

RESEARCH FRAMEWORK FOR STUDYING DRIVER DISTRACTION ON POLISH CITY HIGHWAYS

Kelwyn A. D'Souza¹, Sharad K. Maheshwari¹, Zbigniew A. Banaszak²

¹ Hampton University, Hampton (VA), U.S.A.

² Warsaw University of Technology, Warsaw, Poland

Corresponding author:

Kelwyn A. D'Souza

Eastern Seaboard Intermodal Transportation Applications Center (ESITAC)

Hampton University, Hampton

VA 23668, U.S.A.

phone: (757) 727-5037

e-mail: kelwyn.dsouza@hamptonu.edu

Received: 25 June 2012

Accepted: 5 January 2013

ABSTRACT

Analysis of accidents has found driver distraction to be a significant cause of accidents on the highways [1]. Therefore, studying the causes of driver distraction that impact its risk level is needed for a better understanding of accident occurrences. There is general scarcity of research in this field with no established research framework to study driver distractions. This paper proposes a modular research framework for conducting a driver distraction study on Polish city highways. The framework contains guidelines for distraction studies for wide range of cost and time intervals such as a quick, low cost study like analysis of existing accident databases maintained by the cities to relatively higher cost, longer duration study involving field data collection, statistical modeling, and analysis. A city may choose one or more modules to suit their study requirements including statistical and simulation tools to assess and validate the historical or empirical result. The framework is based on the careful modifications and revisions of an earlier transit bus driver study conducted in the Commonwealth of Virginia, U.S.A., and results from this research are presented for purposes of illustration.

KEYWORDS

accident data analysis, modeling driver distraction, multinomial logistic regression, predicting driver distraction risk, Distraction Risk Index, model validation.

Introduction

The analysis of accident databases has found driver distraction to be a significant cause of accidents on the highways [1]. Hence, the study of driver distraction has grown into an active area of research in the U. S. and other countries as noted from the volume of articles in the transportation safety-related literature [2–6]. Furthermore, the U. S. Government encourages research in this field through the Transportation Research Board and other federal agencies [7].

The impact of technology (e.g., mobile phones and route guidance systems) and non technology-based distractions (e.g., eating, smoking and conversing with passengers) on driving performance should

be examined in order to evaluate the relative influence of these distractions on driving process. Studying the causes of driver distraction and the factors that impact its risk level is needed for an overall understanding of accident occurrences. This paper presents a modular research framework for data collection, analysis, validating, and interpreting results of a driver distraction study. This would help Polish city transportation departments to better comprehend the impact of distracted driving on the delivery of service. Furthermore, it will establish a standardized driver distraction study framework from data collection to result interpretation and application. The objectives are to provide any city with a set of standardized methodologies for studying driver distraction. The Data Collection module consists of

tools for performing an accident database analysis, a survey instrument, and route observations forms. The Analysis module will show how to classify distracting activities, how to develop statistical models that construct relationships between high risk distracting activities and driver characteristics and external factors. The Validation module presents simple observation and discussion methods to sophisticated simulation techniques to check the model results. The final module contains guidelines for Results Interpretation and Usage.

The modular framework offers a city the flexibility of choosing one or more modules for conducting a driver distraction study. Understanding the driver characteristics and external factors that could cause distraction can help to develop effective policies to mitigate risk of accidents. The various components necessary for studying the sources and duration of driver distractions, the risks associated while engaging in potential distracting activities, and visual, manual, and cognitive factors that are believed to be responsible for distraction will be combined together to form the structure of the framework. A city could use these inputs to classify the distracting activities into different risk zones. The distracting activities in high risk zones that pose high safety concerns could be further analyzed using statistical models to quantify the impact of various external factors on driver distraction. Cities will have the option of validating the results using methods like simulation and route observations.

The proposed framework is based on an earlier study conducted on a transit agency in the Commonwealth of Virginia located in the southeastern United States [2–4]. Some of these results are reproduced in this paper for purposes of illustrating the type of outputs obtainable from the framework. These outputs can be modified to suit the transportation safety problems of the U. S. and other countries. For example, Poland has one of the highest traffic death rate per million inhabitants of 146 and 43 per billion motor vehicle km which is more than three times the European Union's average of 13 per billion motor vehicle km [8]. Hence, more detailed output reports are necessary. Finally, the standardized techniques for studying driver distraction can help improve driver training, adaptation of technology, design of driver cabin and dashboard, and development of policies that would help mitigate the accident rates.

Background

Driver distraction represents a significant problem in the personal and public transport sector and

has been studied by several national and international researchers. A study funded by the AAA Foundation [9] identified the major sources of distraction for personal vehicles contributing to crashes, developed taxonomy of driver distractions for the U.S. driving population, and examined the potential consequences of these distractions on driving performance. The source of bus driver distractions at a major Australian public transport company was investigated using ergonomics methods through which, a taxonomy of the sources of bus driver distraction was developed, along with countermeasures to reduce their effects on driver performance ([6]. In an earlier study, Salmon et al. [10] developed taxonomy of distraction sources and duration for bus drivers at the State Transit Authority New South Wales, Australia, but provided insufficient inferential statistical analysis. D'Souza and Maheshwari [2–4] expanded the work of Solmon et al. [10] using multivariate statistical models and simulation to determine the impact of driver and external factors on distracting activities.

Factors such as location, driving hours/week; and driver age, gender, and experience have an impact on public bus driver distraction [11]. A driving route running through a densely populated area would service a greater number of passengers and experience a higher external sources of distraction due to more frequent stops and more other road users or pedestrians [9]. A driver less familiar with the driving routes is more likely to be involved in rear-end accidents at signalized intersections [12]. Studies on the impact of age, gender, driving experience, and driving demands on driving performance suggests that younger (below 25 years) and older (above 70 years) drivers tend to be more vulnerable to the effects of distraction than middle-aged drivers [7, 13]. Blower et al. [14] reported that age, sex, hours driving, trip type, method of compensation, and previous driving records are related to driver errors.

