

INVESTIGATING THE TEMPORAL TRENDS AND RISKS IMPOSED BY DIFFERENT  
DRIVER GROUPS ON OTHER DRIVERS

by

Thomas Thanuvelil Philip

B. Tech., Mahatma Gandhi University, Kerala, India, 2012

A Masters Research Project  
presented to Ryerson University  
in partial fulfillment of the  
requirements for the degree of  
Master of Engineering  
in the Program of  
Civil Engineering

Toronto, Ontario, Canada, 2018

© Thomas Thanuvelil Philip, 2018

## AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A MRP

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions.

I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my MRP may be made electronically available to the public.

# INVESTIGATING THE TEMPORAL TRENDS AND RISKS IMPOSED BY DIFFERENT DRIVER GROUPS ON OTHER DRIVERS

Thomas Thanuvelil Philip  
Master of Engineering - 2018  
Ryerson University

## ABSTRACT

Multi-vehicle traffic collisions usually result in increased injury severities to the more vulnerable drivers involved in those accidents. This research study aims at investigating the temporal trends and risks imposed by different driver groups on other drivers using logistic regression. The study is based on analysing accident data for all light-duty two-vehicle collisions in North Carolina from January 1, 2004 to December 31, 2013. Two logistic regression models are developed for each year. The first model, evaluates the probability that a certain driver sustains at least a visible injury caused by the other driver and the second model, evaluates the probability that a driver will cause at least a visible injury to the other driver. The findings of this research may help decision makers identify driver groups that are more dangerous to other drivers so that necessary precautionary measures can be adopted to make our roads a safer place.

## ACKNOWLEDGEMENTS

I acknowledge with gratitude all those who have encouraged and guided me in making this project a reality.

I would like to thank the Highway Safety Information System (HSIS) for providing me with the collision data for the ten-year analysis period.

I wish to express my sincere gratitude to my Supervisor, Dr. Said Easa (Ryerson University) for his valuable support, encouragement and guidance extended to me for this project. I would also like to thank my Co-Supervisor, Dr. Essam Dabbour (Abu Dhabi University) for his valuable inputs and guidance extended to me for this project.

I am grateful to all my colleagues at Ryerson University and my family members for all the support and encouragement given to me.

Above all, I thank God Almighty for helping me complete the project successfully on time.

## TABLE OF CONTENTS

Author's Declaration for Electronic Submission of a MRP .....	ii
Abstract .....	iii
Acknowledgements .....	iv
List of Figures .....	vi
List of Tables .....	vii
List of Appendices .....	viii
1 INTRODUCTION .....	1
2 LITERATURE REVIEW .....	3
3 DATA COLLECTION .....	5
3.1 Collection of Collision Data .....	5
3.2 Creation of a Consolidated Database .....	6
4 METHODOLOGY .....	9
4.1 Logistic Regression Concept .....	9
4.2 Creation of Subset Databases for Logistic Regression Models .....	9
4.3 Logistic Regression Model Details and Software Used .....	11
5 RESULTS .....	14
6 DISCUSSIONS .....	25
7 CONCLUSIONS .....	27
APPENDIX .....	29
REFERENCES .....	39

## LIST OF FIGURES

<b>Figure 1</b> – Sample Screenshot from Accident File for 2004 .....	5
<b>Figure 2</b> – Sample Screenshot from Vehicle File for 2004 .....	5
<b>Figure 3</b> – Sample Screenshot from Consolidated Database for 2004 .....	6
<b>Figure 4</b> - Driver gender details identified in HSIS database.....	6
<b>Figure 5</b> - Vehicle Types Identified in HSIS database .....	7
<b>Figure 6</b> - Attributes related with driver's physical condition identified in HSIS database .....	8
<b>Figure 7</b> - Driver injury details identified in HSIS database .....	8
<b>Figure 8</b> – Flow Chart for the developed Logistic Regression Models .....	10
<b>Figure 9</b> – Sample Screenshot from Subset File (Driver Age – 2004).....	11
<b>Figure 10</b> – Screenshot of Logistic Regression Calculator (Pezzullo 2015).....	11
<b>Figure 11</b> – Inputs for Logistic Regression Calculator (Pezzullo 2015).....	12
<b>Figure 12</b> – Total Collisions from 2004 to 2013.....	14
<b>Figure 13</b> – Two Vehicle Collisions vs Other Collisions (2004 to 2013) .....	14
<b>Figure 14</b> – Temporal Trend for Driver Impairment – Models a and b .....	19
<b>Figure 15</b> – Temporal Trend for Driver Gender – Models a and b.....	20
<b>Figure 16</b> – Temporal Trend for Vehicle Age – Models a and b .....	22
<b>Figure 17</b> – Temporal Trend for Driver Age – Models a and b.....	23

## LIST OF TABLES

<b><i>Table 1</i></b> - Number (and Percentage) of collisions based on number of vehicles involved.....	15
<b><i>Table 2</i></b> - Coefficients (and Standard Errors) of developed Logistic Regression Models.....	16
<b><i>Table 3</i></b> - Odds Ratio (Upper and Lower 95% CI) .....	17
<b><i>Table 4</i></b> - Average Value of Coefficient and Odds Ratio (2004 – 2013) for each Explanatory Variable.....	18

## LIST OF APPENDICES

<b>Figure 18</b> - Results for Year 2004 obtained from Logistic Regression Calculator .....	29
<b>Figure 19</b> - Results for Year 2005 obtained from Logistic Regression Calculator .....	30
<b>Figure 20</b> - Results for Year 2006 obtained from Logistic Regression Calculator .....	31
<b>Figure 21</b> - Results for Year 2007 obtained from Logistic Regression Calculator .....	32
<b>Figure 22</b> - Results for Year 2008 obtained from Logistic Regression Calculator .....	33
<b>Figure 23</b> - Results for Year 2009 obtained from Logistic Regression Calculator .....	34
<b>Figure 24</b> - Results for Year 2010 obtained from Logistic Regression Calculator .....	35
<b>Figure 25</b> - Results for Year 2011 obtained from Logistic Regression Calculator .....	36
<b>Figure 26</b> - Results for Year 2012 obtained from Logistic Regression Calculator .....	37
<b>Figure 27</b> - Results for Year 2013 obtained from Logistic Regression Calculator .....	38



# 1 INTRODUCTION

Traffic collisions result in approximately 1.3 million deaths every year (World Health Organization 2015). For collisions involving two different types of vehicles, such as collisions between heavy vehicles and motorcycles, certain driver groups (in this case, the motorcyclist) are at more risk of being severely injured because of those collisions. However, there is lack of quantified measures that help us estimate the risk imposed by different driver groups on other drivers, such as the risk imposed by impaired drivers on other drivers. This is where the concept of risk externality comes into play. When an individual (say Y) performs an activity which affects the welfare of another individual (say Z) where the person performing the activity (Y) does not consider the impacts of his activity on the other individual (Z), it can be termed as an externality. (Elvik 2008). In terms of transportation safety, the concept of risk externality can translate to the risk imposed by impaired drivers on other drivers, risks imposed by aged drivers on other drivers etc.

This study provides a quantitative investigation on the level of the risk imposed by different driver groups on other drivers. The research is based on analysing the accident data for all light duty two-vehicle collisions that occurred in North Carolina from January 1, 2004 until December 31, 2013 using logistic regression. Two logistic regression models are developed for each year. The first model, evaluates the probability that a certain driver sustains at least a visible injury caused by the other driver and the second model, evaluates the probability that a driver will cause at least a visible injury to the other driver. The vehicle classes considered as light vehicles are passenger cars, pickup trucks, light trucks, panel trucks, sports utility vehicles, and vans. To eliminate the random effect of the number of passengers in the vehicle, only those light duty two-vehicle collisions where driver was the only occupant in the vehicle was considered for the analysis. The various factors identified (also known as explanatory variables) for the analysis include the driver's age, the age of the vehicle, the gender of the driver and the impairment level of the driver. Out of the explanatory variables considered, driver's age and vehicle age are continuous whereas driver gender and driver impairment level is dichotomous. The odds ratio (OR) and coefficients are used to interpret the significance of the different explanatory variables considered. Also, the 95%

confidence interval (CI) is utilised to define the upper and lower limit values of the odds ratio with 0.05 significance level.

The second part of the analysis focusses on looking at the temporal trends of the above-mentioned factors for the period of 10 years so as to identify the most significantly-stable factors. The identification of significantly-stable factors could also help decision makers identify those driver groups that increase the severity of injuries of other drivers and this in turn can help in increasing the awareness among people. It can also help in identifying the precautionary measures which has to be taken so as to make our roads a safer place for all its users.

## 2 LITERATURE REVIEW

Various researchers around the world have tried to examine risk externalities by analysing traffic accident data with the help of logistic regression methods. Details pertaining to few of those studies are summarised in this section.

Yau (2004) conducted a study to identify the various factors that can contribute to higher severity of single vehicle collisions in Hong Kong during the years 1999 and 2000. Factors pertaining to three types of vehicles, namely private vehicles, goods vehicles and motorcycles were identified using logistic regression models. Various factors such as vehicle age, gender of the driver, roadway lighting conditions, and seatbelt usage were found to be major contributors to the severity of collisions.

Yan, et al. (2005) studied multi-vehicle rear-end collisions at signalized intersections in Florida for the year 2001 by using multiple logistic regression where they identified several factors that contributed to the severity of those collisions. Those factors include the number of lanes, road surface condition, driver's gender, speed limits, and alcohol/drug use.

Chang and Yeh (2006) analysed single vehicle crash data in Taiwan for the year 2000 and they were able to identify fatality risk factors for motorcyclists and non-motorcycle drivers. Based on the factors identified, they recommended several measures to reduce the fatality rates in collisions, including improving the quality of the roadway surface, proper speed management techniques, and enforcement of seatbelt use.

Harb et al. (2008) analysed freeway work-zone crashes in Florida for the years from 2002 to 2004 and found that the geometry of the road, lighting conditions, driving under the influence, age, and gender are all factors associated with work zone crashes.

Daniels et al. (2010) examined risk externalities at roundabouts using logistic regression to analyze the severity level of drivers' injuries related to 1491 collisions that occurred at 148 roundabouts in Flanders-Belgium. They found that vulnerable road user groups (pedestrians, bicyclists, moped riders and motorcyclists) are more severely affected when colliding with other vehicle types. They

also found that older road users (above the age of 45 years) are more severely injured when colliding with younger road users.

Bham et al. (2012) analysed single vehicle and multi-vehicle collisions on urban U.S. highways. A multivariate analysis was used to identify various factors that result in collisions. Five types of collisions (angular, head-on, rear-end, sideswipe – same direction, sideswipe – opposite direction) were identified. Apart from this, various factors such as wet road conditions, poor lighting conditions, driving under the influence, driver's behaviour (decision making), and geometry of the road were found to increase collision severity.

Another interesting study is the work done by Wenzel (2013), where he explored the possibilities of improved vehicle designs and their results in reducing accident fatalities. Various factors such as presence of side air bags, better alignment of light truck bumpers etc. were found to reduce the fatality rates during collisions.

Yu et al. (2014) studied the effect of microscopic traffic, weather and road geometry on specific crash types on the freeway section I-70 in Colorado, USA. With the help of the automatic vehicle identification and weather detection systems installed in the corridor, the study provided valuable insights on how intelligent transport systems can give more focus on traffic safety improvement and effective traffic management.

Dabbour (2017) analysed the effects of various factors on the severity of drivers' injuries in single-vehicle collisions involving light vehicles by using collision data from North Carolina during the period from 2007 to 2013. Factors related to the driver, the roadway, the vehicle, and the environmental conditions were identified, and temporal stability of these variables were also analysed.

