LOGIT AND CASE-BASED ANALYSIS OF DRIVERS' AGE AS A CONTRIBUTING FACTOR FOR FATAL TRAFFIC CRASHES ON HIGHWAYS AND STATE ROADS IN FLORIDA

Bhuiyan Alam, Ph.D.*
Associate Professor
Department of Geography and Planning
The University of Toledo
Toledo, OH 43606-3390
Tel.: 419-530-7269
Fax: 419-530-7919
E-mail address: bhuiyan.alam@utoledo.edu

Lisa Spainhour, Ph.D., PE
Professor
Department of Civil and Environmental Engineering
FAMU-FSU College of Engineering
Florida State University
Tallahassee, FL 32310
E-mail address: spainhou@eng.fsu.edu

*Corresponding author.

WORD COUNT
Text, acknowledgement and references = 4,120 words
2 table @ 250 words per table = 500 word equivalent
3 figures @ 250 words per figure = 750 word equivalent
Sub-total = 5,370 words
Abstract = 237 words
Total = 5,607 words
This paper consists of two steps. First, it investigates the age and sex distribution of at-fault drivers of fatal crashes on the interstate and state highways in Florida using descriptive statistics. Second, it employs a binary logit model to test the statistical significance of the findings in the first step. The first step explores that the younger (=<24 years) and older (65-74, and >=75 years) drivers of both sexes are more likely to cause fatal crashes compared to the ‘average’ middle-aged drivers. The female older (>=75 years) drivers are at highest level of vulnerability of causing fatal crashes followed by younger male drivers. The logit analysis shows that younger drivers are more likely to cause fatal crashes than drivers of 25-34, 35-44, 45-54, 55-64, and 65-74 age cohorts. On the other hand, it also reveals that older drivers have more likelihood of causing fatal crashes than younger drivers. The logit analysis does not find either of the sex groups to be more likely to cause fatal crashes than the other. The study recommends that frequent public transit services should be provided to the older drivers. It further proposes that effective measures should be taken to make current training and educational programs more effective if in place, and more such programs should be initiated if not in place. The study finds logit model as a useful tool to confirm or disconfirm the results of case-based and descriptive analysis.
INTRODUCTION

Nearly 40,000 drivers and passengers are killed in traffic crashes in the United States every year. The rate of such crashes is also high in the state of Florida; with nearly 40 percent more fatalities per vehicle miles travelled (VMT) than the national average (1). Different levels of exposures to driving by people of different age and sex categories are responsible for a significant proportion of fatal crashes in the manner they occur (2, 3, 4, 5, 6). Although several studies have investigated the impacts of age and sex on risks of being involved in fatal crashes (7, 8, 9, 10, 11, 12, 13, 14), many have analyzed number of crashes caused by each age and sex categories per unit miles driven (15, 16, 17, 18), and yet others have considered crash victims of all ages as a single group. Different age groups need different attention to lower the crash rates by each age group (8). Both younger and older drivers show sign of being responsible for a large majority of fatal crashes in the United States than “average” drivers (19, 20, 21, 22, 23, 24, 25). Researchers are attracted to analyze these two groups for different reasons as drivers of these groups contribute to high fatality rates for different reasons. In case of gender, fewer women than men die in fatal car crashes (11, 26, 27, 28).

The proportion of people aged 65 or older increased in the United States by nearly ten percent in last two decades compared to six percent of people younger than 65. This trend is projected to continue through the year 2030 when the share of ≥65 year old citizens will constitute approximately 22% of total population (29). As a result, the proportion of older drivers on the nation’s highways will also increase and it is very likely that they will contribute to a larger proportion of crashes. Therefore, several researchers have studied older drivers’ crash involvement characteristics and explored that the older drivers, as a group, have one of the highest crash rates per VMT (4, 7, 29, 30, 31). Researchers have focused on different aspects of older drivers’ physical and mental conditions. While some have focused on chronic medical conditions like hypoglycemic attacks (32), others have analyzed functional impairment like vision, cognition, confusion, sudden loss of control, and mobility (24, 33, 34).

Similarly, younger drivers have attracted attention of many researchers partly due to their highest rate of involvement in fatal crashes. Although debate continues to focus on the reasons for such crash rates by this age cohort, most blame for two major causes – willingness to take risk and inexperience in conceiving the traffic conditions (3, 35, 36, 37, 38, 39, 40, 41). Speeding and driving under the influence of alcohol are two major specific factors usually found to be responsible for fatal crashes caused by the younger drivers (25, 42). Driving without a license has also been shown to be an important factor for fatal crashes caused by young drivers (23, 43).