Multivariate statistical models are widely used in transportation to study the relationship between the categorical dependent variable and a set of continuous and categorical independent predictor variables. The multivariate model applied by Yan et al. [12] to study accidents in trucks, identified driver age and gender among the several other factors related to rear-end crashes. Washington et al. [15] developed a multinomial logit (MNL) model consisting of 18 independent variables covering driver factors, traffic flow, distance, and number of signals etc. to study factors that influence drivers' selection of route on their morning commute to work. Yan et al. [16] utilized multinomial logistic regres-

sion (MLR) to study the impact of potential factors such as driver factors, road layout, and environmental conditions on rear-end truck to car, car to truck, and car to car crashes. A MLR model was developed by Morfoulaki et al. [5] to identify the factors contributing to service quality and customer satisfaction (*very satisfied, satisfied, somewhat dissatisfied, and very dissatisfied*) with a public transit service in Greece. Gkritza et al. [17] conducted an empirical study using multinomial logit models to investigate the socio-economic and demographic factors that significantly affect passenger satisfaction with airport security screening process. Petrucci [18] computed the odds ratios for the tasks/variables, along with 95 % confidence intervals (CI) to identify the high risk tasks/variables and the strength of association between the categorical dependent variable and independent variables.

The Monte Carlo simulation method is commonly applied to validate empirical results obtained from conceptual models. Carlson [19] demonstrated the application of Monte Carlo simulation to evaluate proposed component changes on highway crash reduction. The impact of age and cognitive functions on driving performance has been studied extensively to predict cognitive distraction with a computational cognitive model and validating the results through simulation [20].

Researchers have developed methodologies for assessment of transit bus driver distraction which includes the analysis of tasks, identification of distraction sources, and risk assessment [11]. Wong and Huang [21] have proposed a research framework for studying driver's mental process in order to determine how accidents occur. It includes a conceptual framework of driving mental process which is a step towards development of a workable model that can be used to study accident causality. Trick et al. [22] have provided a conceptual framework that combines the two fundamental dimensions of attention selection in order to have a more comprehensive driving theory. Although the work of Wong and Huang [21] and Trick et al. [22] are not directly related to driver distraction, their framework provides useful inputs for development of the research framework in this paper.

Outline of the research framework

An outline of the proposed research framework is presented in Fig. 1. It consists of four modules to study the driver distraction on a city's highway: Data Collection, Analysis, Validation, and Guidelines for Results Interpretation and Usage.

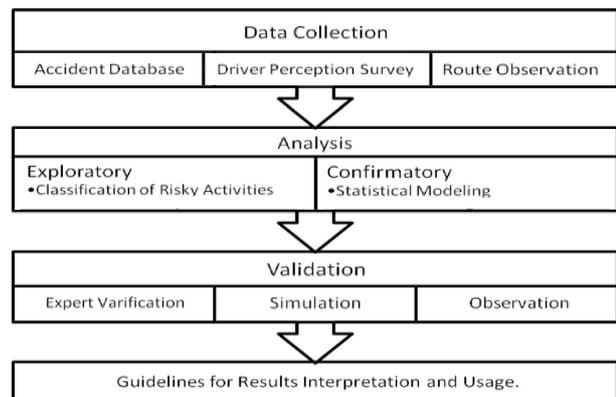


Fig. 1. Outline of the proposed research framework.

The primary research questions addressed by this framework are:

- What are the common sources and duration of distraction?
- Are present driver characteristics such as location, age, experience, gender, and driving hours/week together with internal and external factors related to driving distraction?
- What are the effects of distraction and risks associated with distracted driving?
- How can the risk of distraction for a driver be predicted?
- What impact distractions have on the performance of the driver?
- How can these distraction risks be mitigated?

Data collection

Three different data collection methods are available for a distraction study: Accident Database, Driver Perception Survey, and Route Observation.

Accident database

This data collection module follows the approach of McEvoy et al. [1] who studied the factors associated with distracted driving crashes and reported that 13.6% of all accidents are caused by distraction. Most city traffic departments collect data on auto accidents generated from police reports at different locations in the city. This data could be analyzed to determine regions of higher accidents and establish causes of accidents including distraction related factors.

Driver perception survey

A standard pre-tested survey instrument to study driver perception can be very useful to determine factors (external or/and driver characteristics) that relate to driver distraction and might be easy to administer and analyze. In that context, a self-administered

survey instrument used in the Commonwealth of Virginia needs to be redesigned to collect distraction data from a sample of drivers in other cities (which represents traffic conditions across the other metropolitan areas of considered regions). The reference Driver Distraction Survey has to consist of the following sections:

- Demographic Details, Driving Experience, and Travel Patterns.
- Source and Extent of Distraction.
- Duration of Distraction.
- Perceived Effect of Distraction.

The sample size impacts the accuracy of the survey results. Figure 2 may be used to select the sample size for each city.

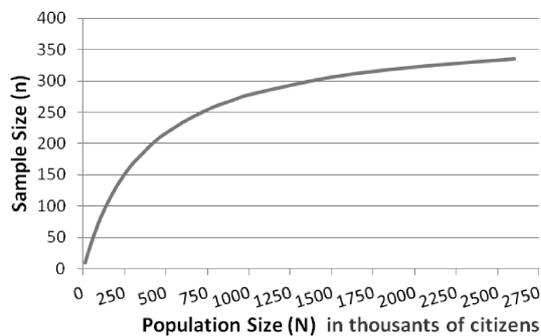


Fig. 2. Sample size determination [23].

Route observation

Data on driver distraction can also be collected via route observation. A standardized format to collect route data will help rapid determination of some distraction factors. Observers can record variety of physical as well as other distractions. Analysis of such data can be used in establishing major causes of distraction thus help in developing training and policy guidelines.

Analysis

The Analysis module consists of two steps that will assist in classification of distracting activities and development of statistical models for distracted driver data. Researchers have reported that 13.6% of all accidents were caused by distracted driving [1]. The exploratory steps will develop a system to classify data and create risk zones and identify the high risk activities using a standardized distraction risk index. The confirmatory steps will develop an appropriate multivariate statistical model for the high risk distracting activity. The MLR model which has been used in the previous Commonwealth study along with other multivariate techniques will be used to determine a set of standardized analysis techniques

for use by Polish cities. Analysis may be restricted based on the available data. For example, data collected using routing observation may be limited in recording type of distractions (e.g., cognitive distractions are not observable) thus limiting very rigorous data analysis.