This research study tries to identify driver related and vehicle related factors that result in increased injury severity for drivers in light duty two vehicle collisions. It also looks at the temporal trends of the various factors identified for a period of ten years. The data collection steps is explained in the next section.

### 3 DATA COLLECTION

#### 3.1 Collection of Collision Data

The collision data corresponding to the ten-year period (2004 – 2013) was collected from the Highway Safety Information System (HSIS). The HSIS is managed by the University of North Carolina Highway Safety Research Center under contract with the Federal Highway Administration. (University of North Carolina Highway Research Center, 2010). For each analysis year, the following two data files were obtained:

- i. Accident File – This file contains all general data pertaining to the accident such as the accident date, time, location details, weather conditions, road surface characteristics, number of vehicles involved etc.

P	Q	R	T	J	K	L	M	N
caseno	alcflag	bikeflag	acc_date	Year	numvehs	acctype	rdsurf	weather
101156429	N	N	4/5/2004	2004	1	19	1	1
101200325	N	N	5/31/2004	2004	1	17	1	1
101156435	N	N	4/5/2004	2004	1	19	1	1
101267942	N	N	8/27/2004	2004	1	18	1	1
101269813	N	N	8/29/2004	2004	1	19	2	3
101327285	N	N	11/5/2004	2004	1	19	1	1
101100381	N	N	1/25/2004	2004	1	19	5	4
101137774	N	N	3/10/2004	2004	1	19	1	1
101215712	N	N	6/20/2004	2004	1	18	1	1
101285618	N	N	9/17/2004	2004	1	2	2	3
101289669	N	N	9/22/2004	2004	1	14	1	1
101320389	N	N	10/24/2004	2004	1	19	2	2
101341651	N	N	11/21/2004	2004	1	17	1	1

*Figure 1 – Sample Screenshot from Accident File for 2004*

- ii. Vehicle File – This file contains the data pertaining to the factors we are trying to analyse such as the vehicle-manufacture year, type of vehicle, driver age, gender of the driver, impairment levels, etc.

A	B	D	L	R	S	T	U	V
caseno	vehno	vehyr	vehtype	physcond	sob_test	prsn_typ	drv_sex	drv_age
100995427	1	1999	1	1	0	1	1	1
100995427	2	1995	1	1	0	1	2	1
101048622	1	2001	1	1	0	1	2	1
101048622	2	1993	1	1	0	1	2	1
101048626	1	1994	4	1	0	1	1	1
101048626	2	1992	1	1	0	1	2	1
101048626	3	1997	2	1	0	1	1	1
101051405	1	2000	5	1	0	1	1	1
101051405	2	1997	1	1	0	1	2	1
101054936	1	1991	1	1	0	1	1	1
101056815	1	1999	10	1	0	1	1	1

*Figure 2 – Sample Screenshot from Vehicle File for 2004*

### 3.2 Creation of a Consolidated Database

Explanatory variables related to the drivers as well as those related to the colliding vehicles were considered for the analysis. The explanatory variables related to the driver were the age, the gender and impairment levels of both colliding drivers. Vehicle's ages were the explanatory variables related to the vehicles. It should be noted that the age of the driver and the age of the vehicle are both continuous variables whereas the gender of driver and impairment level are dichotomous variables. For each year, the necessary data required from both the files were compiled into one database. The case number (which is unique for each accident record) was used to link entries from both the files.

	A	B	C	D	E	F	G	H
1	CASE NO	AGE - D1	AGE - D2	GENDER - D1	GENDER - D2	IMPAIRMENT - D1	IMPAIRMENT - D2	INJURY LVL - D1
2	100995427	54	37	1	2	1	1	5
3	101048622	28	33	2	2	1	1	4
4	101051405	20	48	1	2	1	1	5
5	101057837	22	23	1	1	1	1	3
6	101058479	16	32	2	2	1	1	5
7	101060005	35	35	1	1	1	1	5
8	101061541	36	73	1	1	1	1	5
9	101061542	28	33	1	1	1	1	5
10	101061550	15	43	1	2	1	1	5
11	101061552	58	17	1	2	1	1	5
12	101061563	38	40	1	1	1	1	5
13	101065376	31	29	1	1	1	5	1

*Figure 3 – Sample Screenshot from Consolidated Database for 2004*

In the consolidated database, the coding for gender of the drivers is same as that followed by the HSIS database as shown in Figure 4.

' '	=	'-BLANK-'
'0'	=	'NOT STATED'
'1'	=	'MALE'
'2'	=	'FEMALE'
'3'	=	'NOT OCCUPIED'
'4'	=	'UNKNOWN'

*Figure 4 - Driver gender details identified in HSIS database*

Here male drivers are represented as “1” and female drivers are represented as “2”. When it comes to vehicle type, HSIS database specifies the details pertaining to over 30 types of vehicles (as seen in Figure 5).

```
01  = 'PASSENGER CAR'
02  = 'PICKUP'
03* = 'LIGHT TRUCK (MINI-VAN, PANEL) '
04* = 'SPORT UTILITY'
05  = 'VAN'
06  = 'COMMERCIAL BUS'
07  = 'SCHOOL BUS'
08  = 'ACTIVITY BUS'
09* = 'OTHER BUS'
10  = 'SINGLE UNIT TRUCK (2-AXLE, 6-TIRE) '
11  = 'SINGLE UNIT TRUCK (3 OR MORE AXLES) '
12  = 'TRUCK/TRAILER'
13  = 'TRUCK/TRACTOR'
14* = 'TRACTOR/SEMI-TRAILER'
15* = 'TRACTOR/DOULBES'
16* = 'UNKNOWN HEAVY TRUCK'
17  = 'TAXICAB'
18  = 'FARM EQUIPMENT'
19  = 'FARM TRACTOR'
20  = 'MOTORCYCLE'
21  = 'MOPED'
22  = 'MOTOR SCOOTER OR MOTOR BIKE'
23  = 'PEDALCYCLE'
24  = 'PEDESTRIAN'
25  = 'MOTOR HOME/RECREATIONAL VEHICLE'
26  = 'OTHER'
27* = 'ALL TERRAIN VEHICLE (ATV) '
28* = 'FIRETRUCK'
29  = 'EMS VEHICLE, AMBULANCE, RESCUE SQUAD'
30  = 'MILITARY'
31* = 'POLICE'
32* = 'UNKNOWN'
33** = 'NOT STATED'
34** = '2, 4 DOOR SEDAN'
35** = 'STATION WAGON (PASSENGER) '
36** = 'STATION WAGON (TRUCK) '
37** = 'TRUCK WITH FOUR AXLES'

* Categories present in 2000 and later years.
** Categories present in years prior to 2000.
```

*Figure 5 - Vehicle Types Identified in HSIS database*

As there is significant differences between vehicle and driver characteristics of heavy vehicles and those of light vehicles, only collision data pertaining to light vehicles (passenger cars, pickup trucks, light trucks, panel trucks, sports utility vehicles, and vans) were considered for the analysis. The effect of heavy vehicles and other vehicle types were not considered in this study as they have investigated in several previous studies. In addition to this, to eliminate the random effect of the number of passengers in the vehicle, only those light duty two-vehicle collisions where driver was the only occupant in the vehicle was considered for the analysis.

The HSIS database specifies various attributes related to the physical condition of the driver as shown in Figure 6.

```
01 = 'APPARENTLY NORMAL'
02 = 'ILLNESS'
03 = 'FATIGUE'
04 = 'FELL ASLEEP, FAINTED, LOSS OF CONSCIOUSNESS'
05 = 'IMPAIRMENT DUE TO MEDICATIONS, DRUGS, ALCOHOL'
06 = 'MEDICAL CONDITION'
07 = 'OTHER PHYSICAL IMPAIRMENT'
08 = 'RESTRICTION NOT COMPLIED WITH'
09 = 'OTHER'
10 = 'UNKNOWN'
```

**Figure 6** - Attributes related with driver's physical condition identified in HSIS database

When it comes to the injury level of the driver, the HSIS database identifies injury details as given in Figure 7.

```
1 = 'K KILLED'
2 = 'A TYPE INJURY (DISABLING) '
3 = 'B TYPE INJURY (EVIDENT) '
4 = 'C TYPE INJURY (POSSIBLE) '
5 = 'O NO INJURY'
6 = 'N UNKNOWN'
7 = 'NOT STATED'
8 = 'NOT OCCUPIED'
```

**Figure 7** - Driver injury details identified in HSIS database



## 4 METHODOLOGY

### 4.1 Logistic Regression Concept

The risk imposed by the driver groups is found out using logistic regression, which is a generalized linear model where the occurrence of an event is predicted by fitting the data to a logit function. The event is coded as a dichotomous outcome variable and its value is dependent on the various explanatory variables (Dabbour 2017). The logit function takes the general form:

$$f(z) = e^z / (1 + e^z)$$

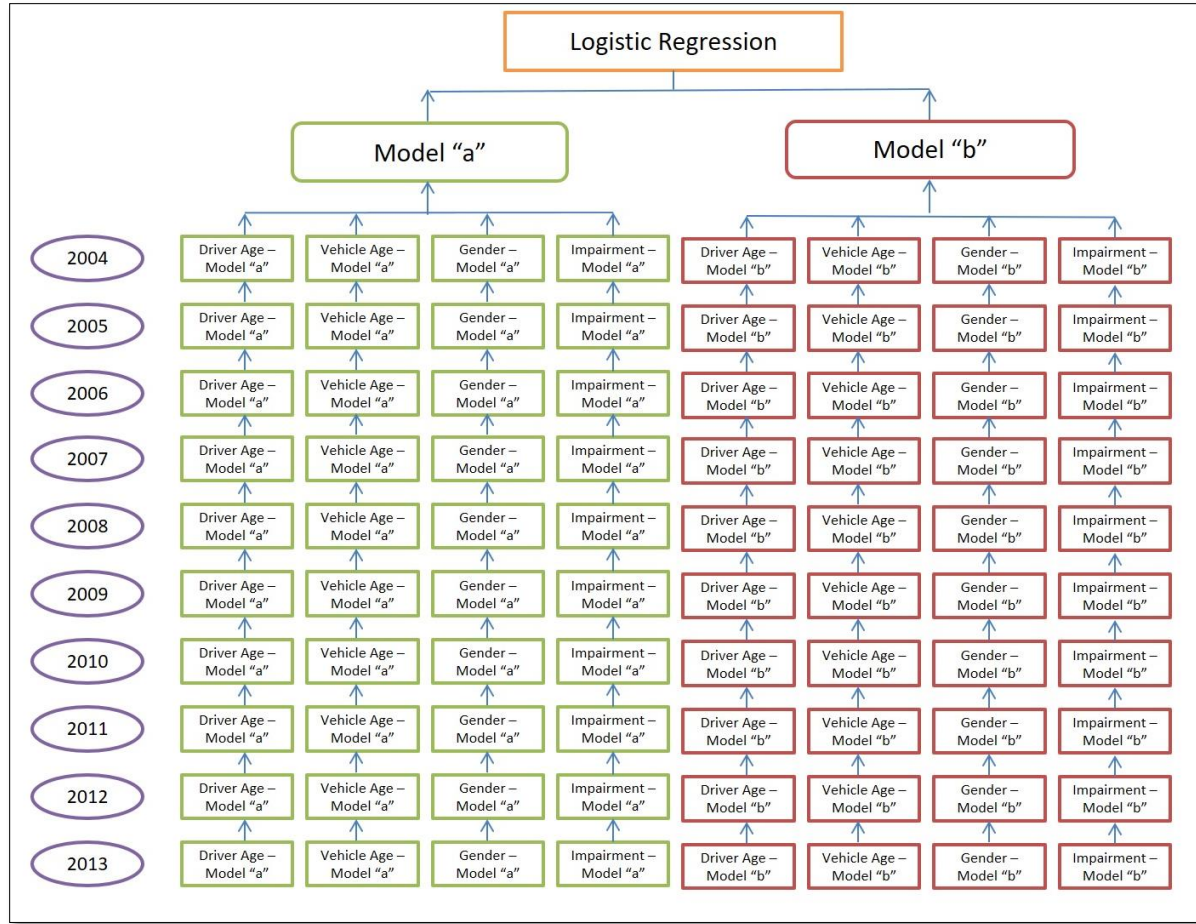
Where  $z$  denotes the logit function, and  $f(z)$  represents the dichotomous variable which is assumed to follow Bernoulli distribution. This represents the probability that the driver is killed or seriously injured provided that the collision has occurred and has resulted in an injury to the driver. It can take a value of “1” if the driver injury is fatal or serious and “0” if it is a minor injury or no injury. The corresponding logit function can be represented as:

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Where  $x_1, x_2, \dots, x_k$  represent the explanatory variables,  $\beta_1, \beta_2, \dots, \beta_k$  represent the regression coefficients and  $\beta_0$  is the intercept.

