

CRASH RISK ANALYSIS OF DISTRACTED DRIVING BEHAVIOR: INFLUENCE OF SECONDARY TASK ENGAGEMENT AND DRIVER CHARACTERISTICS

Osama A. Osman¹, Mengqiu Ye¹, and Sherif Ishak
Louisiana State University
Baton Rouge, Louisiana, United States
Email: Othabel@lsu.edu

Summary: Distracted driving has long been acknowledged as one of the main contributors to crashes in the US. According to past studies, driving behavior proved to be influenced by the socioeconomic characteristics of drivers. However, only few studies attempted to quantify that influence. The study proposed a Crash Risk Index to estimate the crash risk associated with the socioeconomic characteristics of drivers and their tendency to experience distracted driving. The analysis is conducted using data from the SHRP 2 Naturalistic Driving Study (NDS). The proposed Crash Risk Index (CRI) is developed based on a grading system of three measures: the crash risk associated with performing secondary tasks during driving, the effect of socioeconomic attributes (e.g. Age) on the likelihood of engagement in secondary tasks, and the effect of specific categories within each socioeconomic attribute (e.g. Age>60) on the likelihood of engagement in secondary tasks. Logistic Regression analysis was performed on the secondary tasks, socioeconomic attributes, and the specific socioeconomic characteristics. The results identified the significant secondary tasks with high crash risk and the socioeconomic characteristics with significant effect on determining drivers' involvement in secondary tasks among each tested parameter. These results were used to quantify the grading system measures and hence estimate the proposed CRI. This index indicates the relative crash risk associated with the socioeconomic characteristics of drivers and considering the possibility of engagement in secondary tasks. The proposed CRI and the associated grading system are plausible methods for estimating auto insurance premiums.

INTRODUCTION

Distracted driving has captured the attention of many researchers and transportation officials due to its significant impact on traffic safety. Several studies showed that distracted driving is likely to increase the reaction time of drivers and their response time (Harbluk et al. 2007). When analyzing the impact of specific secondary tasks, studies have shown that: (a) talking on a handheld cellphone impairs the drivers' ability to maintain their speed and position on the road (Narad et al. 2013); (b) texting increases braking reaction times and increases lane-position variability with no change in speed (Hosking et al. 2005). In another study by Klauder et al. (2014), they investigated the crash risk associated with performing secondary tasks. The results indicated that crash risk significantly increased for novice drivers when they were dialing a cellphone, texting, reaching for objects, looking at roadside objects, and eating. On the other

¹ Osman O. A. and Mengqiu Ye Contributed Equally to this Work.

hand, for experienced drivers, the crash risk increased significantly only when drivers were dialing cellphone.

According to Elander et al. (1993), the unsafe driving behavior is a type of driving style that is developed over time. This unsafe driving behavior becomes a habit that differs from one driver to another according to some socioeconomic characteristics (Elander et al. 1993). Based on a detailed survey of 834 licensed drivers, Poysti et al. (2005) concluded that younger and male drivers tend to use phones more often compared to older and female drivers. The survey also showed that driving for longer distances increases the likelihood of cellphone use. More so, people tend to use cellphones more often when they perceive themselves as skilled drivers. Based on a survey conducted by Stayer et al. (2006), most drivers may not be aware of their impaired driving behavior while engaged in distracted driving. This study suggested that the inability to correctly perceive own driving skills and actual behavior is influenced by socioeconomic characteristics.

The literature shows the significance of driver's engagement in secondary tasks in determining roadway safety. Past research also shows that socioeconomic attributes have a significant impact on the crash risk and the likelihood of performing secondary tasks. To the authors' knowledge, there is no previous published work on quantification of the effect of the socioeconomic characteristics on drivers' involvement in secondary tasks and crash risk. Thus, this study uses the SHRP 2 Naturalistic Driving Study (NDS) data to develop an approach (grading system) to quantify the crash risk associated with the driving behavior based on drivers' socioeconomic characteristics. The study then develops a Crash Risk Index measure based on that grading system.