Accident database analysis

The accident data can be very useful in conducting exploratory as well as confirmatory data analysis to determine the impact of driver distraction. However, quality and extent of analysis will depend upon type of data collected and available for analysis (not all collected data is always available due to legal or other reasons). An analysis of historical accident data for the past two to three years is to be conducted to identify causes of accidents in the city's different locations (for example Northside and Southside of Hampton Roads). The accidents are to be classified as being either preventable or non-preventable. The non-preventable accidents are not caused by the bus driver. For example, the bus maybe hit by another vehicle. The preventable accidents could have been avoided (for example the bus hit another vehicle) if the bus driver had exerted more caution. Some of the preventable accidents have been caused by driver distraction but the proportion is unknown.

In a study conducted of the accident databases of a transit agency in the Commonwealth of Virginia, accident reports rarely recorded any distracting factors during the accident and certain data like driver's age was not legally available for analysis. The Two-Way Contingency Table 1 shows for each location the estimated number of accidents due to driver distraction and other causes.

Table 1
Distracted driving accidents (Transit Agency 2008–2011 Accident Database).

Location of accident	Driver distraction	Other causes	Total
Northside	105	663	768
Southside	227	1442	1669
Total	332	2105	2437
% of total accidents	14%	86%	100%

This accident related data can be utilized to get some estimate of the distracted driver activities. Uniform method of data collection and extraction will help conduct a quick accident data analysis as illustrated in Figs. 3, 4, and 5. The Fig. 3 summarizes preventable and non-preventable accident data at two distinct locations in Hampton Roads. There is statistically significant difference ($p < 0.05$) in total number of accidents in Northside and Southside as well as preventable and non-preventable accidents. Since,

preventable accidents are related to driver distraction, accidents due to distraction can be assumed to be higher in the Southside as compared to Northside.

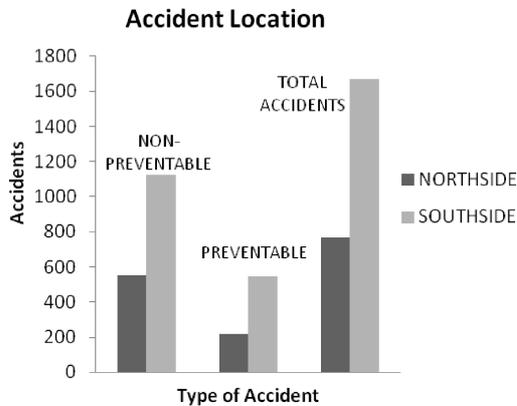


Fig. 3. Accident analysis for different locations in Hampton Roads (Transit Agency 2008–2011 Accident Database).

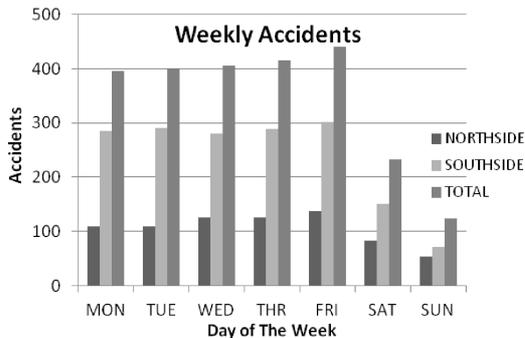


Fig. 4. Accident database analysis for days of the week (Transit Agency 2008–2011 Accident Database).

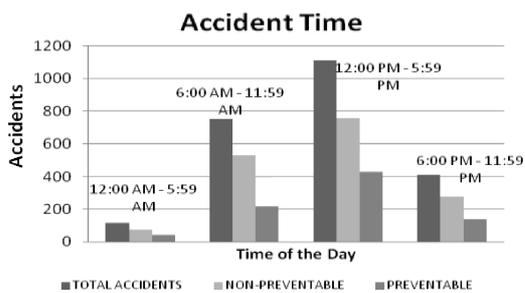


Fig. 5. Accident analysis for time of the day (Transit Agency 2008–2011 Accident Database).

The following approach of Agresti [24] is recommended for analyzing the accident database to predict the probability of accidents due to driver distraction.

Joint, marginal, and conditional probabilities (Refer to Table 2)

Let X = the explanatory (independent) categorical variable having I levels. $I = 2$ rows.

Let Y = the response (dependent) categorical variable having J levels. $J = 2$ columns.

The I, J combinations of outcomes are displayed in the table from which the predictive probabilities can be computed. Suppose a driver is selected at random and then classified on the basis of X and Y .

$p_{ij} = P(X = i, Y = j)$ is the joint probability of X and Y . Where $\sum_{i,j} p_{i,j} = 1$.

p_{i+} is the marginal probability representing the row total ($i+$).

p_{+j} is the marginal probability representing the column total ($+j$).

n_{ij} = cell count, where total sample size $n = \sum_{i,j} n_{i,j}$

$$p_{ij} = (n_{ij}/n).$$

It is clear from the Table 2 data, that the overall probability of the accidents as well as the joint probability of accidents with distractions is higher in the Southside compared to Northside. In addition to location, the number of accidents is dependent on the days of the week with Fridays having the highest number of accidents in the Southside compared to Northside (Fig. 4). Being the end of the week, it is expected that Friday would have a lot more distraction than other days due to fatigue. Therefore, the highest number of accidents due to driver distraction is on a Friday.

Table 2

Distracted driving events (Transit Agency 2008–2011 Accident Database).

Location of accident	Driver distraction (event b_1)	Other causes (event b_2)	Total
Northside (event a_1)	$n_{11} = 105$	$n_{12} = 663$	$n_{1+} = 768$
Southside (event a_2)	$n_{21} = 227$	$n_{22} = 1442$	$n_{2+} = 1669$
Total	$n_{+1} = 332$	$n_{+2} = 2105$	$n = 2437$

The time of the day for the highest number of accidents is between 12:00 to 6:00 PM (preventable and non-preventable). Assuming that the accidents caused due to driver distractions are uniformly distributed across the hours of the day, it could be said that the highest number of accidents caused by distraction is between 12:00 to 6:00 PM (Fig. 5).

