



16th Road Safety on Four Continents Conference
Beijing, China 15-17 May 2013

Predicting Drivers' Crash Risk Based-on Previous Crash History

Xiaoduan Sun, Ph.D., and P.E. (Corresponding author)

Professor
Civil Engineering Department
University of Louisiana
Lafayette, LA 70504
337-739-6732
337-739-6688 (fax)
xsun@louisiana.edu

Subasish Das
PhD. Candidate
Civil Engineering Department
University of Louisiana

Rasel Sk
Graduate Student
Civil Engineering Department
University of Louisiana

Fan Wang
Graduate Student
Civil Engineering Department
University of Louisiana

*Submitted to the 16th International Conference "Road Safety on Four Continents"
May 15, 2013, Beijing, China*

ABSTRACT

Crash-prone drivers should be effectively targeted for various safety education and regulation programs because their over-involvement in crashes presents a big adverse effect on highway safety. By analyzing seven-years of crash data from Louisiana, this paper investigates crash-prone drivers' characteristics and estimates their risk to have crashes in the seventh year based on these drivers' crash history of the past six years.

The analysis results show that quite a few drivers repeatedly had crashes; seven drivers had 13 crashes in seven years; and the maximum number of crashes occurring in a single year to a single driver is eight. Actually, the 5% of drivers who had multiple crashes were responsible for 35% of the crashes that occurred in the seven years in Louisiana. Crash injury rate is also higher for drivers with multiple crashes. The



16th Road Safety on Four Continents Conference

Beijing, China 15-17 May 2013

probability of having crash(es) in any given year is closely related to a driver's crash history; less than 4% of drivers with no crash in the previous six years; and slightly higher than 30% for drivers with nine or more crashes in the previous six years. There are variations in drivers' age, gender, crash contribution factors, and type of crashes by the number of crashes.

Based on the results, several suggestions are made on how to improve roadway safety through reducing crashes committed by drivers with much higher crash risk as identified by the analysis.

Key words: roadway safety, crash-prone driver, crash risk, probability

1 INTRODUCTION

Although conservatively speaking more than 50% of crashes are caused by human errors each year based on highway crash reports, engineers are always trying to make roadways more forgiving and vehicles more crashworthy, which has made considerable impact on highway safety. Under the persistent effort, highway fatal crashes in U.S. have finally reached the lowest number since 1960. Much of the effort has been spent on implementing crash countermeasures on highway facilities through enhancing safety on roadway geometric features and traffic control devices. Safety education and enforcement, the other two elements in the 4E approach (emergency services the fourth), also made strides in educating the general public on various safety risks and enforcing traffic safety laws.

To fulfill the hefty goal established by the AASHTO Highway Safety Strategy to cut traffic fatalities in half by 2020 and by Louisiana Strategic Highway Safety Plan for Destination Zero Deaths, it is important to have effective safety education and regulation programs while continually improving the highway infrastructure's safety. Since crash-prone drivers present a big adverse effect on highway safety, they should be effectively targeted in various safety education and enforcement programs. It is generally known that very young and very old drivers have the highest fatal crash rates, but it does not mean that these two groups commit most crashes. People with similar personal traits could have very different crash risk. Identifying high risk drivers and studying their characteristics are critical in further reducing the number of crashes through targeted safety education and enforcement programs. Thus, a project was conducted at the University of Louisiana to study the impact of crash-prone drivers on safety and to predict how a driver's past crash history could affect his/her crash occurrence(s) in the upcoming year. Most importantly, the study is to provide evidence for developing better and efficient safety education programs and supporting targeted traffic laws or programs on these crash over-involved drivers.

16th Road Safety on Four Continents Conference
Beijing, China 15-17 May 2013

2 LITERATURE REVIEW

Investigating crash-prone drivers' characteristics, exploring the relationship between drivers' past crash/citation history and their crash risk, and predicting drivers' future crash occurrences by their previous crash history were the focus of many past studies.