CRASH RISK INDEX DEVELOPMENT

The proposed research in this study is performed in three main steps: (a) extract data for all secondary tasks and socioeconomic attributes from the NDS database; (b) select the records with high crash risk and significant effect on drivers' involvement in secondary tasks; and (c) develop a grading system and a Crash Risk Index for quantification of crash risk.

Extracting Secondary Tasks and Socioeconomic Attributes

The NDS data provides detailed demographic and history questionnaires for each participant of the study. The demographic questionnaires provide information on the participant's personal background, while the history questionnaires provide information on their driving record. The questionnaires were used to extract several socioeconomic attributes including Age, Gender, Marital Status, Work Status, Average Annual Miles Travelled, Years (has been) Driving, Annual Household Income, Education Level, Vehicle Classification, and the State (participants' location). Although some of these attributes may not be considered socioeconomic (e.g. Vehicle Classification, gender ... etc.), they were included in the analysis as socioeconomic attributes for abbreviation purposes. To consolidate the data, each attribute was classified into socioeconomic categories based on the common practice in the literature.

The NDS data provided more than 24,000 events with unique IDs, each assigned and linked to a participant in the questionnaires. This dataset included details on the event severity and the associated secondary task for drivers prior to each event. Event severity is classified as either crash (1488), near crash (2997), or baseline (19672 events). Baseline events are those with no crashes or near-crashes. More than 60 secondary tasks were recorded including cellphone use, texting, personal hygiene, reaching for an object, and no-secondary task among many others. Since this study focuses on crash risk assessment, the different secondary tasks and socioeconomic attributes were extracted for the 1488 crash events only.

Selection of Secondary Tasks and Socioeconomic Attributes

Secondary Tasks Selection. This study focuses on identifying the crash risk associated with drivers’ socioeconomic characteristics based on their tendency to perform secondary tasks. As such, only secondary tasks with significantly high crash risk were selected. In addition, only secondary tasks with high frequency were included to select a large enough sample size for each socioeconomic attribute. Thus, secondary tasks with a frequency lower than 90 were removed. For each high-risk secondary task, the socioeconomic attributes with a significant effect on drivers’ involvement in this secondary task were identified.

To identify the secondary tasks with high crash risk, Logistic Regression analysis was performed on the categorical data at 5% level of significance. Event severity, represented by 1 for crashes and 0 for baseline, was treated as a dependent variable, while the secondary tasks drivers were involved in were treated as independent variables. As recommended by Young et al. (2015), the no-secondary task class was used as a control group since crashes may also occur with no secondary tasks associated. Based on the corresponding frequencies for the control group, the output of the analysis was whether this secondary task significantly affected the crash risk or not. The p-values and odds ratios were obtained for each secondary task, as shown in Table 1. If the resulting p-value is less than 0.05 and the odds ratio is greater than 1, this secondary task is classified as high crash risk. The odds ratio is calculated by eq (1) with the control group defined as no-secondary task. Moreover, the odds ratios can be used to quantify the crash risk associated with this secondary task.

$$\text{odds ratio of crashes involving a secondary task} = \frac{\text{odds of crash events involving a secondary task}}{\text{odds of crash events not involving that secondary task}} \quad (1)$$

Table 1. Logistic regression analysis results for high crash risk secondary tasks

<i>Type of secondary task</i>	<i>Total Frequency</i>	<i>Pr > ChiSq (crash risk)</i>	<i>Odds ratio (crash risk)</i>
No Secondary Tasks	10028	-	1.00
<i>Adjusting/monitoring other devices integral to vehicle</i>	133	<.0001	4.47
<i>Cellphone, holding</i>	366	<.0001	2.19
<i>Cellphone, texting</i>	387	<.0001	2.77
<i>Looking at an object external to the vehicle</i>	143	<.0001	6.01
<i>Object in vehicle, other</i>	230	0.0408	1.61
<i>Other personal hygiene</i>	256	0.0001	2.13
<i>Reaching for object, other</i>	95	<.0001	10.22