The average driving experience of a driver was 12 years with Southside having more experience drivers (15 years) as compared to Northside (8 years). The drivers with the least experience (0 to 5 years) have the highest number of accidents (preventable and non-preventable) and correspondingly a higher number of accidents caused by distracted driving (Fig. 6).

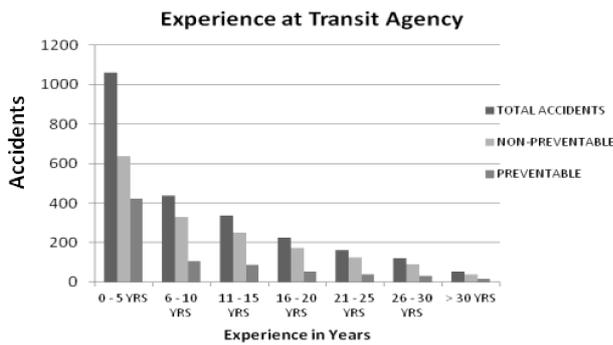


Fig. 6. Accident analysis for driver experience (Transit Agency 2008–2011 Accident Database).

Analysis of driver perception data

Data collected using a survey instrument will be more extensive for the analysis of driver distraction factors as well as distraction prediction. The survey instrument is designed for the purpose will collect data tailor-made for distraction related factors. The driver survey will collect data on sources and duration of driver distraction and perceived risks associated with a particular distracting activity along with independent factors including: location, driving hours/week, age, gender, and driving experience.

Exploratory analysis of the survey data

Based on the drivers' response to the various manual, visual and cognitive distractions activities, a classification of distraction activities will be performed. The ratings and durations for each activity will be averaged and then each activity ranked based on average rating as well as average duration [10, 25]. The distracting activities involving perceived visual, manual, and cognitive effects risk to drivers are to be ranked based on the aggregate count. The activities belonging to the top five average distraction rating, average distraction duration, and driver's perception of risk would be graded at 90%, 70%, and 50% of the highest average values. The graded scores are to be used to compute a Distraction Risk Index (DRI) for each risk zone activity similar to the "Hazard Index" developed by Smith et al. [26] for each driver and secondary task. The DRI estimates the potential risk associated with each risk zone activity. The range of rating for Risk Zone I can be set at the DRI above 70%. Similarly, the range for Risk Zone II set at the DRI between 60% and 70%, and the range for Risk Zone III set at the DRI 50% to 60%. The range for Risk Zone IV can be set for DRI below 50%. An illustration of the graded scores (%) of all the distracting activities with DRI and assigned risk zones are shown in Table 3. The DRI considers the rating, duration, and perceptions of each distracting activity

resulting in classification of high risk activities into Risk Zones I and II, III and IV.

Table 3
Illustration of graded score and distraction risk index for each distracting activity [4].

Source	Avg. rating (% of highest)	Avg. duration (% of highest)	Eyes off the road (% of highest)	Mind/ attention off the road (% of highest)	Physical interference (% of highest)	Avg. (distraction risk index)	Risk zones
Passengers using a mobile phone	100%	100%	26%	100%	45%	74%	I
Passengers	84%	84%	58%	88%	45%	72%	I
Passengers not following etiquette	95%	69%	47%	82%	45%	68%	II
Passengers talking to driver	90%	74%	21%	100%	45%	66%	II
Ticket Machine	61%	56%	89%	58%	64%	66%	II
Fatigue/Sickness	85%	50%	21%	64%	100%	64%	II
Other Road Users	79%	84%	32%	64%	55%	63%	II
Pedestrians	71%	69%	42%	64%	64%	62%	II
On-board rattles	75%	68%	37%	67%	36%	57%	III
Passengers with Infants	76%	59%	47%	82%	18%	56%	III
Climate Controls	56%	34%	58%	52%	64%	53%	III
Reading (e.g. Route Sheet)	57%	27%	100%	45%	27%	51%	III
Disabled passengers	56%	38%	47%	52%	64%	51%	III
Audible alerts	67%	46%	21%	79%	36%	50%	III
General Broadcasts	71%	57%	5%	85%	27%	49%	IV
Personal Broadcasts	67%	48%	21%	67%	36%	48%	IV
Driver's Mobile Phone	64%	28%	26%	52%	27%	39%	IV
Advertisements/Water Activity	51%	24%	16%	48%	18%	31%	IV
Others (please specify)	20%	6%	5%	27%	9%	13%	IV

Confirmatory analysis of survey data

The MLR is suitable to model the high risk distracting activities in Risk Zone I and II using levels of distraction as the dependent variable and correlating it with the factors as independent/predictor variables. For example, in the Commonwealth study, categorical dependent variable (driver distraction) had more than two levels: Not Distracted, Slightly Distracted, Distracted, and Very Distracted. The independent variables included categorical variables: gender and location, and continuous variables: age, driving experience, and driving hours per week.

The MLR could be modeled as an extension of the binary logistic regression [27, 28] by comparing each level of distraction (Slightly Distracted, Distracted, and Very Distracted) with a reference level of distraction (Not Distracted), thus producing three binary logistic regression outputs. A stepwise procedure includes all the selected factors in the model initially; non-significant factors are eliminated until a good fit was achieved with significant factors.

The general MLR model proposed by Moutinho and Hutcheson [27] is expressed as:

$$\log\left(\frac{\Pr(Y = j)}{\Pr(Y = j')}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k, \quad (1)$$

where j is the identified distraction level, and j' is the reference distraction level.

Logit model 2 comparing Slightly Distracted with Not Distracted is stated as:

$$\log\left(\frac{\Pr(Y = \text{Slightly Distracted})}{\Pr(Y = \text{Not Distracted})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k. \quad (2)$$

Logit model 3 comparing Distracted with Not Distracted is stated as:

$$\log\left(\frac{\Pr(Y = \text{Distracted})}{\Pr(Y = \text{Not Distracted})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k. \quad (3)$$

Logit model 4 comparing Very Distracted with Not Distracted is stated as:

$$\log\left(\frac{\Pr(Y = \text{Very Distracted})}{\Pr(Y = \text{Not Distracted})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k. \quad (4)$$

The logit models 2, 3, and 4 provide three estimates for the impact each independent variable has on the dependent variable, allowing the impact of dependent variable X_k to be computed for each logit model and for the whole model.