### 4.2 Creation of Subset Databases for Logistic Regression Models

Figure 8 shows the flow chart of the two logistic regression models (annotated as model “a” and model “b”) that was developed for the project. As discussed earlier, we have four explanatory variables which we are trying to analyse (driver age, vehicle age, driver gender and driver impairment). For each of these variables, we are looking at two aspects (or in other words developing two logistic regression models). A first model (“a”), which evaluates the probability that a certain driver sustains at least a visible injury caused by the other driver and a second model (“b”), which evaluates the probability that a driver will cause at least a visible injury to the other driver.



**Figure 8 – Flow Chart for the developed Logistic Regression Models**

As shown in figure 8, subset databases were created from the consolidated databases for each of the explanatory variables. For each of the subset database, the injury level of the driver is the output variable. Injury levels 1, 2 and 3 (in the HSIS database – Figure 7) is coded as “1” which represents that the driver was at least visibly injured and the rest was coded as “0”. The gender of the driver is coded “1” for female and “0” for male. The impairment level is coded “1” for impaired (which represents the impairment due to medications, drugs, alcohol) and “0” for not impaired. Before the analysis was carried out, those records which have incomplete/irrelevant data were eliminated. For the analysis, 40 subset files were created for the 10 year analysis period.

CASE NO	AGE - D1	AGE - D2	INJURY LVL - D1	Dummy 1	INJURY LVL - D2	Dummy 2
100995427	54	37	5	0	5	0
101048622	28	33	4	0	5	0
101051405	20	48	5	0	5	0
101057837	22	23	3	1	5	0
101058479	16	32	5	0	5	0
101060005	35	35	5	0	4	0
101061541	36	73	5	0	5	0
101061542	28	33	5	0	5	0

**Figure 9** – Sample Screenshot from Subset File (Driver Age – 2004)

### 4.3 Logistic Regression Model Details and Software Used

As mentioned above, two logistic regression models were developed for each year from the subset files. An online logistic regression calculator (as seen in Figure 10) was used for the analysis.

**Logistic Regression**  
 by John C. Pezzullo  
 Revised 2015-07-22: Apply fractional shifts for the first few iterations, to increase robustness for ill-conditioned data.

This page performs logistic regression, in which a dichotomous outcome is predicted by one or more variables. The program generates the coefficients of a prediction formula (and standard errors of estimate and significance levels), and odds ratios (with confidence intervals).

**Instructions:**

For detailed examples by Kevin M. Sullivan, [click here](#).

1. Enter the number of data points:  (or, if summary data, the number of lines of data).
2. Enter the number of predictor variables:
3. Enter the confidence level:  %
4. If you're entering [summary data](#), check here ☐
5. Type or paste data in the window below.  
 Predictor variable(s) first, then outcome variable (1 if event occurred, 0 if it did not occur).  
 If summary data box checked (Step 4), enter outcome as 2 columns: # of non-occurrences, then # of occurrences.  
 Columns must be separated by commas or tabs.  
 See [Kevin Sullivan's page](#) for more examples of how to enter data.

5. Click the [Solve](#) button; results will appear in the window below:

**Figure 10** – Screenshot of Logistic Regression Calculator (Pezzullo 2015)

**Instructions:**

For detailed examples by Kevin M. Sullivan, [click here](#)

1. Enter the **number of data points**:  (or, if summary data, the number of lines of data).
2. Enter the **number of predictor variables**:
3. Enter the **confidence level**:  95 %
4. If you're entering **summary data**, check here ☐
5. Type or paste data in the window below.  
Predictor variable(s) first, then outcome variable (1 if event occurred; 0 if it did not occur).  
If summary data box checked (Step 4), enter outcome as 2 columns: # of non-occurrences, then # of occurrences.  
Columns must be separated by commas or tabs.  
See [Kevin Sullivan's page](#) for more examples of how to enter data.

5. Click the  button; results will appear in the window below:

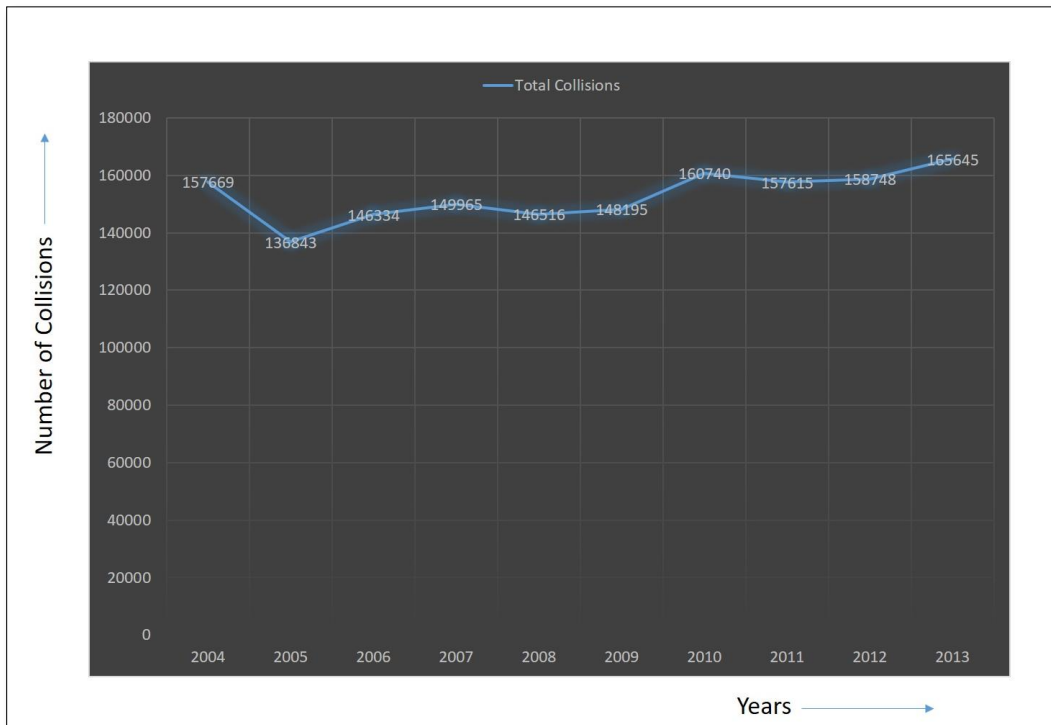
**Figure 11 – Inputs for Logistic Regression Calculator (Pezzullo 2015)**

Corresponding to each of the explanatory variable, inputs were given to the logistic regression calculator. “Number of data points” was inputted based on the number of entries (rows) in the corresponding subset file. The “number of predictor variables” was 2 as we are giving details pertaining to either driver 1 and driver 2 or vehicle 1 and vehicle 2 in the subset file. The “confidence interval” is 95% as shown and the corresponding values from the subset file were copy-pasted into the software and the analysis was carried out. (See sample output in appendix)

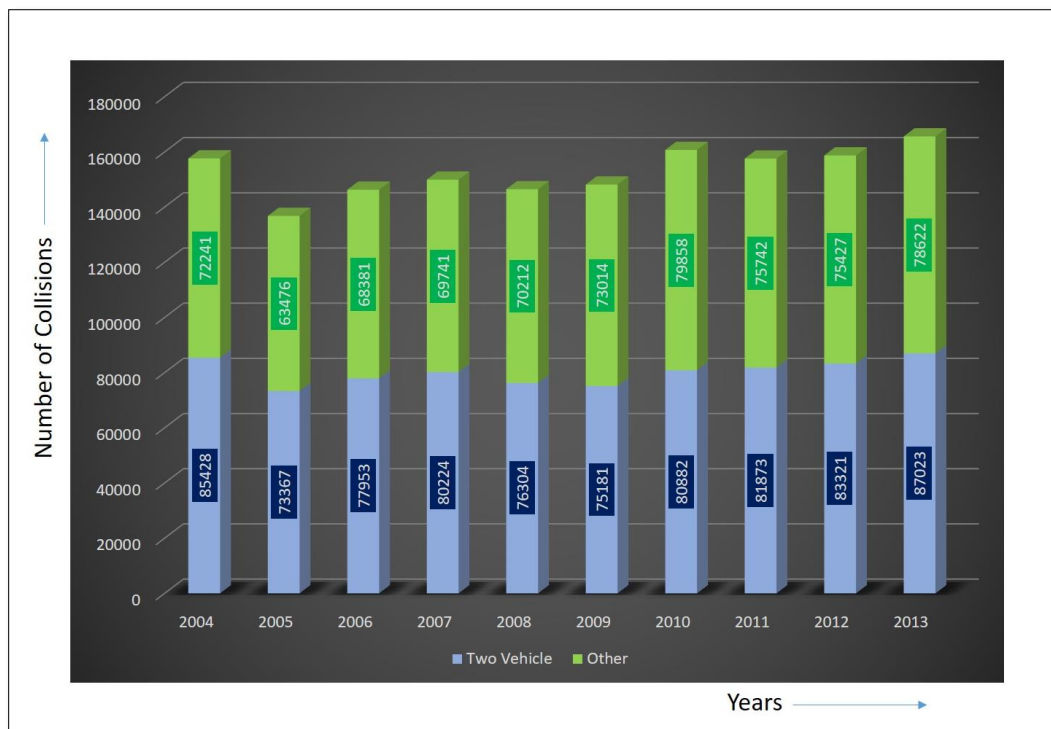
The first part of the analysis focusses on the odds ratio (OR) and coefficients which is used to interpret the significance of the different explanatory variables considered. Also, the 95% confidence interval (CI) is utilised to define the upper and lower limit values of the odds ratio with 0.05 significance level. The second part of the analysis focusses on looking at the temporal trends

of the identified explanatory variables for the period of 10 years. By looking at the temporal trend of each explanatory variable for such a long period of time, we are able to identify those explanatory variables that are consistently significant in causing injuries to the drivers in the event of an accident.

## 5 RESULTS



*Figure 12 – Total Collisions from 2004 to 2013*



*Figure 13 – Two Vehicle Collisions vs Other Collisions (2004 to 2013)*

Figure 12 shows us the year wise split up of the total number of collisions that occurred in North Carolina from 2004 – 2013. Figure 13 gives the split up of the total value and gives us the contribution of two vehicle collisions and other collisions (single vehicle or three or more vehicles). Table 1 gives the corresponding values.