The results show that only seven secondary tasks had significant high crash risk. As shown in Table 1, the odds ratios indicate that Reaching for Object had the highest crash risk with an odds value of 10.22, while Object-In-Vehicle had the lowest crash risk with a value of 1.61. According to the NDS definitions, Object-In-Vehicle task means “the driver clearly is looking at, handling, holding, or manipulating an object (visible or not) or something located in the vehicle”, and the object does not include a moving object, insect, pet, passenger, or child. Adjusting/Monitoring-Other-Devices task is adjusting devices other than climate and radio control. Personal Hygiene includes but not limited to “checking oneself in mirror without the preceding tasks, or trying to get something out of one's eye”. Reaching for Object is reaching for any object except for food, drink, cellphone, cigarettes or personal body-related item.

Selection of Socioeconomic Attributes. For secondary tasks with high crash risk, the socioeconomic attributes with significant association to drivers’ involvement in that task were determined. Logistic Regression with Type III analysis in the SAS GENMOD statistical procedure was performed at 5% level of significance. For each secondary task, the analysis was performed considering the likelihood of drivers’ involvement in this secondary task, represented by 1 for the secondary task and 0 for no secondary task, as a dependent variable, and the various socioeconomic attributes as independent variables. The p-values obtained from this analysis determine whether a specific socioeconomic attribute has a significant association with the likelihood of drivers’ involvement in the secondary task. Based on the resulting p-values, each attribute is described as either Significant (p-value < 0.05) or Insignificant. Insignificant attributes were then excluded from the rest of the analysis. The p-values associated with the socioeconomic attribute for each of the seven secondary tasks with high crash risk are presented in Table 2. The results show that none of the socioeconomic attributes have significant association with engagement of activities such as Adjusting Devices, Looking at Object, and Object in Vehicle. Therefore, these secondary tasks were removed from the rest of the analysis. For the remaining four secondary tasks, some of the attributes were significant as highlighted in bold in Table 2. For instance, Cellphone Texting has three significant attributes including Age, Annual Miles and State. This means that drivers’ involvement in cellphone texting is affected by their age, the number of annual miles, and the State of residence. The results also show some socioeconomic attributes with no significant association with the likelihood of drivers’ involvement in secondary tasks. These attributes include Work Status, Years Driving, Education, and Vehicle Class. Since they do not have any effect on determining drivers’ involvement in secondary tasks, these attributes were not considered further in the analysis.

Table 2. P-values for socioeconomic attributes’ effect on each secondary task

<i>Secondary Task</i>	<i>Socioeconomic Attribute</i>									
	Age	Gender	Marital Status	Work Status	Annual Miles	Years Driving	Income	Education	Vehicle Class	State
<i>Adjust Device</i>	0.504	0.8222	0.0944	0.7337	0.1934	0.2936	0.8632	0.4212	0.7413	0.1626
<i>Cellphone, Texting</i>	<.0001	0.4033	0.2839	0.2579	<.0001	0.1261	0.0991	0.6736	0.7155	0.0004
<i>Cellphone, Holding</i>	0.0114	0.0317	0.0007	0.6415	0.174	0.3553	0.81	0.0512	0.4689	<.0001
<i>Looking At Object</i>	0.5807	0.3516	0.8864	0.1589	0.953	0.6702	0.1458	0.2426	0.7506	0.2466
<i>Object In Vehicle</i>	0.3214	0.7443	0.2197	0.0846	0.4125	0.2741	0.1135	0.7068	0.2671	0.8702
<i>Personal Hygiene</i>	0.3539	0.8935	0.0751	0.2931	0.0038	0.9149	0.0892	0.2395	0.7988	0.7393
<i>Reaching For Object</i>	0.0798	0.9753	0.8244	0.1647	0.7019	0.8114	0.0480	0.7683	0.0556	0.3104

Grading System and Crash Risk Index

In this section, a grading system is developed to quantify the crash risk associated with drivers based on their socioeconomic characteristics. This is accomplished in three consecutive steps performed at three different levels as shown in Figure 1. First, the crash risk associated with each secondary task is quantified. Then, the effect of the different socioeconomic attributes on the likelihood of performing that secondary task is quantified. Finally, the effect of each socioeconomic attribute is broken down by the different categories within that attribute. This is to determine the relative effect of each category for a socioeconomic attribute on the likelihood of performing the secondary task in step 1. The effects in the three steps are labeled as (a) Crash Risk Coefficient for the first step, (b) Significance Level Coefficient for the second step, and (c) Category Contribution Coefficient for the third step.