The intercept β_0 is the value of Y when all the independent variables are equal to zero. $\beta_1, \beta_2,$

β_3, \dots, β_k are the regression coefficients of $X_1, X_2, X_3, \dots, X_k$. Each of the regression coefficients describes the size of the contribution of risk factor X_i relative to a reference category. A positive regression coefficient means that the explanatory variable increases the probability of the outcome, while a negative regression coefficient means that the variable decreases the probability of that outcome [15]. Similarly, a large regression coefficient means that the risk factor strongly influences the probability of that outcome, while a near-zero regression coefficient means that that risk factor has little influence on the probability of that outcome [18].

An illustration of the multinomial dependent variable Y_i (*logit*) which measures the total contribution of the five factors (independent variables) is expressed as:

$$Y_i = \beta_0 + \beta_1 * \text{LOCAT} + \beta_2 * \text{SEX} + \beta_3 * \text{AGE} + \beta_4 * \text{EXP} + \beta_5 * \text{DRIVING/WK}, \quad (5)$$

where

LOCAT: Location of driver, a categorical variable, 1 = Northside, 2 = Southside.

SEX: Gender of driver, a categorical variable, 1 = Male, 2 = Female.

AGE: Reported age of driver in years, a continuous variable.

EXP: Number of years of experience driving a bus, a continuous variable.

DRIVING/WK: Weekly driving hours, a continuous variable.

The coefficients computed by the MLR models are relative to the reference category and can be utilized to predict the probability of the extent that a driver finds an activity distracting versus the reference category from the following binary logistic function [28]:

$$f(Y_i) = \left(\frac{1}{1 + e^{-Y_i}}\right), \quad (6)$$

where, $f(Y_i)$ is the probability of a driver getting Slightly Distracted, Distracted, or Very Distracted.

The SPSS 17.0 [29] or higher package is recommended for solving the MLR model producing three output tables. Each level is referenced with Not Distracted. The event Y is very unlikely to occur if $f(Y_i)$ is close to 0 and very likely to occur if it is close to 1. The output is split into three tables since the dependent categorical variables are compared in pairs. An illustration of the statistical test ratios and parameter estimates is presented in Table 4 for Risk Zone I distracting activity (Passengers).

Table 4
 Illustration of MLR model outputs for passengers [4].

Independent predictor variables and interactions	Coeff β (SE)	Wald Statistic	Odds ratio Exp (B)	95% CI includes 1
Slightly distracting vs. Not distracting				
Intercept	N/S	–		
LOCAT = 1	-2.20 (1.04)**	4.44	< 1	No
LOCAT = 2	0.00			
SEX =1	16.05 (6.04)**	7.07	> 1	No
SEX = 2	0.00			
AGE	N/S	–		
EXP	2.57 (1.23)**	4.34	> 1	No
DRIVING/WK	0.13 (0.07)*	3.64	> 1	Yes
AGE*DRIVING/WK	N/S	–		
SEX=1*DRIVING/WK	-0.34 (0.13)****	6.87	< 1	No
AGE*EXP	N/S	–		
Distracting vs. Not distracting				
Intercept	-224.35 (6.95)****	1042.79		
LOCAT = 1	N/S	–		
LOCAT = 2	0.00			
SEX =1	235.99 (1.53)****	23736	> 1	No
SEX = 2	0.00			
AGE	N/S	–		
EXP	0.20 (0.10)**	3.79	> 1	Yes
DRIVING/WK	4.53 (0.10)****	1947	> 1	No
AGE*DRIVING/WK	N/S	–		
SEX=1*DRIVING/WK	N/S	–		
AGE*EXP	N/S	–		
Very distracting vs. Not distracting				
DRIVING/WK	0.47 (0.21)**	5.00	> 1	No

*p < 0.10; **p < 0.05; ***p < 0.01; ****p < 0.001. N/S = Not Significant

Route observation data analysis

Type of analysis of this type data will depend upon the extent of data collected. As noted earlier, observation data has some inherent limitations. However, this type of data is very useful for quick distraction study as well as for triangulation and validation. A standardized format to collect route data will help rapid determination of some distraction factors. For example, Passengers Talking to the Driver is a high risk distracting activity in a transit agency. But this type of distraction is commonly observed in some route such as Route 111 (Fig. 7). The passengers spoke to the driver for over 70% of the time. They were standing next to the driver’s cab and talking continuously to the driver causing

distraction. Such cases could be investigated further by the city.

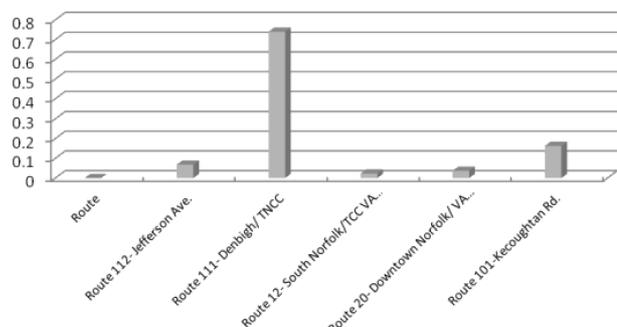


Fig. 7. Passenger talking to driver (transit agency route observations).

It should also be noted here that the observers' understanding of distraction may be very different than the understanding of bus drivers especially for cognitive and visual distractions. Observers may be allowed to speak with drivers to confirm the validity of observation or conducted without the knowledge of bus driver to avoid any observer effect in performance.

Validation

The Validation module will verify the statistical model results. Expert verification is the starting point for validation and could include safety managers in the participating transit agencies. The process of Monte Carlo simulation technique will be documented for the city use. Standardized route observation forms will also be developed for validation purposes.