**Table 1** - Number (and Percentage) of collisions based on number of vehicles involved

Year	Two Vehicle	Other	Total
2004	85428 [54.18%]	72241 [45.82%]	157669 [100%]
2005	73367 [53.61%]	63476 [46.39%]	136843 [100%]
2006	77953 [53.27%]	68381 [46.73%]	146334 [100%]
2007	80224 [53.50%]	69741 [46.50%]	149965 [100%]
2008	76304 [52.08%]	70212 [47.92%]	146516 [100%]
2009	75181 [50.73%]	73014 [49.27%]	148195 [100%]
2010	80882 [50.32%]	79858 [49.68%]	160740 [100%]
2011	81873 [51.94%]	75742 [48.06%]	157615 [100%]
2012	83321 [52.49%]	75427 [47.51%]	158748 [100%]
2013	87023 [52.54%]	78622 [47.46%]	165645 [100%]
<b>Average</b>	<b>80156 [52.45%]</b>	<b>72671 [47.55%]</b>	<b>152827 [100%]</b>

As we can see, two-vehicle accidents constitutes more than 52% of the total collisions that occur. And hence it is worthwhile trying to understand the level of risk imposed by different driver groups on other drivers in this category.

The results obtained from the developed logistic regression models for the analysis period of 2004 – 2013 is summarised in tables 2 and 3. Table 2 gives the values of coefficients (and standard errors) and Table 3 gives us the corresponding odds ratio (with higher and lower 95% confidence intervals) for each of the explanatory variables. We can see that for each of the explanatory variables, there are 2 entries. The first entry (annotated as “a”) corresponds to the first logistic regression model which evaluates the probability that a certain driver sustains at least a visible injury caused by the other driver and the second entry (annotated as “b”) corresponds to the second model which evaluates the probability that a driver will cause at least a visible injury to the other driver. All the entries which are not significant (p value greater than 0.05) were replaced with the letters “NS” (annotated as “c”).

**Table 2 – Coefficients (and Standard Errors) of developed Logistic Regression Models**

Explanatory Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Driver's age <sup>(a)</sup>	0.0042 (0.0007)	0.0051 (0.0007)	0.0049 (0.0007)	0.0054 (0.0007)	0.0061 (0.0008)	0.0048 (0.0009)	0.0045 (0.0009)	0.0064 (0.0009)	0.0053 (0.0008)	0.0081 (0.0007)
Driver's age <sup>(b)</sup>	0.0021 (0.0007)	0.0019 (0.0007)	0.0022 (0.0007)	0.0022 (0.0007)	0.0032 (0.0008)	NS <sup>(c)</sup>	0.0027 (0.0009)	0.0044 (0.0009)	0.0026 (0.0008)	0.0024 (0.0008)
Vehicle's age <sup>(a)</sup>	0.0298 (0.0019)	0.0260 (0.0019)	0.0232 (0.0018)	0.0188 (0.0016)	0.0141 (0.0015)	0.0154 (0.0017)	0.0224 (0.0019)	0.0158 (0.0015)	0.0173 (0.0014)	0.0178 (0.0014)
Vehicle's age <sup>(b)</sup>	0.0151 (0.0020)	0.0121 (0.0021)	0.0144 (0.0019)	0.0097 (0.0018)	0.0049 (0.0017)	0.0052 (0.0021)	0.0149 (0.0021)	0.0118 (0.0016)	0.0099 (0.0016)	0.0117 (0.0016)
Driver's gender <sup>(a)</sup>										
<i>Female</i>	0.1031 (0.0234)	0.0614 (0.0252)	0.0901 (0.0247)	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>
<i>Male</i> (base category)	-	-	-	-	-	-	-	-	-	-
Driver's gender <sup>(b)</sup>										
<i>Female</i>	-0.2622 (0.0240)	-0.3044 (0.0259)	-0.2693 (0.0254)	-0.3061 (0.0258)	-0.2385 (0.0268)	-0.3054 (0.0322)	-0.2855 (0.0313)	-0.2775 (0.0308)	-0.2311 (0.0270)	-0.2507 (0.0267)
<i>Male</i> (base category)	-	-	-	-	-	-	-	-	-	-
Driver's impairment <sup>(a)</sup>										
<i>Impaired</i>	1.4591 (0.0625)	1.4604 (0.0663)	1.5131 (0.0651)	1.5737 (0.0646)	1.4452 (0.0701)	1.5554 (0.0794)	1.6619 (0.0792)	1.7623 (0.0736)	1.5766 (0.0683)	1.6823 (0.0662)
<i>Not impaired</i> (base category)	-	-	-	-	-	-	-	-	-	-
Driver's impairment <sup>(b)</sup>										
<i>Impaired</i>	1.1667 (0.0687)	1.2304 (0.0715)	1.1108 (0.0745)	1.3122 (0.0703)	1.1558 (0.0774)	1.2967 (0.0865)	1.2852 (0.0897)	1.4621 (0.0810)	1.2658 (0.0759)	1.4162 (0.0723)
<i>Not impaired</i> (base category)	-	-	-	-	-	-	-	-	-	-

<sup>(a)</sup> Characteristics of the driver who is more severely injured.

<sup>(b)</sup> Characteristics of the driver and the vehicle that caused more injury to the other driver.

<sup>(c)</sup> Result not significant during the observation period (p value greater than 0.05)



**Table 3 - Odds Ratio (Upper and Lower 95% CI)**

Explanatory Variable	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Driver's age <sup>(a)</sup>	1.0042 (1.0029-1.0056)	1.0051 (1.0037-1.0065)	1.0049 (1.0035-1.0063)	1.0054 (1.0040-1.0068)	1.0062 (1.0047-1.0076)	1.0048 (1.0030-1.0066)	1.0045 (1.0028-1.0063)	1.0064 (1.0047-1.0081)	1.0053 (1.0038-1.0068)	1.0081 (1.0067-1.0096)
Driver's age <sup>(b)</sup>	1.0021 (1.0007-1.0035)	1.0019 (1.0004-1.0033)	1.0022 (1.0007-1.0036)	1.0022 (1.0008-1.0037)	1.0032 (1.0017- 1.0047)	NS <sup>(c)</sup>	1.0027 (1.0009-1.0044)	1.0044 (1.0027-1.0061)	1.0026 (1.0011-1.0041)	1.0024 (1.0009-1.0039)
Vehicle's age <sup>(a)</sup>	1.0302 (1.0265-1.0340)	1.0264 (1.0225-1.0302)	1.0234 (1.0199-1.0270)	1.0189 (1.0158-1.0221)	1.0142 (1.0112- 1.0171)	1.0155 (1.0120-1.0190)	1.0227 (1.0189-1.0265)	1.0159 (1.0128-1.0189)	1.0174 (1.0147-1.0202)	1.0179 (1.0151-1.0208)
Vehicle's age <sup>(b)</sup>	1.0153 (1.0113-1.0192)	1.0122 (1.0081-1.0163)	1.0145 (1.0107-1.0182)	1.0098 (1.0062-1.0133)	1.0049 (1.0015- 1.0083)	1.0052 (1.0011-1.0092)	1.0150 (1.0109-1.0191)	1.0119 (1.0086-1.0152)	1.0099 (1.0068-1.0130)	1.0117 (1.0086-1.0149)
Driver's gender <sup>(a)</sup>										
<i>Female</i>	1.1086 (1.0589-1.1605)	1.0634 (1.0121-1.1172)	1.0942 (1.0425-1.1486)	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>	NS <sup>(c)</sup>
<i>Male</i> (base category)	-	-	-	-	-	-	-	-	-	-
Driver's gender <sup>(b)</sup>										
<i>Female</i>	0.7693 (0.7340-0.8064)	0.7376 (0.7010-0.7761)	0.7639 (0.7269-0.8028)	0.7363 (0.7000-0.7746)	0.7878 (0.7474- 0.8303)	0.7368 (0.6917-0.7849)	0.7517 (0.7070-0.7991)	0.7577 (0.7133-0.8048)	0.7937 (0.7528-0.8368)	0.7783 (0.7387-0.8200)
<i>Male</i> (base category)	-	-	-	-	-	-	-	-	-	-
Driver's impairment <sup>(a)</sup>										
<i>Impaired</i>	4.3023 (3.8060-4.8633)	4.3077 (3.7825-4.9057)	4.5410 (3.9970-5.1591)	4.8245 (4.2511- 5.4754)	4.2428 (3.6984- 4.8674)	4.7369 (4.0539-5.5349)	5.2695 (4.5119-6.1544)	5.8258 (5.0433-6.7296)	4.8385 (4.2326-5.5311)	5.3780 (4.7238-6.1229)
<i>Not impaired</i> (base category)	-	-	-	-	-	-	-	-	-	-
Driver's impairment <sup>(b)</sup>										
<i>Impaired</i>	3.2115 (2.8071- 3.6741)	3.4225 (2.9749-3.9375)	3.0367 (2.6239-3.5144)	3.7142 (3.2358- 4.2633)	3.1767 (2.7295- 3.6971)	3.6573 (3.0869-4.3331)	3.6156 (3.0325-4.3107)	4.3152 (3.6815-5.0578)	3.5460 (3.0561-4.1144)	4.1215 (3.5773-4.7485)
<i>Not impaired</i> (base category)	-	-	-	-	-	-	-	-	-	-

<sup>(a)</sup> Characteristics of the driver who is more severely injured.

<sup>(b)</sup> Characteristics of the driver and the vehicle that caused more injury to the other driver.

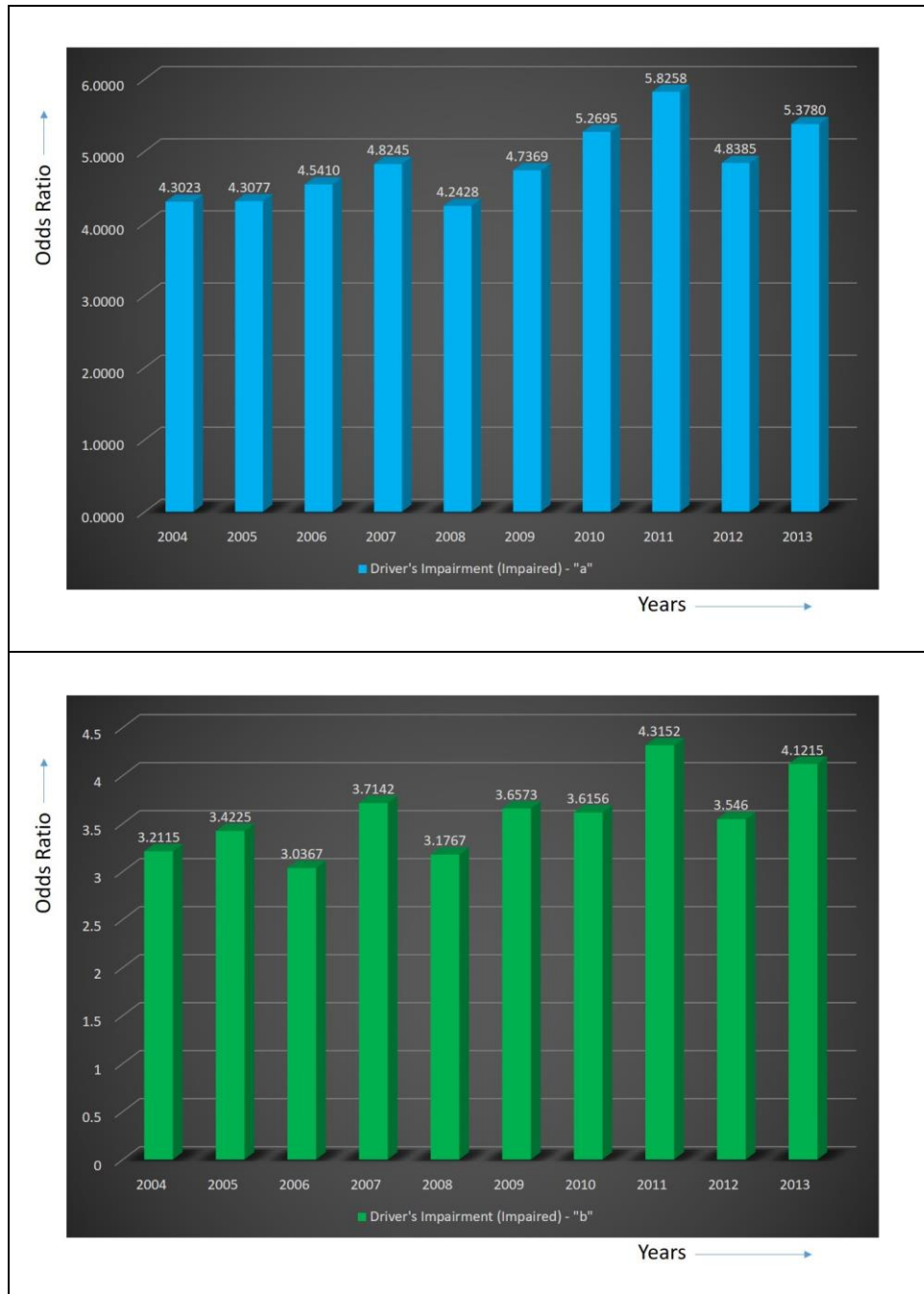
<sup>(c)</sup> Result not significant during the observation period (p value greater than 0.05)