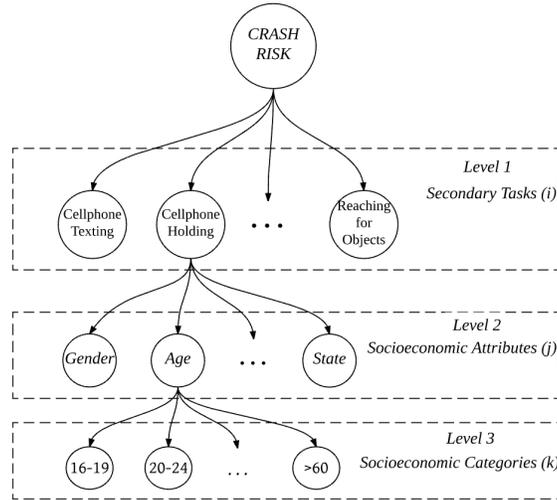


Figure 1. Crash risk quantification tree

Step 1: The Crash Risk coefficient (R_i) was calculated based on the odds ratios obtained for each task (i) shown in Table 1. For each secondary task, a higher odds ratio means higher crash likelihood while performing this secondary task. In order to quantify the relative crash risk for each secondary task, the odds ratios for the different secondary tasks were normalized relative to the secondary task with the highest odds ratio. Based on this normalization, the coefficient R_i takes values from 0 to 1, with 1 denoting highest crash risk.

Step 2: The Significance Level Coefficient (a_{ij}) quantifies the effect a socioeconomic attribute (j) has on the likelihood of drivers' involvement in secondary task (i). To quantify the level of significance of each socioeconomic attribute, chi-square values were obtained as part of the Type III analysis. Chi-square values quantify the effect a Significant Attribute may have on drivers' involvement in a specific secondary task. For each attribute, this coefficient is calculated as the ratio between Chi-square value for a specific Significant Attribute divided by the summation of the chi-square values for all Significant Attributes as in eq (2)

$$a_{ij} = \frac{\text{chi-square for Significant Attribute } (j)}{\sum_j \text{Chi-square values for the Significant Attributes in Secondary Task } (i)} \quad (2)$$

Step 3: The Category Contribution coefficient (b_{ijk}) is measured for the category (k) within each Significant Attribute (j). To identify the crash risk from socioeconomic characteristics, the relative significance of each socioeconomic category (Table 3) is determined. In other words, the contribution of each category (k) within an attribute (j) to the likelihood of drivers' involvement in secondary task (i) is quantified. This is determined using the odds ratios of each category (k) within a socioeconomic attribute (j). The odds ratios were obtained by dividing the odds of performing a secondary task (i) by the socioeconomic category (k) by the odds of performing the same secondary task by a control category as shown in the example in eq (3). The bold-faced categories in Table 3 are considered control groups for each attribute.

$$\text{odds ratio of married people}|_{\text{texting}} = \frac{\text{odds that married people perform texting}}{\text{odds of a control category perform texting}} \quad (3)$$

To calculate the coefficient b_{ijk} , the odds ratios determined using eq (3) are normalized based on the category with the highest odds ratio. Similar to the Crash Risk coefficient, the Category Contribution takes values between 0 and 1, where 1 denotes the highest contribution to the significance level of a socioeconomic attribute.