The likelihood ratio test using model fitting information for Passengers is illustrated in Table 4. It shows that the difference in the Log Likelihood between the intercept only (without any independent variables) and the final model (with all the independent variables) provides the chi-square (χ^2) = 36.61 signifying a good improvement in the model fit. It follows that the independent variables contribute significantly to the outcome of the distraction level. The values of the AIC initial/final values (114.22/104.16); the BIC initial/final values (145.06/140.14) gets smaller during the stepwise process indicating a good fit for the final model. The model's Goodness of Fit as indicated by multiple statistics such as: the p-values for Pearson and Deviance (both test the same results) chi-square (χ^2) = 1.00 ($p = 1$) proving no significance. Hence, it can be inferred that the predicted values of the model are not significantly different from the observed values at all outcome levels i.e. the model fits the data well. The measures of Pseudo R^2 (0.59, 0.65, and 0.32) are reasonably similar and high values of R^2 indicating a good fit.

The Table 4 further presents outputs from the three binary logistic regression models along with the coefficients, Wald Statistic, and Odds Ratio and 95% CI values which are truncated to $<$ or $>$ 1 and includes or excludes 1 from the 95% CI.

The MLR models for the risk zone activities provide estimates of the current levels of driver distraction in the city. Are the results generated from the MLR for each risk zone activity listed in Table 3 valid for a large random population of transit bus drivers? Simulation of the models using probabilistic distributions generates driver distraction events that would occur in practice over a range of random

factors. Monte Carlo simulation using discrete distribution that incorporate random variability into the model can be applied to validate a model's output results.

Following the approach of Smith et al. [26], the independent variables could be simulated for 1,000 replications one at a time, keeping the remaining variables constant. For each set of 1,000 replications, the simulation spreadsheet model generated average probability values for Y_{ij} . The impact of independent variable coefficients on high risk activities can be validated by comparison with the simulated outputs. For each risk zone activity, the simulation spreadsheet model could generate average probability values for 1,000 drivers getting Slightly Distracted, Distracted, and Very Distracted by the impact of the factors. As an illustration, the results for Location and Driving Experience are presented as follows:

LOCATION. The MLR results for the five risk zone activities presented in Table 5 indicates that the Southside drivers have a higher chance of getting Slightly Distracted and Distracted versus Not Distracted. The simulation output (Fig. 8) validates these results for all the passenger-related activities and Ticket Machine. In the case of Ticket Machine, the probability of getting Slightly Distracted is the same for Southside and Northside drivers.

DRIVING EXP. The MLR results for the three risk zone activities (Table 6) indicates that more experienced drivers have a higher chance of getting Slightly Distracted and Distracted by Passenger-related activities and Ticket Machine. The simulation output presented in Fig. 9 validates these results for Passenger Using Mobile Phone. In the case of Passenger and Ticket Machine, the probability of getting Slightly Distracted and Distracted is the same for more experienced and less experienced drivers.

Table 5
MLR results for location [3].

Risk Zone Activity	MLR Results – Location	Simulation Results (Fig. 8)
Passengers Using Mobile Phone	Southside drivers are more likely to get Slightly Distracted followed by Distracted	Probability of Slightly Distracted and Distracted is higher for Southside drivers
Passengers	Southside drivers are more likely to get Slightly Distracted.	Probability of Slightly Distracted is higher for Southside drivers
Passengers Not Following Etiquette	Southside drivers are more likely to get Slightly Distracted followed by Distracted	Probability of Slightly Distracted and Distracted is higher for Southside drivers
Ticket Machine	Southside drivers are more likely to get Slightly Distracted.	Probability of Slightly Distracted Equal for both locations

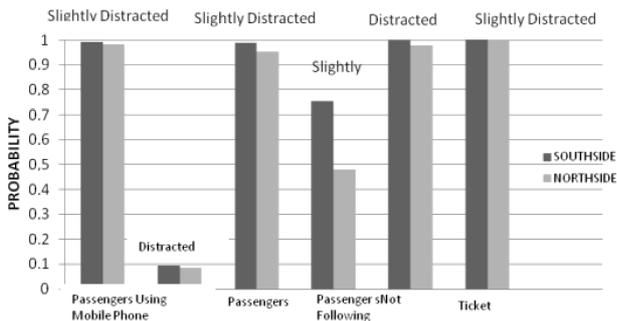


Fig. 8. Simulation results for location [3].

Table 6
MLR results for driving experience [3].

Risk Zone Activity	MLR Results – Driving Exp	Simulation Results (Fig. 9)
Passengers Using Mobile Phone	Drivers with more driving experience are more likely to get Slightly Distracted followed by Distracted	Probability of Slightly Distracted and Distracted increases with driving experience
Passengers	Drivers having more driving experience are more likely to get Distracted	Probability of Distracted equal for all driving experience
Ticket Machine	Drivers having more driving experience are more likely to get Slightly Distracted.	Probability of Slightly Distracted equal for all driving experience

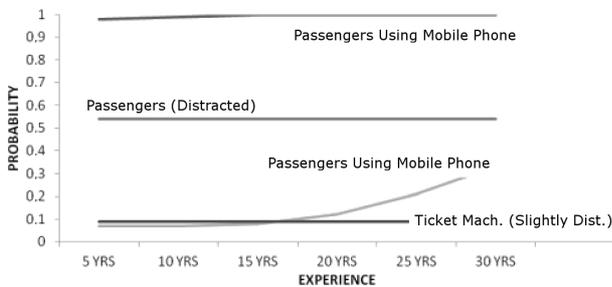


Fig. 9. Simulation results for driving experience [3].

Comparison of predicted results generated by the model with simulated outputs must be made to show similarities for the model results. For example, simulation and model results could match for certain distraction activities like Passenger Using Mobile Phone and Passengers for all independent variables namely sex, age, location, experience and driving per week. Two other distraction activities – Passenger Not Following Etiquette and Passengers match simulation and predicted results for all but one independent variable. However, distraction activity Ticket Machine does not show any convergence between predicted and simulated results for any independent variable. It is possible that survey respondents have confounded between two distract-

tion activities: Ticket Machine and Passenger. In practice most of the Ticket Machine distraction can be attributed to Passenger.

