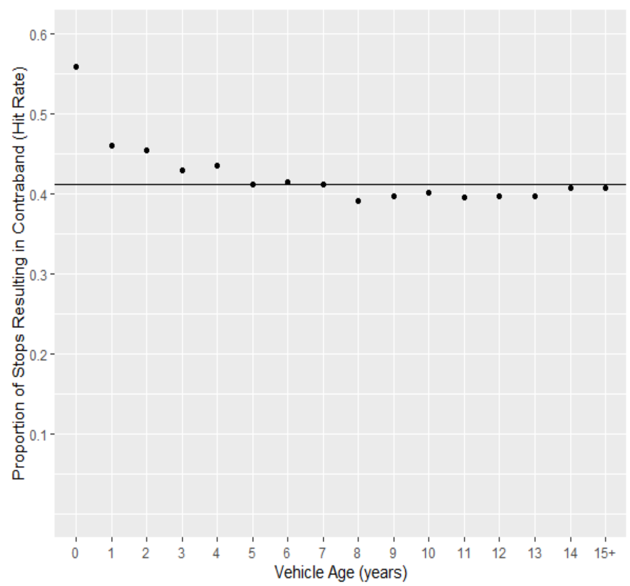
On top of providing the search rate value for 24 distinct combinations of driver and vehicle characteristics over a six-year period, the final graph depicts another surprising and important result. Compared to black males driving an old car who are searched after ~5.7% of traffic stops, black males driving an old ***luxury*** car are searched after ~7.7% of traffic stops. The estimated coefficient for luxury cars in the holistic regression was 0.001, which means the specification that someone is driving a luxury vehicle should correspond to a .1% increase in the likelihood they are searched. However, in this case, that one specification corresponded to a 2% increase in the probability of a search, almost 20 times the expected percentage. A similar phenomenon occurs with Hispanic male drivers. When these individuals were driving an old car, they were searched after ~3.5% of stops. However, Hispanic males driving an old ***luxury*** car were searched after ~5% of stops, about 15 times the expected increase in search likelihood. Furthermore, for black and Hispanic male drivers, the reverse is true when vehicle perception specificity is added to new cars. In these cases, the black and Hispanic male drivers are searched more when the car is cheap and less when the car is luxury. For white males and female drivers of all races, vehicle perception specificity does not seem to significantly affect the likelihood of a search. This explains why the clear effect of this variable on black and Hispanic male drivers is not represented in the holistic regression. The results of this figure progression present one more question: are these disparities a result of routine policing? In other words, do some of these groups differ in their tendencies for criminality, and if so, do the search rate disparities between them correspond to matching disparities in criminality?

I will first compare the search and hit rates of the 24 groups defined by the driver's race/sex, vehicle age, and vehicle perception. In this graph, the y-axis will still represent the overall characteristics of the driver and car, but the x-axis has been modified to depict the "Proportion of Stops Resulting in Contraband," a number more commonly known as the hit rate. Once again, there is a vertical line that intercepts the x-axis at the mean hit rate across all observations in the dataset (~41%). This figure provides the necessary context to determine if the search rate figure illustrated police bias:

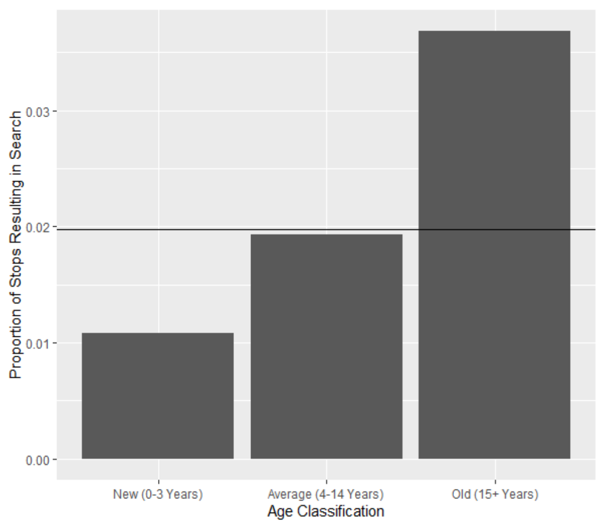


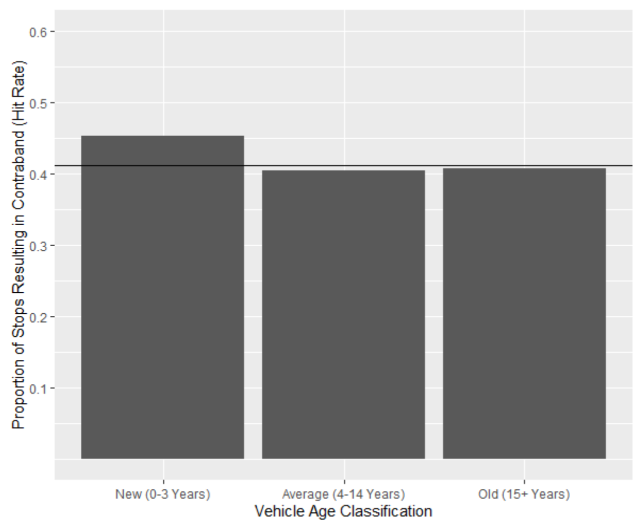
This graph offers strong evidence that the disparities observed in the search rates of 24 specific combinations of driver and vehicle characteristics can be explained, at least in part, by implicit police bias. Again, the comparison of black and Hispanic males to white males supports this interpretation. Black males driving an old luxury car, a group that is searched nearly 4 times more than the average driver, is found with contraband at a rate barely above the average. The same is true for Hispanic males driving old luxury cars. Despite being found with contraband after ~32% of traffic stops, a hit rate that is almost 25% lower than the average of ~41%, this group was more than 1.5 times more likely to be searched than the average driver. Additionally, both black and Hispanic male drivers driving old cheap cars had less-than-average hit rates, but were subjected to searches over 2 times more than the average driver. Another interesting aspect of this graph is that the two groups of black females that exhibited the most criminality of all 24 groups were searched well below the average rate. However, the drivers that get off the easiest are white females in new luxury cars. Even though the hit rate for this group was ~80% of the average, the search rate was only ~12.5% of the average. Finally, the graph seems to depict a general bias against old cars. In the search rate figure, the order of the y-axis was almost entirely divided by age classification, with old cars having the higher search rates. In the hit rate figure, this division disappears, which signals that age classification has a greater impact on search rates than hit rates. To test this conjecture, I will present search and hit rate scatterplots with vehicle age (by year) as the independent variable:





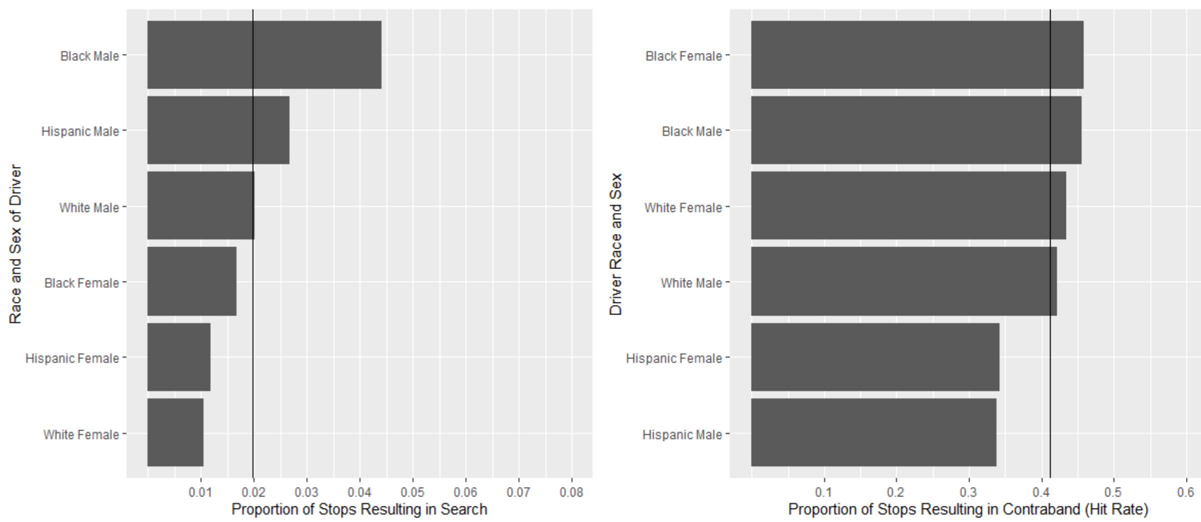
By itself, the first plot identifies a stunning connection between a vehicle's age in years and the probability that it will be searched. When both plots are considered together, it confirms that this tight association was not formed through evidence-based policing. In the search rate plot, there is an initial dip in search rates from the vehicle ages of 0-3, followed by an observable linear increase for every vehicle age from 3-14. For vehicles aged 15+ years, the search rate jumps, departing from the established trend. All vehicle ages from 0-8 were searched less than 2% of the time, while all vehicle ages from 9-15+ were searched more than 2% of the time. These clear search rate disparities in vehicle age are not present in the hit rate scatterplot. In fact, the hit rate scatterplot reveals a small negative association between vehicle age and hit rates, which is quite different than the substantial positive association between vehicle age and search rates. This discrepancy is most noticeable when comparing search and hit rates for vehicles aged 0 and 15+, the newest and oldest types of cars. Despite having a search rate ~25% lower than the average, brand-new vehicles are about 1.35 times more likely to carry contraband than the average vehicle. At the other end of the spectrum, cars 15+ years old are searched about 1.85 times more than average, which is consistent with the results from previous sections. The holistic regression model serves as a reminder that driving an old car has a major predictive effect on search rates second only to the driver race/sex combination of black male. For drivers of old cars, this increased search likelihood is not based on a corresponding level of criminality. The hit rate scatterplot supports my conjecture that law enforcement officers are suspicious of old vehicles. This suspicion is also unjustified, as drivers of these aging cars are found with contraband at the same rate as the average driver. Side-by-side presentation of the vehicle age classification search and hit rate graphs visually simplifies this conclusion:





The final relevant figures demonstrate the overall racial and gender bias in Texas Highway Patrol traffic stop searches from 2011-2016. In order to reach this conclusion, I had to examine the possible association between vehicle and driver-based characteristics and consider the argument that for certain groups, above-average search rates could be the result of above-average criminality. Figure 1 revealed clear search rate *disparities* between certain driver race/sex combinations, but I wanted to understand if this was due to implicit *discrimination*. The regression models contradicted my hypothesis that vehicle-based characteristics would diminish the effect of driver demographics on search rates. Vehicle age appeared as a major predictor of search rates while neither inflating nor deflating the predictive values of race/sex combinations.

Additionally, some driver race/sex combinations seemed to exacerbate the search rate disparities for specific vehicle types. For these reasons, I am confident that significant differences in the correspondence of search and hit rates for driver race/sex would represent implicit biases. According to my findings, this is the case for black males, Hispanic males and white females:



Discussion

When I formulated my research questions for this observational study, I was primarily interested to investigate the extent to which vehicle-based characteristics would account for race and gender disparities in traffic stop search rates. I hoped to fill a gap in the literature, but my principal goal of identifying potential racial and gender bias was similar to many other previous studies. However, as I gathered my initial results, it became clear that I would have to analyze unanticipated and complex questions. First, once I uncovered the disparities predicted in my hypothesis, I had to assess whether or not they demonstrated police bias. This process took several steps to complete, and the first major result was that the vehicle age and driver race/sex variables did not covary in any meaningful way. This was an essential piece to my conclusion that there was some observable bias because it eliminated an omitted variable problem that could have clouded my interpretation of the search rates for these two variables. Next, I demonstrated that the most of the incongruence between search and hit rates for black males, Hispanic males, and white females persisted even as additional vehicle-based variables were accounted for. In other words, the racial and gender biases continued to manifest themselves throughout the subgrouping of all relevant variables. Finally, I observed indicators of vehicle age bias during the subclassification process, so I then compared search and hit rates for the vehicle age variable alone, both by individual year and age classification. Both of these comparisons pinpointed a lack of correspondence between search and hit rates for the vehicle age variable that I previously suspected. Completing these steps provided the necessary context for a correct interpretation of the results. By looking beyond the basic disparities and eliminating alternative explanations, I can support the conclusion that for the Texas Highway Patrol from the years 2011-2016, implicit biases regarding black males, Hispanic males, white females, new cars, and old cars influenced the decisions of whether a post-stop search was warranted.