The existence of crash-prone drivers was first recognized in 1920 by Greenwood and Yule. In their published paper (1), crash-prone drivers are defined as the drivers with higher than expected number of crashes. In a 2003 study by Blasco, crash-prone drivers are described as the drivers with recurring crashes that were caused by human error, not by coincidence (2). In 1971, Peck et al. concluded that it is quite difficult to accurately identify which driver will or will not cause crashes because of the statistical nature of crash frequencies (3). After analyzing five years of Kentucky crash data (1993–1997), Stamatiadis found that about 2.1% of the licensed drivers who were charged with six or more points in past 2 years accounted for nearly 5.3% of all crashes (4).

Predicting a driver's crash risk based on his/her past crash and traffic offence history was the topic of many investigations. Through examining older drivers' previous conviction records and crash data, Daigneault in 2002 concluded that prior crashes are a better predictor for crash risk than prior convictions (5). In a study published in 1991, Hauer determined that if the prediction model makes the right use of the driver's past crash records, the performance of their multivariate model for a crash would be improved (6). A logistic regression model was developed by Chen in 1995 to identify crash prone drivers based on their records prior to their at-fault crash involvements (7), which discovered that a model using prior at-fault crash data can recognize up to 23% more drivers who will have one or more at-fault crash involvements in the next 2 years than a model that uses the conviction information. After studying 17 logistic regression models, Gebers in 1999 concluded that his models could correctly classify crash-involved drivers up to 27.6% (8). By deploying canonical correlation techniques in a subsequent research effort, Gebers and Peck in 2003 achieved an accuracy level up to 27.2% from their best model to identify crash-prone drivers (9). Although no model can be considered perfect, the modeling progress can be seen in much research especially from the Californian studies (7 and 8). However, some researchers have voiced their skepticism over crash-prone drivers' prediction (3 and 8).

Although the past studies on crash-prone drivers have yielded some interesting results, the small sample size and lack of detailed data on drivers' characteristics and crashes have limited the research findings.

3 DATA ANALYSIS

Seven years of crash data, 2004-2010, in Louisiana are used for this analysis. There are more than one million crashes recorded in this seven year period, but about 10% of these records do not have the computer generated driver ID (meaning no driver license information collected at crash scene or hit-and-run crashes), which makes the cross-year analysis impossible. The crash records with no driver ID were removed from the analysis.

16th Road Safety on Four Continents Conference

Beijing, China 15-17 May 2013

The at-fault drivers, i.e., drivers responsible or mainly responsible for a crash listed as driver-1 in the data based, were selected for the analysis. In general about 4% of licensed drivers in Louisiana are involved in at least one crash each year. The number of drivers having crashes is summarized in Table 1, which reveals some drivers had crashes repeatedly within one year. The annual maximum number of crashes to a

Table 1: Number of Drivers with Crashes

Number of Crashes	2004	2005	2006	2007	2008	2009	2010
1	128,396	123,169	122,469	121,242	120,366	121,498	114,063
2	6,110	5,567	5,848	5,827	5,382	5,364	4,833
3	457	392	441	433	428	376	316
4	49	38	60	61	42	40	22
5	7	9	10	12	6	6	11
6	1	3	1	0	0	4	1
7	2	0	0	1	1	0	1
8	1	0	1	1	0	0	1
Total Number of Drivers	135,023	129,178	128,830	127,577	126,225	127,288	119,248
Number of Crashes	142,246	135,694	135,792	134,514	132,619	133,568	124,841

single driver is eight. About 10% of crashes occurred to drivers having multiple crashes annually. By examining the seven years of data together, we found that the maximum number of crashes by a single driver is 13 as shown in Table 2.