In traffic safety study, it is common to analyze at-fault drivers’ proportion normalized by some measurement units. Such normalizing measurement units could be number of total population in each age or sex category (11, 29, 44), number of licensed drivers in each age or sex category, and VMT (7). Another approach of investigating crash data by different age and sex groups is by environmental conditions like time of the day, day of the week, wet/dry road, snowed/icy roads, presence of animals on roads, poor visibility, etc. (8, 18), and yet another measurement unit is driving violations (18). Researchers have used different datasets to analyze traffic data. Fatality Analysis Reporting Systems (FARS), National Personal Transportation Survey (NPTS), Crash Analysis Reporting Environment (CARE) are three databases – to name few.

The objectives of this paper are twofold. First, investigate the age and sex distribution of at-fault drivers of fatal crashes on the interstate and state highways of Florida in 2000 using descriptive statistics like RFs. Second, explore the likelihood of drivers in different age cohorts
of causing (hence, to be at fault in causing) fatal crashes compared to the reference group, younger drivers, i.e., <=24–year old. Using a binary logit model, the second objective, in turn, verifies the results of RFs. The binary logit model is also used to analyze the female drivers’ likelihood of causing fatal crashes compared to male drivers.

DATASET AND METHODOLOGY
The original dataset included a total of 2,082 fatal crashes in Florida in 2000. This included 3,825 drivers of which 1,874 were at-fault, 1,935 were not-at-fault, and the rest 16 were unknown. Among these 3,825 drivers, a total of 1,913 drivers were driving automobiles, 282 passenger vans, 599 pick-up/light truck, 79 medium truck, 161 heavy truck, 475 truck tractor (cab), and the rest 316 drivers were driving motor home, bus, bicycle, motorcycle, moped, and other types of vehicle. The crashes included one vehicle, two vehicles, and multi vehicles. This paper deals with all crashes irrespective of the type and number of vehicles involved. Among 1,874 at-fault drivers the ages and/or other information of 134 drivers were unknown. These drivers have been left out of the study keeping the rest 1,740 at-fault drivers in the database. Similarly, out of 1,935 not-at-fault drivers, the ages and/or other information of 170 drivers were unknown leaving a toll of 1,765 not-at-fault drivers in the dataset. The median age of the drivers was 38 years.

The research team entered data on fatal crashes from paper and computer resources to develop an electronic database. The developed database goes beyond the data then currently available from the Florida Traffic Crash Report (FTCR), incorporating information available from Traffic Homicide Investigation (THI) reports, i.e., so-called police reports. The database included 17 major data categories, including crash, roadway, vehicle, driver, passenger, pedestrian, truck, environment, and violations data. It also included factors contributing to the crashes and fatalities.

A diverse team of traffic safety specialists were trained to study, analyze, and reconstruct crashes on a case-by-case basis before starting data entry and analysis of overall trends. The research team, consisting of experts in homicide investigation, traffic and safety engineers, traffic crash researchers, and crash reconstructionists, scrutinized the available data in each crash. The THI reports indicated who at-fault driver(s) was/were in each crash. However, the research team did not rely only on the THI reports since it is widely believed that such reports could be seriously biased against the younger drivers. The research team identified the at-fault driver(s) in each crash after thoroughly studying and analyzing the case, i.e., based on the detailed investigation of photographic evidence, officer and witness statements, posted speed limits, actual vehicle speeds/positions/travel lanes, watching video logs, etc. All data entry and analysis was supervised by the team leader, which helped to establish and maintain inter-rater reliability and ensure the high quality of the dataset.

To achieve the first objective stated earlier, descriptive statistical techniques like RFs and bar charts were used to analyze the data. When necessary, the age distribution was expressed in terms of percentages. The inferences for the highs and lows were drawn based on the investigations of these specific cases. While Chi-squared tests measure the association between two categories of variables, the RFs show the weight of one specific group of object compared to the ‘average’ group. The RFs have been used in this research to find out the vulnerability of different age groups for fatal crashes caused by different contributing factors compared to the vulnerability of an ‘average’ driver. The RFs are different than Odds Ratio (OR) and Relative
Risk (RR) in that while the ORs and RR measure association in 2 X 2 tables, the RFs are used for any number of columns and rows \((24)\).