Guidelines for interpretation and usage of results

In the last module, guidelines will be created for the cities for the interpretation of results and application of those results in predicting driver distraction, developing policies, determining training needs, designing cabin or adopting technology etc.

Interpretation and usage of results

The results of the MLR models, simulation, and route observations will indicate the related activities classified under Zone I and II most distracting to the driver. It is therefore a challenge for the cities to develop effective policies for handling driver behavior, so that they are less likely to undertake distracted behavior. Training should focus on drivers who are more likely to get distracted by passengers. Educational training program on the proper use of technological devices mounted in the cab or issued to the driver, and hazards associated with utilizing these devices while driving should focus on older and female drivers who are likely to get distracted with technological devices. The design of control panel, and other devices must be user-friendly, and not require long glances away from the forward roadway.

How could the cities use the MLR models developed in this study? They could be applied to predict distraction for varying driver characteristics and driving patterns. It is observed that drivers with differing factors are affected differently by distracting activities which may be possibly corrected through proper training. For example, the city’s transit agency could develop driver MLR models for each risk zone activity from its existing driver database. These models could be used for predicting the probability of a new driver getting distracted by the risk zone activities (Table 7). If the probability value is high, the new driver could be scheduled for related training.

As an illustration, consider the MLR dependent variable Y_i (*logit*) which measures the total contribution of the five factors (independent variables) is expressed as:

$$Y_i = \beta_0 + \beta_1 * LOCAT + \beta_2 * SEX + \beta_3 * AGE + \beta_4 * EXP + \beta_5 * DRIVING/WK, i = 1, 2, 3, 4 \quad (7)$$

$$f(Y_i) = \left(\frac{1}{1 + e^{-Y_i}} \right),$$

where $f(Y_i)$ is the probability of a driver getting Slightly Distracted, Distracted, or Very Distracted.

Table 7
 Predicting the driver distraction levels [4].

Driver	LOC	Age YRS	SEX	EXP YRS	HRS/WK	SD-MOB PHONE	DIS PASSEN	V DIS PASSEN	V DIS FATIG	SD ETIQUÉ	DIS ETIQUÉ	V DIS ETIQUÉ	SD TIC MACH
1	1	52	0	21	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
2	1	38	0	0.7	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
3	1	56	0	6	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
4	1	50	1	0.5	40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
5	1	32	0	6	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
6	1	45	0	5	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
7	1	51	1	3	40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
8	1	46	0	5	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
9	1	67	0	18	40	1.00	0.00	1.00	0.50	0.01	1.00	0.50	1.00
10	0	47	0	0.5	40	1.00	0.00	1.00	0.50	0.50	1.00	0.50	1.00
11	0	32	0	6	40	1.00	0.00	1.00	0.50	0.50	1.00	0.50	1.00
12	0	43	0	15	40	1.00	0.00	1.00	0.50	0.50	1.00	0.50	1.00
13	0	51	0	15	40	1.00	0.00	1.00	0.50	0.50	1.00	0.50	1.00
14	0	45	1	4	10	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00
15	0	46	1	0.5	40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
16	0	29	1	1	40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
17	0	30	1	1	40	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
18	0	23	1	1	25	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

The Y_i (logit) and the probability $f(Y_i)$ are computed for each driver listed in Table 7. It can be noted that Driver #4 and #14 has high probability of distraction for all the distracting activities and requires a rigorous training program. While Driver #8 and #10 have a fewer distracting activities.

Conclusions

Studying the causes of driver distraction and the factors that impact its risk level is needed for an overall understanding of accident occurrences. To help safety policy makers fully comprehend and implement the various components of a driver distraction study in any city or state, this paper proposes a research framework for modeling, analysis, and predicting driver distraction. Understanding the driver and external factors that could cause distraction can help to develop effective policies to mitigate risk of accidents.

In this context, the paper has attempted to combine independent procedures for studying driver distraction into a comprehensive framework. It is one of only a few studies to consolidate methodologies for data collection, analysis, validation, and interpretation of results into a framework. It would have wide applications in other metropolises around the World with fresh inputs from an expanded study covering sample of drivers from different European as well as Asian cities (which represents traffic conditions across metropolitan regions located in different climatical zones and following diverse habits and traffic regulations).

The results from this study could be applied to reduce driver distraction and improve overall transit performance. Only five factors: location, driving hours/week, and driver gender, age, and driving experience were included as the independent variables. Other variables, however such as environmental, vehicle type, roadway designs etc could also have an impact on driver distraction.

The MLR model could be used to predict the probability of an existing or new driver getting distracted by a risk zone activity. Relevant training programs can then be developed to mitigate risk of distraction. Training needs to be focused towards male and younger drivers who are more likely to get distracted by passengers-related activities. Educational training program should be designed on the proper use of technological devices mounted in the cab or issued to the driver, and hazards associated with utilizing these devices while driving. It seems that the particular technique, or sub-set of techniques, employed, however, will depend on the particular aspect of the Human Machine Interface [30] to be assessed, and in particular on the form of distraction (e.g., visual, physical etc) that is imposed on the driver by that aspect of the interface. The findings of this search suggest that using a range of distraction measurement techniques, rather than a single technique, would be appropriate in evaluating the Human Machine Interface design concepts and prototypes in vehicles.