16th Road Safety on Four Continents Conference
Beijing, China 15-17 May 2013

Table 2: Drivers' Basic Information

Number of Crashes	Number of Drivers	Gender		Licensed By		% of Total Crashes
		Male	Female	Louisiana	Other	
13	7	85.71%	14.29%	100.00%	0.00%	0.01%
12	5	100.00%	0.00%	100.00%	0.00%	0.01%
11	6	66.67%	33.33%	100.00%	0.00%	0.01%
10	23	73.91%	26.09%	100.00%	0.00%	0.02%
9	31	61.29%	38.71%	100.00%	0.00%	0.03%
8	95	77.89%	22.11%	100.00%	0.00%	0.08%
7	223	78.92%	21.08%	100.00%	0.00%	0.17%
6	573	71.73%	28.27%	99.83%	0.17%	0.37%
5	1,675	70.93%	29.07%	99.28%	0.72%	0.89%
4	6,139	68.99%	30.85%	80.24%	19.76%	2.61%
3	23,414	64.60%	35.40%	98.51%	1.49%	7.48%
2	105,621	59.36%	40.62%	97.09%	2.91%	22.49%
1	618,388	56.13%	43.54%	90.21%	9.79%	65.84%

As expected, the majority of crashes involved drivers holding a Louisiana driver license. About 66% and 34% of crashes are blamed on drivers with a single crash and with multiple crashes in seven years, respectively. These 34% crashes were repeatedly committed by 137,812 drivers that accounts for 5% of licensed drivers in the state. As also shown in Table 3 and Figure 1, drivers with multiple crashes tend to have higher injury rate before the final fatal crash.

Table 3: Distribution of Crash Severity

Severity	Number of Crashes in Seven Years												
	13	12	11	10	9	8	7	6	5	4	3	2	1
Fatal Crashes	0	0	0	1	1	1	1	4	14	34	133	525	2,429
Injury Crashes	22	4	14	46	64	128	302	658	1,507	4,272	11,879	34,569	99,534
PDO	67	56	51	183	213	630	1,256	2,768	6,832	20,182	58,034	175,554	514,534

16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013

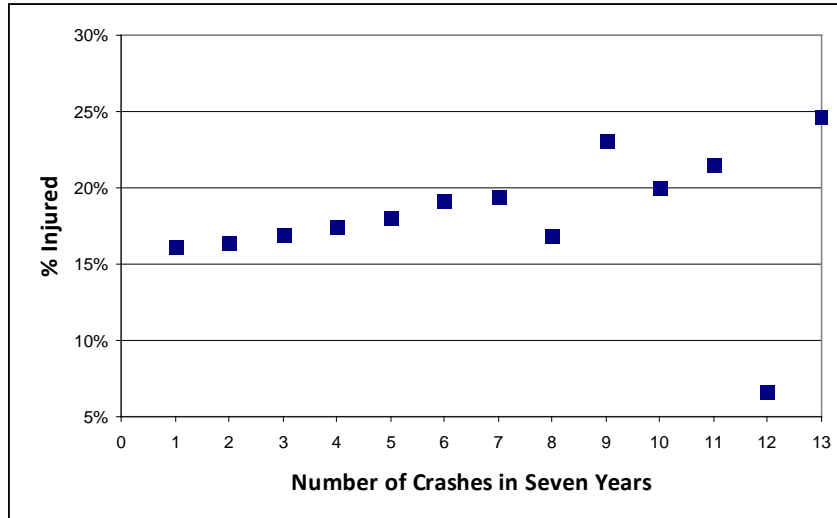


Figure 1: Distribution of injury crashes

The initial data analysis reveals a big variation in crash frequency among licensed drivers in Louisiana. A driver's crash risk in the coming year is closely related to the driver's safety performance. Based on a driver's crash history, we can predict the probability of having a crash or crashes in the coming year. For that purpose, two crash matrixes were developed, with each row presenting i crashes in the previous six years and each column presenting j crashes in the year 2010. Each cell in the first matrix presents the probability of having j crashes in the seventh year and i crashes in the previous six years $P(j \cap i)$. The summation of each column and row represents the probabilities $P(i)$ and $P(j)$. The second matrix consists of conditional probability $P(j/i)$ that is calculated by:

$$P(j/i) = \frac{P(j \cap i)}{P(i)} \quad (1)$$

16th Road Safety on Four Continents Conference
Beijing, China 15-17 May 2013

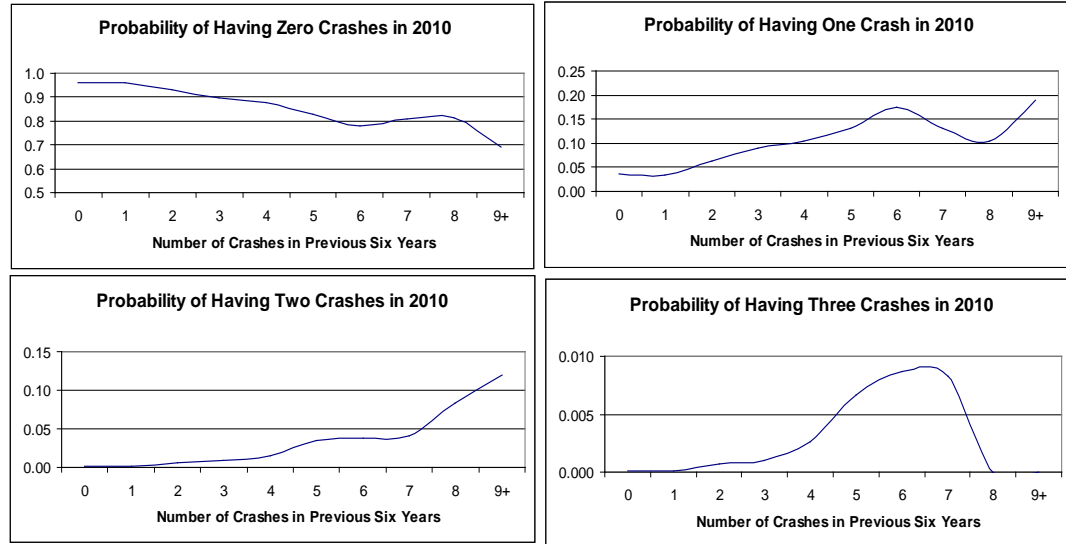


Figure 2: Probabilities of drivers having zero, one, two and three crashes

With the second conditional probability matrix, the probability of having any given number of crashes, j , in 2010 varies by drivers' history as illustrated in Figure 2.

Thus, the probability of having crash(s) in the seventh year can be estimated as:

$$P(j_7 \geq 1/i) = \sum_{j=1}^m \frac{P(j \cap i)}{P(i)} \quad (2)$$

for $i = 0, 1, 2, \dots, 9^+$

Based on this equation, how the probability of having crash(es) varies is illustrated in Figure 3.

16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013

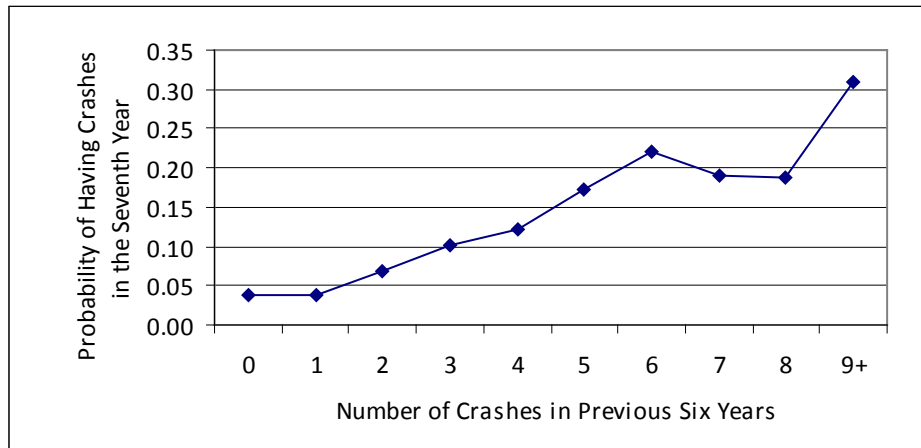


Figure 3: Conditional probability of having crash(es) in 2010

It is expected that drivers with no crashes during previous six years have the lowest probability of having crash(es) in the seventh year as shown in Figure 3 and the probability of having zero crashes decreases as drivers' past crash frequency increases as shown in Figure 2. The probability of having one or two crashes in the seventh year increases as the crash frequency increases in previous years. The probability curve of having three crashes in the seventh year is different; it's increasing first and then decreases as the number of previous crashes increases.