A binary logit model, as shown by Equation 1, was employed with dependent variable “at-fault” (i.e., whether a driver was at fault) and independent variables “age distribution” and “sex.” The dependent variable could assume only two values: 1 indicating that the driver was at fault and 0 indicating otherwise. The independent variable “age distribution” is categorical in nature and has seven age cohorts: \(<24, 25-34, 35-44, 45-54, 55-64, 65-74, \text{ and } \geq 75\) years. The age cohort \(<24\) years is used as the reference category. The other explanatory variable “sex” is binary in nature: 1 indicating a male driver and 2 indicating a female driver. The male category is considered as the reference group for this study.

\[
\text{logit } (Y) = \ln(\text{odds}) = \ln \left( \frac{\Pi}{1-\Pi} \right) = a + b_1X_1 + b_2X_2 \quad \cdots \cdots \cdots \cdots \quad (1)
\]

If we take the antilog of both sides of Equation 1, we get the equation for predicting the probability of occurrence of the expected outcome which can be shown by Equation 2 below:

\[
\Pi = \text{Probability}(Y=\text{outcome of an event } \mid X= x, \text{ a value of X})
\]

\[
\frac{\exp(a + b_1X_1 + b_2X_2)}{1 + \exp(a + b_1X_1 + b_2X_2)} \quad \cdots \cdots \cdots \cdots \quad (2)
\]

Where, \(\Pi\) is the probability of occurrence of an event; hence, whether a driver was at fault, \(a\) represents the \(Y\)-intercept, \(b_1\) and \(b_2\) are regression coefficients, and \(\exp\) is the base of natural logarithm, which equals 2.71828. Typically, \(b_1\) and \(b_2\) values are estimated employing maximum likelihood method. Equation 1 shows a linear relationship between \(\text{natural } \log(\text{odds})\) and \(X\). For this study, \(X_1\) represents drivers’ age while \(X_2\) represents drivers’ sex. The null hypothesis for this study would be to state that \(b_1\) and \(b_2\) values equal zero, meaning that there is no linear relationship between predictor variables age and sex with logit of the dependent variable —whether a driver was at fault.

RESULTS AND DISCUSSION

Figure 1 explores that the proportion of at-fault drivers is highest in the \(<24\)-year age cohort and gradually decreases until 65-74-year when it increases for \(\geq 75\)-year cohort. The figure indicates that the age distribution of at-fault drivers takes a shape similar to a ‘U’, indicating that the proportions of at-fault younger and older drivers are typically higher than that of some middle-aged drivers. Given the wide range of middle-aged groups’ age band \((25 – 74 = 50\) years), Figure 1 shows that the proportion of at-fault middle-aged drivers is relatively less than that of the younger and older drivers: 34.7% in two cohorts \((<24\) and \(\geq 75\)) vs. 65.3% in five cohorts.

Three-fourth of the at-fault drivers in Florida are males while the rest are female. Figure 2 shows the distribution of fatal crashes by age and sex categories. It explores that the distribution of sex among different age categories are fairly uniformly distributed. None of the age categories has peak bar for any of the sexes. This implies that the male and female drivers are approximately equally distributed over the age groups. However, Figure 2 shows that there are
slightly more younger male drivers compared to younger female drivers while little more older female drivers compared to older male drivers.

FIGURE 1  Age distribution of at-fault drivers in different age groups.

FIGURE 2  Distribution of crashes by sex and age categories.
Figure 3 shows the RFs of the at-fault drivers based on age and sex categories. It explores that both younger and older drivers of both sexes are overrepresented in the data sets with RF values 2.0 and 1.8 for male drivers of $\leq 24$ and $\geq 75$ years while with RF values 1.9, 1.5, and 2.6, respectively, for female drivers of age groups $\leq 24$, 65-74, and $\geq 75$ years. These are shown by the ‘U’ shaped curves for both the male and female drivers. The figure tells us that younger and older drivers cause more fatal crashes in Florida highways compared to the ‘average’ drivers irrespective of the sex of the drivers. Scrutinizing more, Figure 3 explores that the younger male drivers are slightly more likely to cause fatal crashes compared to the younger female drivers with RF values 2.0 and 1.9, respectively, for $\leq 24$ years, and that the older female drivers are more vulnerable to cause fatal crashes compared to the older male drivers with RF values 1.5 vs. 1.0 for 65-74 age group, and 2.6 vs. 1.8, for age group $\geq 75$ years.

The statistical package for the social sciences (SPSS v. 17) was used for logit analysis. The Hosmer-Lemeshow Test investigates the null hypothesis that there is no linear relationship between the explanatory variables and the log odds of the dependent variable. The SPSS output for Hosmer-Lemeshow Test of this study shows a Chi-square value of 4.284 with 6 degrees of freedom and a $p$ value of 0.638. This non-significant Chi-square value reveals that the data fit the model well. Similarly, the Omnibus Tests of Model Coefficients output of SPSS shows a Chi-square value 159.091 with a $p$ value of 0.000 indicating that the model used in this study is significantly better than a model with only the intercept. The Classification Table (Table 1) shows the sensitivity, specificity, and overall success rate of the model in correctly predicting log odds of the dependent variable. The proportion of occurrences correctly predicted is called the sensitivity of prediction while the proportion of non-occurrences correctly predicted is known as the specificity of prediction. Table 1 shows that the sensitivity of the binary logit model of this study is 39.5% ($=688/(1052+688)$) while the specificity is 77.8% ($=1374/(1374+391)$). Overall the model correctly predicts 2062 out of 3505 cases with an overall success rate of 58.8%.