References

- [1] McEvoy S.P., Stevenson M.R., Woodward M., *The Prevalence of, and Factors Associated with, Seri-*

- ous Crashes Involving a Distracted Activity, Accident Analysis and Prevention, 39, 475–482, 2007.
- [2] D'Souza K.A., Maheshwari S.K., "Improving performance of public transit buses by minimizing driver distraction", *Proceedings of the Urban Transport 2012 Conference*, A Coruña, Spain, May 13–16, 2012. [Eds.] J.W.S. Longhurst and C.A. Brebbia, Wessex Institute of Technology Press, Southhampton, U.K., 281–293, 2012a.
- [3] D'Souza K.A., Maheshwari S.K., "An analysis of transit bus driver distraction using multinomial logistic regression models", *On-line Proceedings of the Journal of the Transportation Research Forum*, Tampa, FL, March 15–17, www.trforum.org/forum/2012, 2012b.
- [4] D'Souza K.A., Maheshwari S.K., *Multivariate Statistical Analysis of Public Transit Bus Driver Distraction*, Journal of Public Transportation: Special Edition: Rural and Intercity Bus, 15, 3, 1–23, 2012c.
- [5] Morfoulaki M., Tyrinopoulos Y., Aifadopoulou G., *Estimation of Satisfied Customers in Public Transport Systems: A New Methodological Approach*, Journal of the Transportation Research Forum, 46, (1), 63–72, Spring 2007.
- [6] Salmon P.M., Young K.L., Regan M.A., *Distraction 'On the Buses': A Novel Framework of Ergonomics Methods for Identifying Sources and Effects of Bus Driver Distraction*, Applied Ergonomics, 42, 602–610, 2011.
- [7] U.S. Department of Transportation, "An Examination of Driver Distraction as Recorded in NHSTA Databases", *DOT HS 811 216*, September 2009.
- [8] ECORYS, Study on Strategic Evaluation on Transport Investment Priorities under Structural and Cohesion funds for the Programming Period 2007–2013, No 2005.CE.16.0.AT.014. Country Report Poland Final, Client: European Commission, DG-REGIO. ECORYS Nederland BV, Rotterdam, The Netherlands, pp. 14, September 2006.
- [9] AAA Foundation for Traffic Safety, "Distraction in Everyday Driving", *University of North Carolina at Chapel Hill, Highway Safety Research Center*, June, 2003. www.aaafoundation.org.
- [10] Salmon P.M., Young K.L., Regan M.A., "Bus Driver Distraction Stage 1: Analysis of Risk for State Transit Authority New South Wales Bus Drivers", *Final Report*, Monash University Accident Research Center, Victoria, Australia, June 2006.
- [11] Young K.L., Regan M.A., Lee J.D., *Factors Moderating the Impact of Distraction on Driving Performance and Safety*, Driver Distraction: Theory, Effects, and Mitigation, Edited by M.A. Regan, J.D. Lee, and K.L. Young, CRC Press, Taylor and Francis Group, 335–351, 2009.
- [12] Yan X., Radwan E., Abdel-Aty M., *Characteristics of Rear-End Accidents at Signalized Intersections Using Multiple Logistic Regression Model*, Accident Analysis and Prevention, 37, 983–995, 2005.
- [13] Oxley J., Charlton J., Fildes B., Koppel S., Scully J., Congiu M., Moore K., "Crash Risk of Older Female Drivers", *Report No. 245*, Monash University Accident Research Center, Victoria, Australia, November 2005.
- [14] Blower D., Green P.E., Matteson A., "Bus Operator Types and Driver Factors in Fatal Bus Crashes: Results from the Buses Involved in Fatal Accidents Survey", *U.S. DOT Federal Motor Carrier Safety Administration*, University of Michigan Transportation Research Institute, Ann Arbor Michigan, June 2008.
- [15] Washington S.P., Karlaftis M.G., Mannering F.L., *Logistic Regression, Discrete Outcome Models and Ordered Probability Models*, Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall/CRC, 303–359, 2011.
- [16] Yan X., Radwan E., Mannila K.K., *Analysis of Truck-Involved Rear-End Crashes Using Multinomial Logistic Regression*, Advances in Transportation Studies an International Journal, Section A, 17, 39–52, 2009.
- [17] Gkritza K., Niemeier D., Mannering F., *Airport Security Screening and Changing Passenger Satisfaction: An Exploratory Assessment*, Journal of Air Transport Management, 12, 213–219, 2006.
- [18] Petrucci C.J., *A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression*, Journal of Social Service Research, 35 (2), 193–205, 2009.
- [19] Carlson W.L., *Application of Simulation to Assessment of Component Changes on rear End Highway Accidents*, Transportation Science, 6 (1), 1–14, February 1972.
- [20] Salvucci D.D., Chavez A.K., Lee F.J., "Modeling Effects of Age in Complex Tasks: A Case Study in Driving", *26th Annual Conference of the Cognitive Science Society*, 2004.
- [21] Wong Jinn-Tsai, Shah-Hsuan Huang, *Modeling Driver Mental Workload for Accident Causation and Prevention*, Journal of the Eastern Asia Society for Transportation Studies, 8, 1890–1905, 2009.
- [22] Trick L.M., Enns J.T., Mills J., Vavrik J., *Paying Attention Behind the wheel: a framework for studying the role of attention in Driving*, Theoretical Issues in Ergonomics Science, 5 (5), 385–424, 2004.

- [23] Orcher L.T., Conducting a Survey: Techniques for a Term Project, *Pyrczak Publishing, P. O. Box 250430, Glendale, CA 91225*, pp. 46–47, 2007.
- [24] Agresti A., An Introduction to Categorical Data Analysis. *John Wiley and Sons, Inc., New York, NY, 2nd Ed.*, 2007.
- [25] Cross G.W., Hanna H., Garrison T., McKee C., “Distracted Driving in Mississippi: A Statewide Survey and Summary of Related Research and Policies”, *Final Report by Family and Children Research Unit, MSU*, The Center for Mississippi Health Policy, September 2010.
- [26] Smith D.I., Chang J., Cohen D., Foley J., Glassco R., “A Simulation Approach for Evaluating the Relative Safety Impact of Driver Distraction During Secondary Tasks”, *12th World Congress on ITS*, San Francisco, 6–10 November, 2005, Paper 4340.
- [27] Moutinho L., Hutcheson G., “Multinomial Logistic Regression”, *The SAGE Dictionary of Quantitative Management Research*, SAGE Publications Ltd., pp. 208–212, 2011.
- [28] Field A., *Logistic Regression*, Discovering Statistics Using SPSS, SAGE Publications Ltd., 3rd Edition, 264–315, 2009.
- [29] SPSS 17.0. SPSS Inc., Chicago, IL 60606-6412, 2008.
- [30] Young K.L., Regan M.A., Hammer M., *Driver Distraction: A Review of the Literature*, Monash University Accident Research Centre, Victoria, Australia, Report No. 206, Holden Ltd. November 2003.