It is interesting to know that these multiple crashes did not occur uniformly along the seven-year time period. The crash data is used here to investigate the recurrent time interval or time gap, the time between a previous traffic crash and involvement in a newer one. The gap between crashes is much smaller than the uniformly distributed gaps. As displayed in Figure 4, the estimated average gap between crashes is two years for drivers who had two crashes in the seven-year time period, which is much smaller than the uniform 3.5 years. The estimated average gap between crashes is 0.46 compared to the uniform gap of 0.54 for drivers having 13 crashes in seven years.

16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013

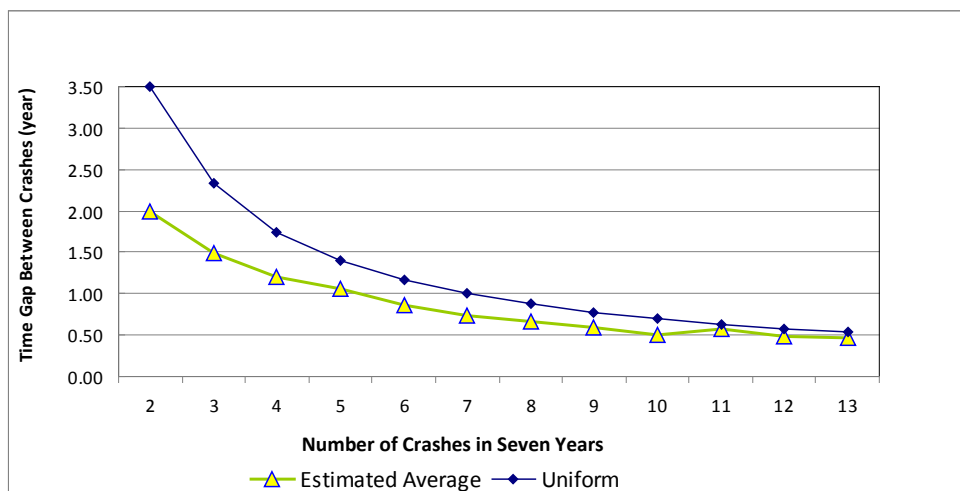
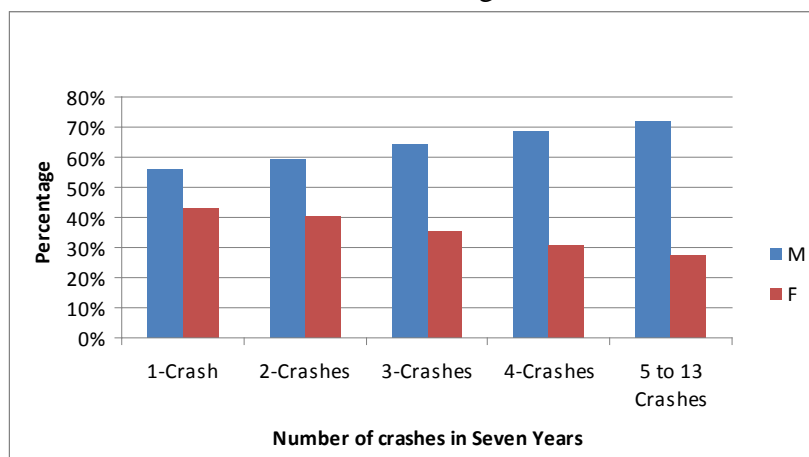


Figure 4: Time gap between crashes

Crash-prone drivers apparently present a significant adverse effect on highway safety. Investigating crash-prone drivers' characteristics can help to develop effective safety strategies targeting these drivers. Through the data analysis, we found the following crash characteristics are worth noting. First, the relative difference between genders by number of crashes in seven years was explored with the results shown in Table 2, which indicates the percentage of male driver increases as the number of crashes increases. About 71% of drivers with five or more crashes in seven years are male drivers while the percentage is 55% in one crash case as shown in Figure 5.



16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013

Figure 5: Distribution of drivers' gender

The drivers' age distribution shown in Figure 6 illustrates the differences among different age groups. One national study (10) has indicated that the very young and the very old have the highest fatal crash rate. This study reveals that the 20 to 40 age group has the highest crash involvement rate and their rate increases as the number of crashes increases. Comparing with another age group (40 to 65 years old drivers), the over-involvement in multiple crashes by the 20-40 age group is more evident.