**FIGURE 3** Risk factors of different age categories for sex.
Table 2 shows the logistic regression predicting the status of at fault from age and sex. The odds ratio for the drivers of each age cohort compares the likelihood of a driver in a specific cohort of being at fault in a fatal crash compared to the drivers of the reference group =<24 years. Similarly, the odds ratio for female drivers indicates the likelihood of a female driver's to be at fault with respect to the reference group – male drivers. The table explores that drivers of age cohorts 25-34, 35-44, 45-54, 55-64 and 65-74 are less likely (indicated by the associated –b values) to be at fault in fatal crashes compared to a driver of age cohort =<24. In other words, the drivers of =<24 years cohort are more likely to be at fault in fatal crashes compared to the drivers of ages 25-74 years. On the other hand, the older drivers of age cohort >=75 years are 1.661 times more likely (indicated by the associated +b value) to be at fault in fatal crashes than that of =<24-year cohort. This finding confirms the results of case-based analysis of crashes discussed earlier. However, the logit analysis shows that the female drivers are not more or less likely to be at fault in fatal crashes with respect to their male counterpart. This implies that sex does not play any statistically significant role in causing fatal crashes. Table 2 shows another statistic called Wald Chi-square that tests the unique contribution of a specific explanatory variable holding other predictors constant, i.e., removing any overlaps between the predictor variables. All the Wald Chi-square values shown in Table 2 are statistically significant at 95% confidence level, except for sex.

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>Sensitivity, Specificity, and Overall Success Rate of the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted At Fault</td>
</tr>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>At Fault</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 2</th>
<th>Logistic Regression Predicting the Status of At Fault from Age and Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor</td>
<td>b</td>
</tr>
<tr>
<td>Constant</td>
<td>0.569</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.019</td>
</tr>
<tr>
<td>Age</td>
<td></td>
</tr>
<tr>
<td>25 – 34</td>
<td>-0.760</td>
</tr>
<tr>
<td>35 – 44</td>
<td>-0.796</td>
</tr>
<tr>
<td>45 – 54</td>
<td>-1.018</td>
</tr>
<tr>
<td>55 – 64</td>
<td>-0.806</td>
</tr>
<tr>
<td>65 – 74</td>
<td>-0.520</td>
</tr>
<tr>
<td>&gt;= 75</td>
<td>0.507</td>
</tr>
</tbody>
</table>

Page 8 of 12
CONCLUSIONS

The descriptive analysis explores that younger and older drivers of both sexes are more likely to cause fatal crashes compared to the ‘average’ drivers. Among these, the female older (>=75 years) drivers are at highest level of vulnerability of causing fatal crashes followed by younger male drivers. The younger drivers of both sexes and older female drivers are more vulnerable to such crashes compared to their counterparts. The logit analysis confirms the findings of descriptive analysis that younger and older drivers have more likelihood of causing fatal crashes than drivers of other age cohorts. However, the logit model shows that sex does not play any statistically significant role in contributing fatal crashes, i.e., female drivers do not have more or less likelihood of causing fatal crashes than male drivers and vice versa, which is in contrary to the findings of Travis et al. (11), NCHS (26), Baker et al. (27), and Owsley et al. (28).

Although it cannot be claimed with certainty, detail investigation of individual crash cases and the THI reports indicate that the older drivers tend to be confused on busy streets while the younger drivers are motivated by their immature attitude. However, it is not realistic that drivers of these age cohorts be prohibited to drive on the streets. Therefore, special target-oriented measures should be taken for these drivers to lessen the fatal crash rates. An ideal solution would be to provide efficient and frequent public transit services to the older drivers. The paper proposes that effective measures should be taken to make current training and educational programs more effective if in place, and more such programs should be initiated if not in place – for both younger and older drivers.

The implications of the findings of this article are enormous since the study is based on detailed case-based and logit analysis. It is expected that the results will be used to guide future design standards as well as to develop education and enforcement programs for different segments of the society differentiated by age and gender. In brief, the findings will be helpful for the transportation planners and policy makers to make informed decisions to reduce fatal crashes on our streets.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the generous financial support of the State Safety Office of the Florida Department of Transportation for the project from which this research is extracted. They are also thankful to Florida Highway Patrol for supplying the Traffic Homicide Reports. The research would not be possible without the assistance of either agency. The authors are also grateful to the anonymous reviewers for their comments and suggestions. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Florida Department of Transportation.
REFERENCES