In this paper, I presented a full regression model that calculated the overall predictive value of driver and vehicle-based variables. In the case of the driver race/sex and vehicle age variables, some specific values were shown to have significant effects on search rates. However, both levels of the vehicle perception variable, luxury and cheap, had very little total influence. I was surprised by this finding; it contradicted an element of my hypothesis that expected luxury cars to be searched less often than cheap cars. Because I doubted the fact that vehicle perception played no role in determining search probability, I decided to create a progression of figures that showed the partial effects of the three variables on 24 distinct subgroups of driver and vehicle characteristics. The results of this particular analysis further supported the presence of racial and gender bias and also demonstrated that in some specific cases, the specification of whether a vehicle is luxury or cheap can have large effects on search rates. This was especially true for black and Hispanic males. There are no clear explanations for why the search rates of black and Hispanic males driving an old car spike when the vehicle is a luxury brand, but this is most likely to due to the fulfillment of a stereotype that associates these subgroups with higher criminality.

While the results of this analysis were quite compelling, further research will be needed to test these conclusions due to limitations in the dataset and the overall lack of research that considers these driver and vehicle-based variables in tandem. The database lacked information concerning both the stop purpose and driver age, both of which are variables that have been studied in relation to this subject. Furthermore, for the sake of simplicity, I did not examine the effects of county or time of day in this analysis. With an already large quantity of figures and a total of 24 subgroups in my most in-depth graph, I did not want to unnecessarily complicate my findings. The exclusion of the county variable was also based on an assumption that it would have a negligible effect, considering that the stops were conducted by the Highway Patrol rather than a municipal department patrolling different neighborhoods. The exclusion of the time of day variable was also based on an assumption that no specific type of driver or vehicle is disproportionately represented on the road at certain times. Future researchers should test these assumptions and conduct similar analyses on datasets from other states and departments to determine the external validity of my conclusions.

Conclusion

The purpose of this paper was to gain insight on potential police bias with respect to both driver and vehicle-based variables. A database with more than 12 million observations from the Texas Highway Patrol allowed me to examine a variety of relationships and obtain highly-specified results unseen in previous studies. My original hypothesis contained three separate parts. First, I expected luxury cars to be searched less than cheap cars. Second, I expected new cars to be searches less than old cars. Finally, I expected these two vehicle-based variables to coincide with the effect of driver race and gender on search probability. I found no evidence to support the first element, as luxury and cheap cars were searched at similar rates. Furthermore, vehicle perception had very little total predictive effect on search rates. However, as I created more subgroups that included additional variables, I found that vehicle perception can have a large influence on search rates for specific combinations of driver race/sex and vehicle age. The second element of my hypothesis was confirmed by a number of figures and the regression models. New cars were searched less often than old ones, but this was a result of an unjustified bias. In fact, new cars were more likely to be found with contraband than older ones. Additionally, the analysis uncovered an amazing correlation between vehicle age and search rates when vehicle age was presented in yearly increments. Finally, the third element of my hypothesis was disproven despite my strong belief that it would be confirmed. The regression models revealed that vehicle and driver race/sex are both strong predictors of search probability, but that they influence officer decisions independent of each other. A main finding of this analysis is that drivers of old cars are suspected of criminality at similar levels to black male drivers, a finding established through the comparison of search and hit rates.

References

Anderson, Hannah. “It’s Time To Start Collecting Stop Data: A Case For Comprehensive Statewide Legislation.” *The Policing Project*, 30 September 2019.” Retrieved from https://www.policingproject.org/news-main/2019/9/27/its-time-to-start-collecting-stop-data-a-case-for-comprehensive-statewide-legislation

Harris, David. “Driving While Black: Racial Profiling On Our Nation’s Highways.” *American Civil Liberties Union,* June 1999. Retrieved from https://www.aclu.org/report/driving-while-black-racial-profiling-our-nations-highways

Schafer, Joseph. “Decision making in Traffic Stop Encounters: A Multivariate Analysis of Police Behavior.” *Police Quarterly,* January 2006. Retirved from https://journals.sagepub.com/doi/abs/10.1177/1098611104264990

Sides, John. “What data on 20 million traffic stops can tell us about ‘driving while black.’” *The Washington Post*, 17 July 2018. Retrieved from https://www.washingtonpost.com/news/monkey-cage/wp/2018/07/17/what-data-on-20-million-traffic-stops-can-tell-us-about-driving-while-black/

Weir, Kirsten. “Policing in black & white.” *American Psychological Association,* December 2016. Retrieved from https://www.apa.org/monitor/2016/12/cover-policing