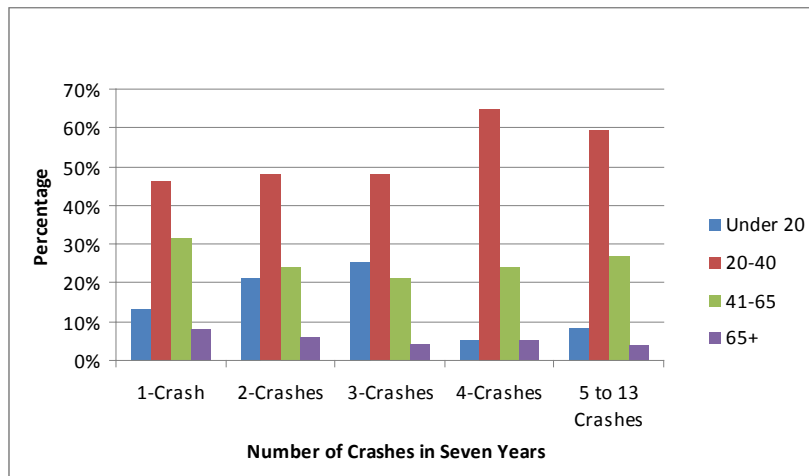


Figure 6: Distribution of drivers' age

To answer the question of what are major contributing factors to these multiple crashes in seven years, we examined drivers' condition, type of violations and distractions, and type of crashes. The box-and-whisker plot (Figure 7) displays that distractions involved with drivers in multiple crashes (2 or more crashes in seven years' crash record) are younger in mean ages than the drivers in a single crash in seven years. This figure clearly indicates that distraction related events are predominant in younger crash-prone drivers.

16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013

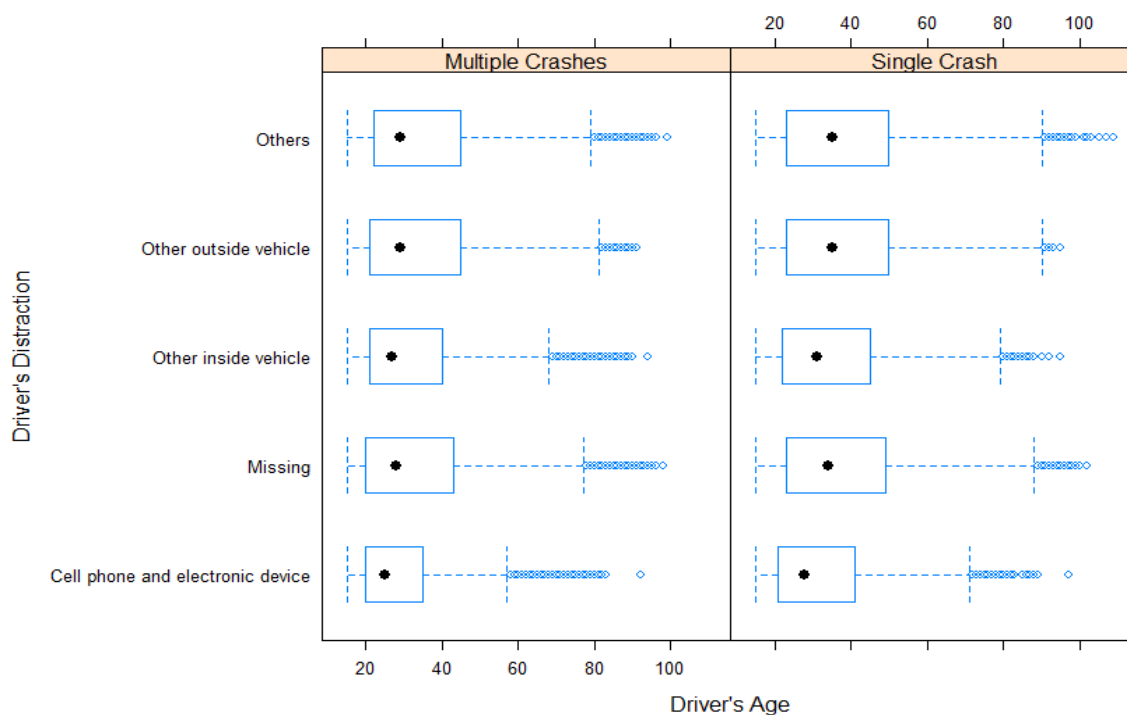
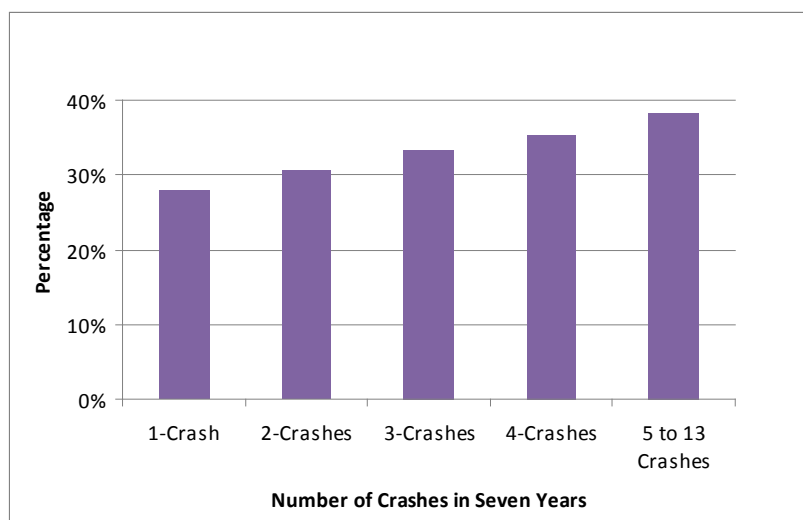