From tables 2 and 3, we see that out of the different explanatory variables considered, impairment level has the most significant coefficient with the highest odds ratio when compared to the other explanatory variables. Driving under the influence of alcohol or illicit drugs is therefore a significant factor that increases the risk of being seriously injured and also to cause severe injuries to other drivers. Table 4 gives the average value of coefficients and odds ratio (for the 10 year analysis period) for both the logistic regression models (“a” and “b”) pertaining to each of the explanatory variables.

**Table 4 - Average Value of Coefficient and Odds Ratio (2004 – 2013) for each Explanatory Variable**

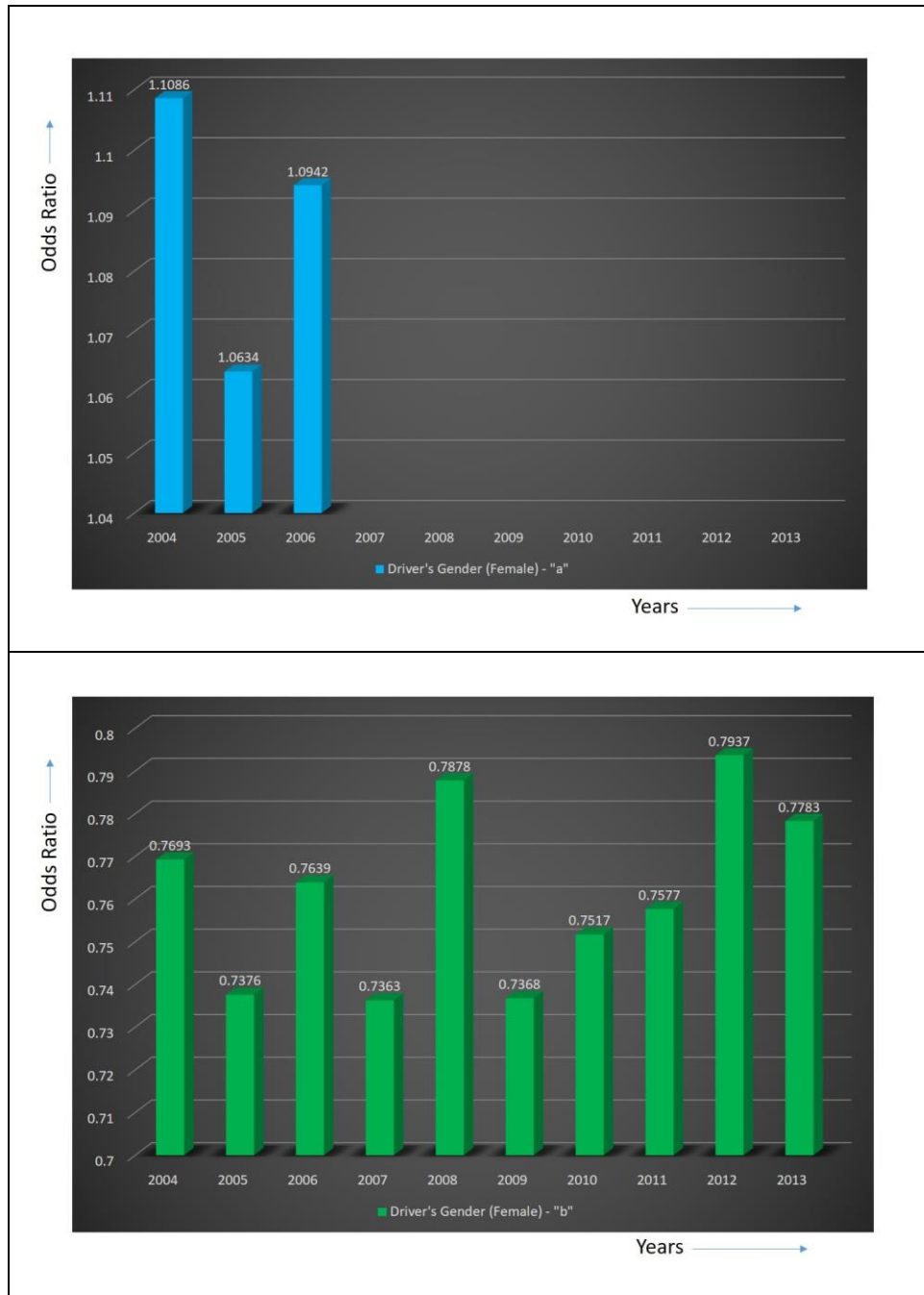
<b>Particulars</b>	<b>Driver Age</b>		<b>Vehicle Age</b>		<b>Driver Gender</b>		<b>Driver Impairment</b>	
	<b>“a”</b>	<b>“b”</b>	<b>“a”</b>	<b>“b”</b>	<b>“a”</b>	<b>“b”</b>	<b>“a”</b>	<b>“b”</b>
<i>Coefficient</i>	0.0055	0.0026	0.0201	0.0110	0.0849	-0.2731	1.5690	1.2702
<i>Odds Ratio</i>	1.006	1.003	1.020	1.011	1.089	0.761	4.827	3.582

From table 4 we see that an impaired driver is 4.83 times more likely to be injured compared to an unimpaired driver when involved in an accident. This result is similar to the findings of Behnood et al. (2014). The possible explanation for higher risk of injury for an impaired driver could be that the driver may not be reacting fast enough (for example normal driver might go to a bracing position faster at the time of accident compared to an impaired driver). Another finding which we see from Table 4 (related to impairment level) is that an impaired driver is 3.58 times more likely to increase the injury severity for other driver. This could possibly be explained by the fact that impaired drivers are less likely to adopt corrective maneuvers (in terms of braking or steering away) and correspondingly the impact speeds can be higher causing higher injury severities to the injured driver. When we look at the temporal trends related with impaired driving (Figure 14), we see that Model “a” has a highest value of 5.8258 during 2011 and a lowest value of 4.2428 in 2008. In the case of Model “b”, we see that the highest value recorded is 4.3152 in 2011 and the lowest value recorded was 3.0362 in 2006. Both the models were consistently significant during the ten year analysis period.



**Figure 14 – Temporal Trend for Driver Impairment – Models a and b**

When it comes to the gender of the driver involved in the accident, we see from tables 2 and 3 that female drivers are less likely to cause more serious injuries to other road users. This can be explained by the fact that women tend to be more cautious drivers compared to their male counterparts. Generally, aggressive driving behaviour, road rage etc. is exhibited more by males than female drivers. (Social Issues Research Centre 2004)

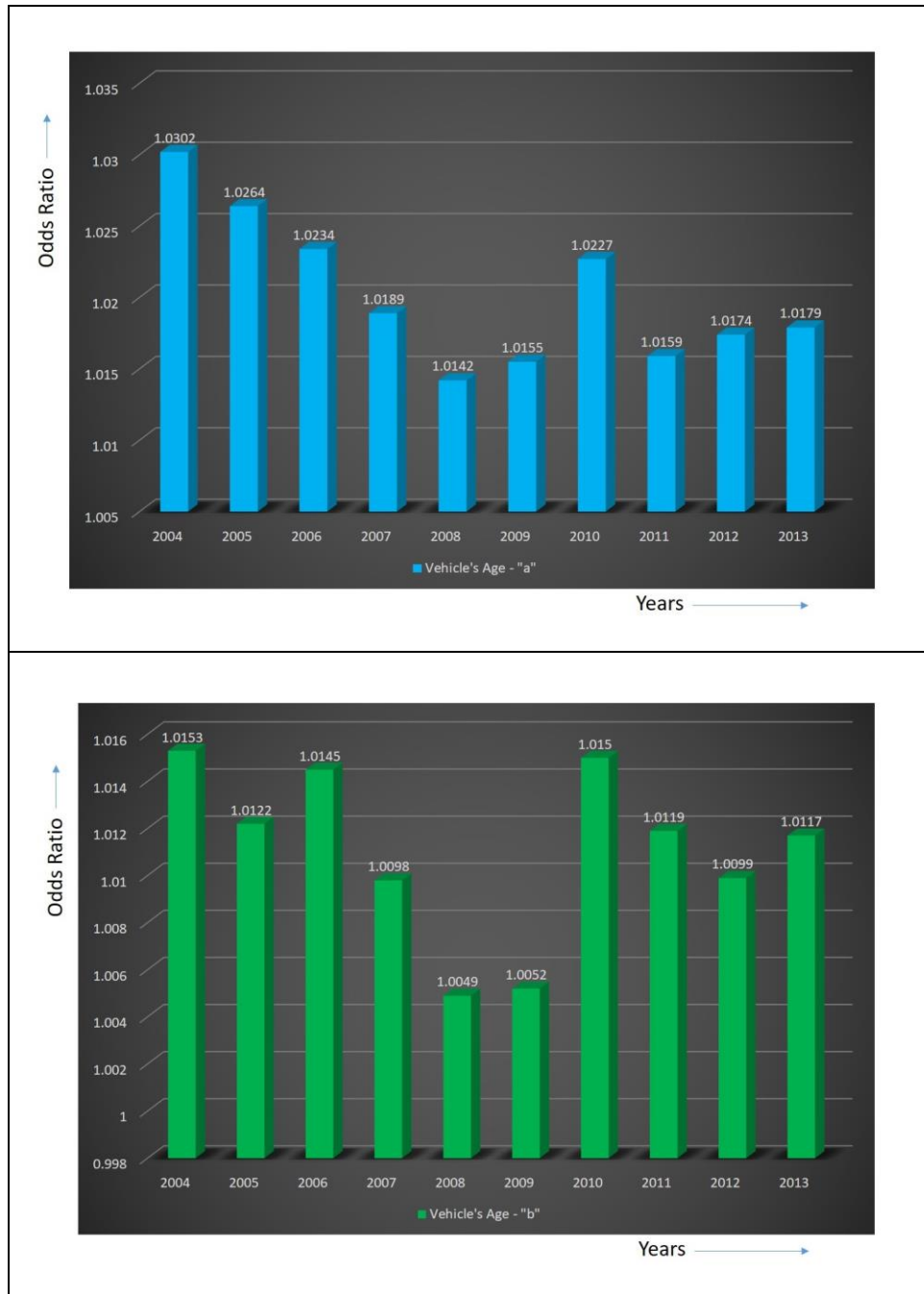


**Figure 15 – Temporal Trend for Driver Gender – Models a and b**

When we look at previous studies related with the gender of the driver, we see discussions related with “which gender is more severely injured?”. Behnood and Mannering (2015) analysed the effect of male drivers on injury severity for the years 2004 to 2012, where they found that for the