Table 3. List of categories within each socioeconomic attribute

Socioeconomic Attribute	Number of categories	List of Categories
Age	10	16-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, >60
Gender	2	Female, Male
Marital Status	5	Divorced, Married, Single, Unmarried-Partners, Widow(er)
Average Annual Miles Travelled	7	< 5000, 5000 – 10000, 10000 - 15000, 15000 – 20000, 20000 – 25000, 25000 – 30000, > 30000
Annual Household Income	7	\$150000 +, \$100000 - \$149999, \$70000 - \$99999, \$50000 - \$69999, \$40000 - \$49999, \$30000 - \$39999, Under \$29000
State	6	Florida, Indiana, North Carolina, New York, Pennsylvania, Washington

These three coefficients represent the grading system for the crash risk associated with performing the different secondary tasks based on people's socioeconomic characteristics. The three coefficients can be used to quantify the crash risk associated with people's tendency to conduct secondary tasks based on their socioeconomic characteristics. This crash risk can be measured as a Crash Risk Index (CRI) that is calculated using equation (4).

$$CRI_i = R_i * \sum_{j \in s(i)} (a_{ij} * b_{ijk}) \quad (4)$$

Where, $s(i)$ is the set of attribute (j) with significant influence on the likelihood of performing secondary task (i) and CRI_i is the CRI value associated with the high-crash-risk secondary task (i). CRI is calculated for each secondary task individually. To calculate the overall CRI value associated with all high-crash-risk secondary tasks, equation (5) is used.

$$CRI_o = \sum CRI_i \quad (5)$$

SUMMARY AND CONCLUSION

This study provided a methodology to quantify the crash risk associated with driving behavior based on drivers' socioeconomic characteristics. The NDS data is used to develop a reliable

approach to evaluate the crash risk associated with driving behavior and investigates the relationship between the socioeconomic characteristics and crash risk. A Crash Risk Index (CRI) is proposed based on a grading system composed of three main measures: the crash risk associated with performing secondary tasks while driving, the effect of socioeconomic attributes on the likelihood of engagement in secondary tasks, and the effect of specific categories within each socioeconomic attribute on the likelihood of drivers being involved in secondary tasks.

Logistic Regression analysis is performed to study the crash risk associated with secondary tasks. The analysis also investigates the effect of socioeconomic attributes (Age, Gender ... etc) on the likelihood of drivers' engagement in secondary tasks. In addition, the analysis investigates the effect of specific socioeconomic categories (Age = 16-19, 20-24,; Gender = Male, Female; ... etc) on the likelihood of drivers' engagement in secondary tasks. Based on the analysis results, the three measures in the proposed grading system are quantified and used to determine the Crash Risk Index associated with the socioeconomic characteristics of drivers. The developed Crash Risk Index and grading system were discussed in detail in the paper. The CRI values give an indication of the relative crash risk associated with the socioeconomic characteristics of drivers and considering the possibility of engagement in secondary tasks. The proposed CRI measure and the associated grading system may assist the auto insurance industry in the process of underwriting policies and determining insurance premiums. The CRI measure also represents a promising tool for policy makers and DOT officials to develop safety programs for mitigating drivers' involvement in risky secondary tasks.

REFERENCES

- Elander, J., West, R., & French, D. (1993). Behavioral correlates of individual differences in road-traffic crash risk: An examination of methods and findings. *Psychological bulletin*, 113(2), 279.
- Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, 39(2), 372-379.
- Hosking, S. G., YOUNG, K., & Regan, M. A. (2005). *The effects of text messaging on young novice driver performance*.
- Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England journal of medicine*, 370(1), 54-59.
- Narad, M., Garner, A. A., Brassell, A. A., Saxby, D., Antonini, T. N., O'Brien, K. M., Epstein, J. N. (2013). Impact of distraction on the driving performance of adolescents with and without attention-deficit/hyperactivity disorder. *JAMA pediatrics*, 167(10), 933-938.
- Pöysti, L., Rajalin, S., & Summala, H. (2005). Factors influencing the use of cellular (mobile) phone during driving and hazards while using it. *Accident Analysis & Prevention*, 37(1), 47-51.
- Strayer, D. L., Drews, F. A., & Crouch, D. J. (2006). A comparison of the cell phone driver and the drunk driver. *Human factors: The journal of the human factors and ergonomics society*, 48(2), 381-391.
- Young, R. (2015). *Revised odds ratio estimates of secondary tasks: A re-analysis of the 100-Car naturalistic driving study data (0148-7191)*.