Figure 7: Relationship between distraction and driver's age

Regarding the type of violation, careless operation was cited as the main reason for the crash and it increases as the number of crashes increase as shown in Figure 8.



16th Road Safety on Four Continents Conference Beijing, China 15-17 May 2013

Figure 8: Careless operations vs. percentage of crashes

Rear-end collision is the most common type of crash regardless of crash frequency level. However, the percentage of rear-end collision does go up as number of crashes increase. As displayed in Figure 9, the proportion of single-vehicle crashes (non-collision with motor vehicle) also increases as number of crashes increases.

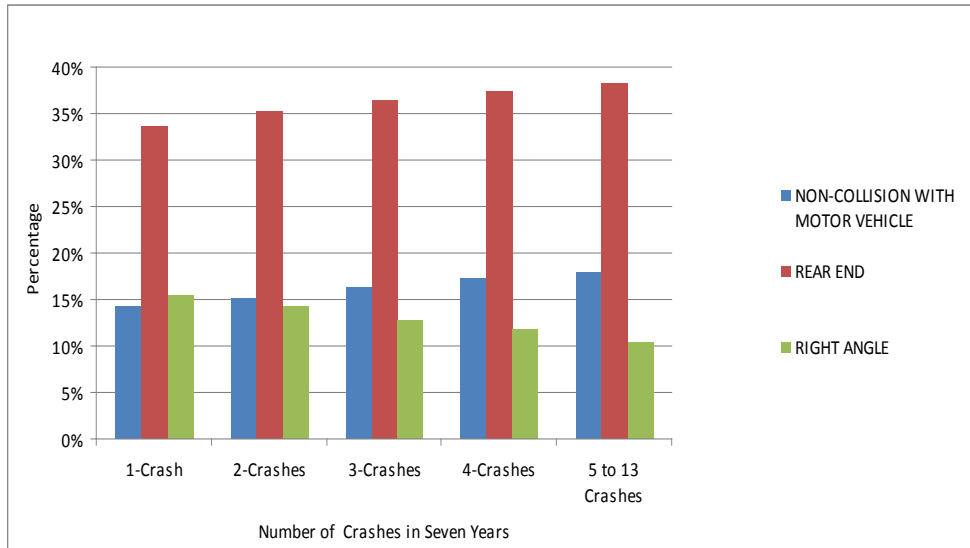


Figure 9: Type of crashes vs. percentage of crashes

4 DISCUSSION AND CONCLUSION

The seven year crash data analysis introduced in this paper has demonstrated that crash-prone drivers need to be carefully targeted in safety education and traffic law enforcement programs because their over-involvement in crashes presents a big adverse effect on roadway safety. The study results quantitatively confirm that 5% of drivers in Louisiana are responsible for 35% of crashes in a seven-year time period. The probability of having crash(es) in the coming year for drivers with a crash history is more than seven times higher than the probability for drivers with zero crashes. These crash-prone drivers are likely to be male in the 20-40 age group.

Drivers with frequent crash history should be targeted for special safety programs regularly through education and regulations. For instance, a state motor vehicle registration office could work with the enforcement agencies to establish a driver's license reviewing program that has authority to send warnings or to suspend a driver's license, or to request the driver take a mandatory safety class if the driver has had multiple crashes within a short time period. These targeted safety classes should focus on distracted driving and even if these multiple crashes are not severe, our analysis shows that the next crash will most likely be severe if not fatal.

16th Road Safety on Four Continents Conference

Beijing, China 15-17 May 2013

The crash-prone driver problem might be worse than is revealed in this study because of the removed datasets for drivers with incomplete information. The drivers with no valid license (suspended license) or a fake license tend to be the drivers with frequent crashes and traffic law violations.