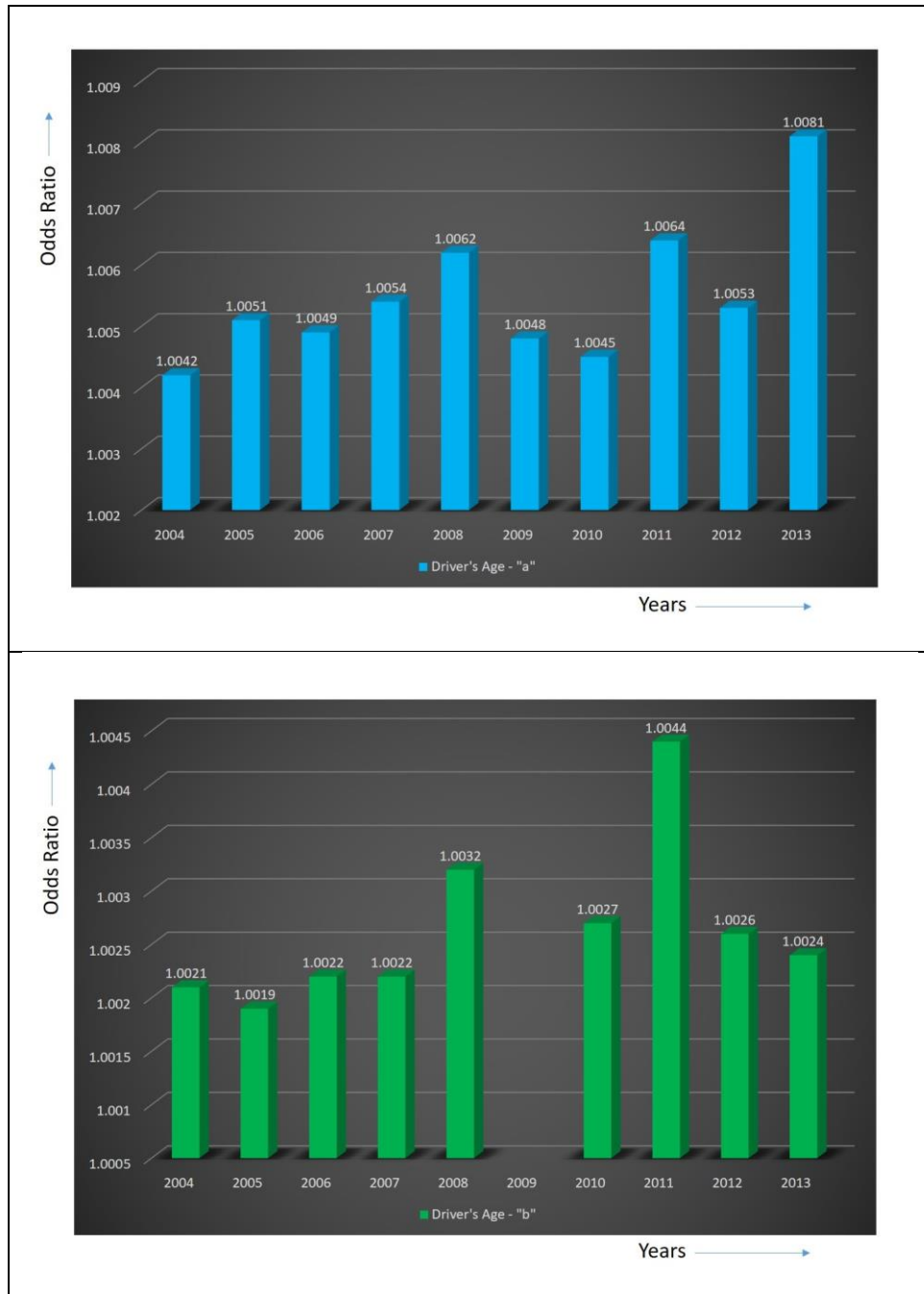
year 2007 and 2009, with regard to severe injuries, male drivers have a greater likelihood of severe injuries. In all the other years it was found that males have a less likelihood of severe injuries. Another study by Kim et al. (2013) showed that male drivers are associated with a greater probability of fatal injuries and it was attributed to the fact that male drivers are usually over-represented in the fatal injury category. When we look at the temporal trends related with driver gender (Figure 15), we see that values obtained for Model “a” was not significant from 2007 to 2013. The highest value obtained in model “a” was 1.1086 in 2004. Based on the temporal trend exhibited by model “a”, we can conclude that gender of the driver (model “a”) is not a consistently significant factor in predicting the probability of driver’s own injury. In the case of Model “b”, we see that the highest value recorded is 0.7937 in 2012 and the lowest value recorded was 0.7363 in 2007. It should also be kept in mind that the coefficient for driver gender for model “b” (obtained from Table 2) is negative indicating that female drive are less likely to cause injuries to the other driver. By looking at the temporal trend, we can conclude that gender of the driver (model “b”) is consistently significant.

The ages of the colliding vehicles were also found to have a consistently significant influence on the severity of drivers’ injuries. It can be seen from table 4 that as the vehicle age increases by one year, the odds of being seriously injured increase by 2.0%. So, we can say that, using a 10-year old vehicle can increase the odds of serious injury by 20%. Another result related to vehicle’s age from table 4 is that as the age of the colliding vehicle increases by 1 year, it increases the odds of serious injuries by 1.1%. So, in short, the older the vehicle, the greater the injury severity for the injured driver. This result is similar to previous studies (Wenzel 2013) and it might be explained by the fact that modern vehicles have more safety features than older vehicles. Also, due to the higher safety standard requirements, modern vehicles are subject to more rigorous crash tests prior to being commercially available. When we look at the temporal trends related with vehicle age (Figure 16), we see that Model “a” has a highest value of 1.0302 during 2004 and a lowest value of 1.0142 in 2008. In the case of Model “b”, we see that the highest value recorded is 1.0153 in 2004 and the lowest value recorded was 1.0049 in 2008. Both the models were consistently significant during the ten year analysis period.



**Figure 16 – Temporal Trend for Vehicle Age – Models a and b**

The age of the driver was also found to be consistently significant and increase the odds of driver's serious injuries. We see from table 4 that for an increase in the driver age by one year, the odds of being seriously injured increase by 0.6%. This is logical because an 80-year-old driver is more likely to be injured because the physical condition deteriorates as the years go by.



**Figure 17 – Temporal Trend for Driver Age – Models a and b**

Another result which we see from table 4 related with the driver age is that as the age of the driver of the colliding vehicle increases by 1 year, it increases the odds of serious injuries by 0.3%. This result is similar to previous studies (Kim et al. 2013) and it could be explained by the fact that the

responses of an older driver (in terms of braking or steering away) could be slower when compared to a young driver and correspondingly it can result in greater injury severity for the injured driver. When we look at the temporal trends related with driver age (Figure 17), we see that Model “a” has a highest value of 1.0081 during 2013 and a lowest value of 1.0042 in 2004. In the case of Model “b”, we see that the highest value recorded is 1.0044 in 2011 and the lowest value recorded was 1.0019 in 2005. For model “b”, the value corresponding to 2009 was not significant. So by looking at the temporal trend, we can say that model “b” is significant and model “a” is consistently significant during the ten year analysis period.



## 6 DISCUSSIONS

By understanding the risk imposed by the different explanatory variables considered for the above analysis, it is important to look at what we can do to minimise the risks for different drivers in the future.

As we saw earlier, driver impairment is the major factor that has to be targeted so as to make the roads safer for all road users. The first step is by proper enforcement of laws which eliminate drunk driving and also the enforcement of strict measures such as license suspensions/cancellations depending on the severity of the offense. Another method of law enforcement is by setting up sobriety checkpoints for breath checks to see if the driver is impaired or not. A third option is by installing ignition interlocks in vehicles. These are devices installed in vehicles that measure the driver's breath before each journey and if the driver is impaired, he/she won't be able to start the vehicle. Generally, these devices are installed in vehicles of drivers who have been convicted multiple times for driving under the influence (DUI). However, making it mandatory for all vehicles can act as a deterrent for people who drive while being impaired.

When it comes to the gender of the driver, we saw that female drivers are less likely to cause serious injury to the other driver compared to their male counterparts. The best option here is to educate drivers the significance of being calm behind the wheel and to take necessary precautionary steps to avoid aggressive driving. Usually, drivers get agitated when they are late. So, planning ahead and giving yourself enough time to reach your destination can definitely help. Traffic congestions is another reason for aggressive driving behaviour. So, the driver should identify alternate routes with less traffic and be ready to use them so as to avoid aggressive driving. Even though such precautionary measures are adopted, the drivers should come to terms with the fact that traffic congestions are part of driving and its best handled with a peaceful attitude.

When it comes to the age of the vehicle, we saw that as the age of the vehicle increases, the severity of injuries increases as well. The average life span of a passenger car ranges from 8 to 10 years. However, there are older/classic cars on the road which have been brought back to life by restorations. While performing restorations/improvements, importance should be given for the safety aspects as well. Many of these classic cars do not even have seatbelt provisions which is a

disadvantage. Combine that with the lack of modern features like airbags and dedicated crumple zones during impact, older vehicles are far behind in terms of modern safety standards. Specific rules should be set apart for older vehicles such that at least the basic safety requirements are met during their restoration.

When it comes to the age of the driver, we saw that as the age increases, the chances of severe injury increases. We also saw that aged drivers can cause more injuries to the other driver. Frequent senior driver assessments can ensure that the senior drivers are physically and cognitively fit to be on the road. Vehicle related improvements can also be made by car manufacturers such that driver restraint systems like seatbelts, airbags etc. are designed specifically for senior citizens so as to reduce injuries. Road infrastructure changes such as better road lighting facilities can help senior citizens see vehicles better during night time driving reducing the chances of accidents. Improved educational and refresher courses related with driving should be made available to old drivers before their license renewals. All the above-mentioned measures can definitely help make our roads much safer.

## 7 CONCLUSIONS

Logistic regression was used to identify and quantify the effects of various explanatory variables that increase the risk of injury in drivers associated with two-vehicle collisions for light duty vehicles. The study was conducted by analysing the accident records of all two-vehicle light-duty vehicle collisions that occurred in North Carolina from January 1, 2004 to December 31, 2013. Many of the previous studies have primarily focussed on single vehicle accidents (eg. Yau 2004, Chang and Yeh 2006 etc.). Those studies which have analysed multivehicle accidents (eg. Yan et.al 2005, Harb et al. 2008 etc.) have done so, by looking at smaller analysis periods and often in these studies we see that data pertaining to various years are combined to a single database for analysis. This method of combining data pertaining to various years could possibly result in aggregation bias which is eliminated in the current study. Each year is analysed separately and the temporal trends pertaining to the various explanatory variables for the ten-year analysis period is looked at. Two logistic regression models were developed for each year. A first model, which evaluates the probability that a certain driver sustains at least a visible injury caused by the other driver and a second model, which evaluates the probability that a driver will cause at least a visible injury to the other driver. Explanatory variables related to the drivers as well as those related to the colliding vehicles were considered for the analysis. The explanatory variables related to the driver were the age, the gender and impairment levels of both colliding drivers. Vehicle's ages were the explanatory variables related to the vehicles.

Driver impairment was found to be the most consistently significant factor with the highest probability of causing injuries to the drivers. It was found that an impaired driver is most likely to be seriously injured and also to cause serious injuries to other drivers. When it comes to driver gender, we saw that the gender of the driver that causes injury to the other driver was found to be consistently significant during the analysis period and that female drivers cause less injury to the injured driver compared to males. Temporal analysis also revealed that gender of the driver is not a consistently significant factor in predicting the probability of driver's own injury. Age of the driver was also found to be a consistently significant factor that increases the odds of serious injuries. The study also found out that modern vehicles are consistently associated with less driver injuries which can be attributed to the improved vehicle safety standards compared to older vehicles.

It is very important to understand that, traffic collisions that result in serious injuries constitute a large burden on the healthcare system. Identifying various factors that increase the risk of drivers' serious injuries can help decision makers take necessary precautionary measures to reduce those risks or to improve legislative policies. The adoption and proper enforcement of necessary policies can definitely make our roads a safer place.



<div> <div>Variable</div> <div>Avg</div> <div>SD</div> </div> <div> <div>1</div> <div>38.5457</div> <div>16.7398</div> </div> <div> <div>2</div> <div>38.5457</div> <div>16.7398</div> </div>	<div> <div>Variable</div> <div>Avg</div> <div>SD</div> </div> <div> <div>1</div> <div>7.5734</div> <div>5.6684</div> </div> <div> <div>2</div> <div>7.5734</div> <div>5.6684</div> </div>
<div>Iteration History...</div> <div>-2 Log Likelihood = 54089.6030 (Null Model)</div> <div>-2 Log Likelihood = 54078.8583</div> <div>-2 Log Likelihood = 54062.4666</div> <div>-2 Log Likelihood = 54047.7773</div> <div>-2 Log Likelihood = 54038.8800</div> <div>-2 Log Likelihood = 54035.1830</div> <div>-2 Log Likelihood = 54034.1606</div> <div>-2 Log Likelihood = 54033.9849</div> <div>-2 Log Likelihood = 54033.9683</div> <div>-2 Log Likelihood = 54033.9677</div> <div>-2 Log Likelihood = 54033.9677</div> <div>-2 Log Likelihood = 54033.9677 (Converged)</div>	<div>Iteration History...</div> <div>-2 Log Likelihood = 54089.6030 (Null Model)</div> <div>-2 Log Likelihood = 54046.6691</div> <div>-2 Log Likelihood = 53982.8281</div> <div>-2 Log Likelihood = 53928.7641</div> <div>-2 Log Likelihood = 53898.5457</div> <div>-2 Log Likelihood = 53886.9121</div> <div>-2 Log Likelihood = 53883.8553</div> <div>-2 Log Likelihood = 53883.3444</div> <div>-2 Log Likelihood = 53883.2968</div> <div>-2 Log Likelihood = 53883.2948</div> <div>-2 Log Likelihood = 53883.2948</div> <div>-2 Log Likelihood = 53883.2948 (Converged)</div>
<div>Overall Model Fit...</div> <div>Chi Square= 55.6353; df=2; p= 0.0000</div>	<div>Overall Model Fit...</div> <div>Chi Square= 206.3082; df=2; p= 0.0000</div>
<div>Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...</div> <div>Variable</div> <div>Coeff</div> <div>StdErr</div> <div>p</div> <div>O.R.</div> <div>Low</div> <div>High</div> <div>1</div> <div>0.0051</div> <div>0.0007</div> <div>0.0000</div> <div>1.0051</div> <div>1.0037</div> <div>1.0065</div> <div>2</div> <div>0.0019</div> <div>0.0007</div> <div>0.0115</div> <div>1.0019</div> <div>1.0004</div> <div>1.0033</div> <div>Intercept</div> <div>-3.2855</div> <div>0.0421</div> <div>0.0000</div>	<div>Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...</div> <div>Variable</div> <div>Coeff</div> <div>StdErr</div> <div>p</div> <div>O.R.</div> <div>Low</div> <div>High</div> <div>1</div> <div>0.0266</div> <div>0.0019</div> <div>0.0000</div> <div>1.0264</div> <div>1.0225</div> <div>1.0302</div> <div>2</div> <div>0.0121</div> <div>0.0021</div> <div>0.0000</div> <div>1.0122</div> <div>1.0081</div> <div>1.0163</div> <div>Intercept</div> <div>-3.3155</div> <div>0.0259</div> <div>0.0000</div>
<div>Predicted Probability of Outcome, with 95% Confidence Limits...</div> <div>X</div> <div>Y</div> <div>Prob</div> <div>Low</div> <div>High</div> <div>54.0000</div> <div>73.0000</div> <div>0</div> <div>0.0534</div> <div>0.0506</div> <div>0.0564</div> <div>27.0000</div> <div>78.0000</div> <div>0</div> <div>0.0473</div> <div>0.0445</div> <div>0.0503</div> <div>16.0000</div> <div>43.0000</div> <div>0</div> <div>0.0421</div> <div>0.0405</div> <div>0.0439</div> <div>39.0000</div> <div>47.0000</div> <div>0</div> <div>0.0475</div> <div>0.0462</div> <div>0.0487</div> <div>75.0000</div> <div>17.0000</div> <div>0</div> <div>0.0535</div> <div>0.0503</div> <div>0.0569</div> <div>44.0000</div> <div>39.0000</div> <div>0</div> <div>0.0479</div> <div>0.0468</div> <div>0.0491</div> <div>20.0000</div> <div>58.0000</div> <div>0</div> <div>0.0441</div> <div>0.0422</div> <div>0.0462</div> <div>43.0000</div> <div>29.0000</div> <div>0</div> <div>0.0469</div> <div>0.0456</div> <div>0.0482</div> <div>20.0000</div> <div>42.0000</div> <div>0</div> <div>0.0429</div> <div>0.0414</div> <div>0.0445</div> <div>30.0000</div> <div>27.0000</div> <div>0</div> <div>0.0438</div> <div>0.0425</div> <div>0.0452</div> <div>73.0000</div> <div>29.0000</div> <div>0</div> <div>0.0542</div> <div>0.0514</div> <div>0.0571</div> <div>22.0000</div> <div>27.0000</div> <div>0</div> <div>0.0422</div> <div>0.0406</div> <div>0.0438</div> <div>19.0000</div> <div>61.0000</div> <div>0</div> <div>0.0444</div> <div>0.0421</div> <div>0.0464</div> <div>28.0000</div> <div>29.0000</div> <div>0</div> <div>0.0436</div> <div>0.0422</div> <div>0.0450</div> <div>34.0000</div> <div>40.0000</div> <div>0</div> <div>0.0457</div> <div>0.0446</div> <div>0.0469</div> <div>45.0000</div> <div>20.0000</div> <div>0</div> <div>0.0466</div> <div>0.0449</div> <div>0.0483</div> <div>18.0000</div> <div>29.0000</div> <div>1</div> <div>0.0415</div> <div>0.0399</div> <div>0.0432</div> <div>32.0000</div> <div>41.0000</div> <div>0</div> <div>0.0454</div> <div>0.0442</div> <div>0.0466</div> <div>43.0000</div> <div>89.0000</div> <div>0</div> <div>0.0521</div> <div>0.0485</div> <div>0.0550</div> <div>34.0000</div> <div>28.0000</div> <div>0</div> <div>0.0448</div> <div>0.0435</div> <div>0.0461</div> <div>43.0000</div> <div>25.0000</div> <div>0</div> <div>0.0465</div> <div>0.0451</div> <div>0.0480</div> <div>16.0000</div> <div>55.0000</div> <div>0</div> <div>0.0431</div> <div>0.0415</div> <div>0.0447</div> <div>34.0000</div> <div>33.0000</div> <div>0</div> <div>0.0452</div> <div>0.0440</div> <div>0.0464</div> <div>35.0000</div> <div>77.0000</div> <div>0</div> <div>0.0491</div> <div>0.0463</div> <div>0.0520</div>	<div>Predicted Probability of Outcome, with 95% Confidence Limits...</div> <div>X</div> <div>Y</div> <div>Prob</div> <div>Low</div> <div>High</div> <div>9.0000</div> <div>13.0000</div> <div>0</div> <div>0.0510</div> <div>0.0495</div> <div>0.0526</div> <div>20.0000</div> <div>1.0000</div> <div>0</div> <div>0.0582</div> <div>0.0551</div> <div>0.0615</div> <div>19.0000</div> <div>14.0000</div> <div>0</div> <div>0.0659</div> <div>0.0628</div> <div>0.0692</div> <div>5.0000</div> <div>11.0000</div> <div>0</div> <div>0.0451</div> <div>0.0438</div> <div>0.0465</div> <div>14.0000</div> <div>1.0000</div> <div>0</div> <div>0.0502</div> <div>0.0482</div> <div>0.0524</div> <div>20.0000</div> <div>20.0000</div> <div>0</div> <div>0.0723</div> <div>0.0678</div> <div>0.0770</div> <div>13.0000</div> <div>5.0000</div> <div>0</div> <div>0.0513</div> <div>0.0498</div> <div>0.0529</div> <div>11.0000</div> <div>4.0000</div> <div>0</div> <div>0.0483</div> <div>0.0469</div> <div>0.0497</div> <div>1.0000</div> <div>10.0000</div> <div>0</div> <div>0.0404</div> <div>0.0389</div> <div>0.0419</div> <div>8.0000</div> <div>5.0000</div> <div>0</div> <div>0.0454</div> <div>0.0442</div> <div>0.0466</div> <div>2.0000</div> <div>3.0000</div> <div>0</div> <div>0.0382</div> <div>0.0367</div> <div>0.0396</div> <div>6.0000</div> <div>10.0000</div> <div>0</div> <div>0.0457</div> <div>0.0445</div> <div>0.0470</div> <div>16.</div>

**Figure 19 - Results for Year 2005 obtained from Logistic Regression Calculator**



















Variable

Avg

SD

1

40.1795

17.2635

2

40.1795

17.2635

Iteration History...

-2 Log Likelihood = 52320.3511 (Null Model)

-2 Log Likelihood = 52294.6546

-2 Log Likelihood = 52255.5630

-2 Log Likelihood = 52220.7115

-2 Log Likelihood = 52199.7238

-2 Log Likelihood = 52191.0421

-2 Log Likelihood = 52188.6472

-2 Log Likelihood = 52188.2360

-2 Log Likelihood = 52188.1972

-2 Log Likelihood = 52188.1956

-2 Log Likelihood = 52188.1955

-2 Log Likelihood = 52188.1955 (Converged)

Overall Model Fit...

Chi Square= 132.1555; df=2; p= 0.0000

Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...

Variable	Coeff.	StdErr	p	O.R.	Low	High
1	0.0081	0.0007	0.0000	1.0081	1.0067	1.0096
2	0.0024	0.0008	0.0014	1.0024	1.0009	1.0039
Intercept	-3.7488	0.0448	0.0000			

Predicted Probability of Outcome, with 95% Confidence Limits...