This paper presents the initial results of a series of studies on crash-prone drivers' characteristics and individual drivers' crash risk modeling. The on-going and future analysis will be extended to non-at-fault drivers involved in a crash but not responsible or less responsible for the crash. We hope that the extended analysis could shed more light on what, how, and why these multiple crashes occur to a single driver over a short time period so that more effective crash-preventive measures can be proposed to reduce recurring crashes. The analysis will also look into the crash history of drivers with a fatal crash in the last year of the analysis time period and their time gap distribution between crashes.

REFERENCES

1. Greenwood, M., Yule, U. (1920). *An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents*. Journal of Royal Statistical Society, Vol. 83 (2), 255–279.
2. Blasco, R.D., Prieto, J.M., Cornejo, J.M. (2003). *Accident probability after accident occurrence*. Safety Science, Vol. 4 (6), 481–501.
3. Peck, R.C., McBride, R.S., Coppin, R.S. (1971). *The distribution and prediction of driver accident frequencies*. Accident Analysis & Prevention, Vol. 15, 371–393.
4. Stamatiadis, N., Agent, K.R., Pigman, J., Ridgeway, M. (1999). *Evaluation of retesting in Kentucky's driver license process*. Research Report KTC-99-Kentucky Transport Cabinet.
5. Daigneault, G., Joly, P., Frigon, J. (2002). *Previous convictions or accidents and the risk of subsequent accidents of older drivers*. Accident Analysis & Prevention, Vol. 34, 257–261.
6. Hauer, E., Persaud, B.N., Smiley, A., Duncan, D. (1991). *Estimating the accident potential of an Ontario driver*. Accident Analysis & Prevention, Vol. 23 (2/3), 133–152.
7. Chen, W., Cooper, P., Pinili, M. (1995). *Driver accident risk in relation to the penalty point system in British Columbia*. Accident Analysis & Prevention, Vol. 26 (1), 9–18, 1995.
8. Gebers, M.A. (1999). *Strategies for estimating driver accident risk in relation to California's negligent-operator point system*. California Department of Motor Vehicles Research and Development Branch, Technical Monograph 183.
9. Gebers, M.A., Peck, R.C. (2003). *Using traffic conviction correlates to identify high accident-risk drivers*. Accident Analysis & Prevention, Vol. 35, 903–912.



16th Road Safety on Four Continents Conference
Beijing, China 15-17 May 2013

10. National Highway Traffic Safety Administration (NHTSA) (2009). *An Examination of Driver Distraction as Recorded in NHTSA Databases*. Traffic Safety Facts, DOT HS 811 216, September.