X	Y	Prob	Low	High	
18.0000	23.0000	0	0.0280	0.0266	0.0294
19.0000	44.0000	0	0.0206	0.0284	0.0309
42.0000	27.0000	0	0.0341	0.0330	0.0352
18.0000	48.0000	0	0.0297	0.0284	0.0310
73.0000	84.0000	0	0.0494	0.0458	0.0533
68.0000	57.0000	0	0.0647	0.0426	0.0469
19.0000	26.0000	0	0.0284	0.0271	0.0297
21.0000	33.0000	0	0.0293	0.0282	0.0305
56.0000	32.0000	0	0.0385	0.0372	0.0398
52.0000	74.0000	0	0.0411	0.0389	0.0433
46.0000	21.0000	0	0.0347	0.0334	0.0360
59.0000	31.0000	0	0.0393	0.0379	0.0408
22.0000	22.0000	0	0.0288	0.0275	0.0302
17.0000	28.0000	1	0.0281	0.0268	0.0294
26.0000	45.0000	0	0.0314	0.0303	0.0325
20.0000	24.0000	0	0.0285	0.0272	0.0298
79.0000	36.0000	0	0.0464	0.0439	0.0491
27.0000	52.0000	0	0.0321	0.0310	0.0333
19.0000	52.0000	0	0.0302	0.0288	0.0315
59.0000	22.0000	0	0.0385	0.0368	0.0402
28.0000	34.0000	0	0.0311	0.0300	0.0321
66.0000	23.0000	0	0.0407	0.0388	0.0428
77.0000	60.0000	0	0.0483	0.0455	0.0512
21.0000	56.0000	0	0.0309	0.0296	0.0324

DRIVER AGE

Variable

Avg

SD

1

0.4673

0.4989

2

0.4673

0.4989

Iteration History...

-2 Log Likelihood = 52320.3511 (Null Model)

-2 Log Likelihood = 52302.3952

-2 Log Likelihood = 52274.7884

-2 Log Likelihood = 52249.6474

-2 Log Likelihood = 52234.0645

-2 Log Likelihood = 52227.4160

-2 Log Likelihood = 52225.5310

-2 Log Likelihood = 52225.2007

-2 Log Likelihood = 52225.1692

-2 Log Likelihood = 52225.1678

-2 Log Likelihood = 52225.1678

-2 Log Likelihood = 52225.1678

-2 Log Likelihood = 52225.1678 (Converged)

Overall Model Fit...

Chi Square= 95.1832; df=2; p= 0.0000

Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...

Variable	Coeff.	StdErr	p	O.R.	Low	High
1	-0.0564	0.0263	0.0321	0.9451	0.8976	1.0952
2	-0.2507	0.0267	0.0000	0.7783	0.7387	0.8200
Intercept	-3.1810	0.0207	0.0000			

Predicted Probability of Outcome, with 95% Confidence Limits...

X	Y	Prob	Low	High	
0.0000	1.0000	0	0.0313	0.0299	0.0328
0.0000	1.0000	0	0.0313	0.0299	0.0328
0.0000	0.0000	0	0.0399	0.0384	0.0415
1.0000	0.0000	0	0.0378	0.0362	0.0394
0.0000	0.0000	0	0.0399	0.0384	0.0415
0.0000	0.0000	0	0.0399	0.0384	0.0415
0.0000	1.0000	0	0.0313	0.0299	0.0328
1.0000	0.0000	0	0.0378	0.0362	0.0394
0.0000	0.0000	0	0.0399	0.0384	0.0415
0.0000	1.0000	0	0.0313	0.0299	0.0328
0.0000	1.0000	0	0.0378	0.0362	0.0394
0.0000	0.0000	0	0.0399	0.0384	0.0415
0.0000	1.0000	0	0.0313	0.0299	0.0328
0.0000	0.0000	0	0.0399	0.0384	0.0415
1.0000	0.0000	1	0.0378	0.0362	0.0394
1.0000	1.0000	0	0.0297	0.0283	0.0311
1.0000	1.0000	0	0.0297	0.0283	0.0311
0.0000	0.0000	0	0.0378	0.0362	0.0394
0.0000	0.0000	0	0.0399	0.0384	0.0415
1.0000	0.0000	0	0.0378	0.0362	0.0394
1.0000	0.0000	0	0.0378	0.0362	0.0394
0.0000	0.0000	0	0.0399	0.0384	0.0415
1.0000	1.0000	0	0.0297	0.0283	0.0311
0.0000	0.0000	0	0.0399	0.0384	0.0415

DRIVER GENDER

Variable

Avg

SD

1

9.4447

6.8940

2

9.4447

6.8940

Iteration History...

-2 Log Likelihood = 52320.3511 (Null Model)

-2 Log Likelihood = 52279.0986

-2 Log Likelihood = 52219.7265

-2 Log Likelihood = 52172.8843

-2 Log Likelihood = 52149.0746

-2 Log Likelihood = 52140.7077

-2 Log Likelihood = 52138.6612

-2 Log Likelihood = 52138.3353

-2 Log Likelihood = 52138.3058

-2 Log Likelihood = 52138.3046

-2 Log Likelihood = 52138.3046 (Converged)

Overall Model Fit...

Chi Square= 182.0465; df=2; p= 0.0000

Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...

Variable	Coeff.	StdErr	p	O.R.	Low	High
1	0.0178	0.0014	0.0000	1.0179	1.0151	1.0208
2	0.0117	0.0016	0.0000	1.0117	1.0086	1.0149
Intercept	-3.6065	0.0250	0.0000			

Predicted Probability of Outcome, with 95% Confidence Limits...

X	Y	Prob	Low	High	
15.0000	6.0000	0	0.0366	0.0355	0.0378
14.0000	5.0000	0	0.0356	0.0345	0.0367
17.0000	1.0000	0	0.0358	0.0344	0.0373
13.0000	2.0000	0	0.0338	0.0326	0.0351
16.0000	20.0000	0	0.0436	0.0418	0.0454
12.0000	6.0000	0	0.0348	0.0338	0.0358
5.0000	9.0000	0	0.0319	0.0310	0.0328
15.0000	8.0000	0	0.0374	0.0364	0.0385
5.0000	5.0000	0	0.0305	0.0295	0.0315
25.0000	4.0000	0	0.0425	0.0404	0.0447
11.0000	16.0000	0	0.0383	0.0371	0.0394
15.0000	10.0000	0	0.0383	0.0373	0.0394
7.0000	9.0000	0	0.0330	0.0321	0.0339
20.0000	17.0000	1	0.0451	0.0433	0.0469
10.0000	7.0000	0	0.0340	0.0331	0.0349
13.0000	14.0000	0	0.0387	0.0376	0.0398
28.0000	7.0000	0	0.0462	0.0437	0.0488
19.0000	6.0000	0	0.0392	0.0378	0.0407
16.0000	20.0000	0	0.0436	0.0418	0.0454
7.0000	16.0000	0	0.0357	0.0346	0.0369
11.0000	20.0000	0	0.0400	0.0385	0.0416
1.0000	1.0000	0	0.0272	0.0260	0.0284
19.0000	15.0000	0	0.0454	0.0418	0.0450
13.0000	7.0000	0	0.0358	0.0348	0.0368

VEHICLE AGE

Variable

Avg

SD

1

0.0104

0.1013

2

0.0104

0.1013

Iteration History...

-2 Log Likelihood = 52320.3511 (Null Model)

-2 Log Likelihood = 52071.4742

-2 Log Likelihood = 51805.6474

-2 Log Likelihood = 51658.6558

-2 Log Likelihood = 51597.5281

-2 Log Likelihood = 51577.7245

-2 Log Likelihood = 51573.0311

-2 Log Likelihood = 51572.2940

-2 Log Likelihood = 51572.2276

-2 Log Likelihood = 51572.2250

-2 Log Likelihood = 51572.2249

-2 Log Likelihood = 51572.2249 (Converged)

Overall Model Fit...

Chi Square= 748.1261; df=2; p= 0.0000

Coefficients, Standard Errors, Odds Ratios, and 95% Confidence Limits...

Variable	Coeff.	StdErr	p	O.R.	Low	High
1	1.6823	0.0662	0.0000	5.3780	4.7238	6.1229
2	2.4162	0.0723	0.0000	4.1215	3.5773	4.7485
Intercept	-3.3852	0.0137	0.0000			

Predicted Probability of Outcome, with 95% Confidence Limits...

X	Y	Prob	Low	High	
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	1	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.0319	0.0336
0.0000	0.0000	0	0.0328	0.	

**Figure 27 - Results for Year 2013 obtained from Logistic Regression Calculator**

## REFERENCES

- Behnood, Ali, and Fred L. Mannering. 2015. "The Temporal Stability of Factors Affecting Driver-Injury Severities in Single-Vehicle Crashes: Some Empirical Evidence." *Analytic Methods in Accident Research* 8: 7–32.
- Behnood, Ali, Arash M. Roshandeh, and Fred L. Mannering. 2014. "Latent Class Analysis of the Effects of Age, Gender, and Alcohol Consumption on Driver-Injury Severities." *Analytic Methods in Accident Research* *younger than 31 years old were considered younger drivers, and those 31 year* 3–4: 56–91.
- Bham, Ghulam H., Bhanu S. Javvadi, and Uday R. R. Manepalli. 2012. "Multinomial Logistic Regression Model for Single-Vehicle and Multivehicle Collisions on Urban U.S. Highways in Arkansas." *Journal of Transportation Engineering* 138(6): 786–97.
- Chang, Hsin-Li, and Tsu-Hung Yeh. 2006. "Risk Factors to Driver Fatalities in Single-Vehicle Crashes: Comparisons between Non-Motorcycle Drivers and Motorcyclists." *Journal of Transportation Engineering* 132(3): 227–36.
- Dabbour, Essam. 2017. "Investigating Temporal Trends in the Explanatory Variables Related to the Severity of Drivers' Injuries in Single-Vehicle Collisions." *Journal of Traffic and Transportation Engineering (English Edition)* 4(1): 71–79.
- Daniels, Stijn, Tom Brijs, Erik Nuyts, and Geert Wets. 2010. "Externality of Risk and Crash Severity at Roundabouts." *Accident Analysis and Prevention* 42(6): 1966–73.
- Elvik, Rune. 2008. "Dimensions of Road Safety Problems and Their Measurement." 40(349): 1200–1210.
- Harb, Rami et al. 2008. "Freeway Work-Zone Crash Analysis and Risk Identification Using Multiple and Conditional Logistic Regression." *Journal of Transportation Engineering* 134(5): 203–14.

- Kim, Joon Ki, Gudmundur F. Ulfarsson, Sungyop Kim, and Venkataraman N. Shankar. 2013. "Driver-Injury Severity in Single-Vehicle Crashes in California: A Mixed Logit Analysis of Heterogeneity due to Age and Gender." *Accident Analysis and Prevention* 50: 1073–81.
- Pezzullo, John. 2015. "Logistic Regression." <http://statpages.info/logistic.html>.
- Social Issues Research Centre. 2004. "Sex Differences in Driving and Insurance Risk." (August): 1–24. <http://www.sirc.org/publik/driving.pdf>
- University of North Carolina Highway Research Center. 2010. "Highway Safety Information System." <http://www.hsisinfo.org/index.cfm> (December 20, 2017).
- Wenzel, Tom. 2013. "The Effect of Recent Trends in Vehicle Design on U.S. Societal Fatality Risk per Vehicle Mile Traveled, and Their Projected Future Relationship with Vehicle Mass." *Accident Analysis and Prevention* 56(2013): 71–81.
- Yan, Xuedong, Essam Radwan, and Mohamed Abdel-Aty. 2005. "Characteristics of Rear-End Accidents at Signalized Intersections Using Multiple Logistic Regression Model." *Accident Analysis and Prevention* 37(6): 983–95.
- Yau, Kelvin K W. 2004. "Risk Factors Affecting the Severity of Single Vehicle Traffic Accidents in Hong Kong." *Accident Analysis and Prevention* 36(3): 333–40.
- Yu, Rongjie, Mohamed A. Abdel-Aty, Mohamed M. Ahmed, and Xuesong Wang. 2014. "Utilizing Microscopic Traffic and Weather Data to Analyze Real-Time Crash Patterns in the Context of Active Traffic Management." *IEEE Transactions on Intelligent Transportation Systems* 15(1): 205–13.