A METHODOLOGICAL APPROACH FOR STUDYING PUBLIC BUS TRANSIT DRIVER DISTRACTION

K.A. D'SOUZA & S.K. MAHESHWARI
Hampton University, USA.

ABSTRACT

The increase in bus transit ridership along with the proliferation of personal electronic control and communication gadgets is causing more distractions for the drivers. For transit vehicles, some distractions are caused by factors beyond the driver's control such as operating additional equipment, attending to passengers, and communicating with the operations center. Several driver distraction studies have been conducted for personal vehicles and commercial vehicles. But bus transit driver distraction has received limited attention in the literature even though bus transit accidents may cause more injuries due to larger number of passengers. Hence, their distraction is not clearly understood; furthermore, no established methodology is available to conduct a detailed study at a transit agency because of inadequate research in the field.

The objective of this paper is to present a detailed modular research framework for studying bus transit driver distractions. The framework provides a transit agency with a set of standardized methodologies for studying distraction over a wide range of cost and time intervals. An agency may choose one or more modules to suit their study requirements. The modules for data collection, analysis, validation, and interpretation and usage of results are designed on the basis of in-depth studies and tests at transit agencies in the Commonwealth of Virginia. The paper provides a detailed process and a set of guidelines to study bus transit driver distraction which will make it easier for any transit agency to conduct such a study. The results of the bus transit driver distraction studies could be used for training bus drivers to mitigate distraction and assist state and city governments to formulate effective regulations to control distracted driving.

Keywords: Bus transit driver distraction, Distraction Risk Index, modular design for studying bus driver distraction, model validation, modeling and predicting driver distraction, Monte Carlo simulation, multinomial logistic regression, route observations.

1 INTRODUCTION

Over the past several years, there have been numerous transit accidents, which were caused due to operator distraction while using portable electronic devices. Transit accidents get abundant attention in the media and public due to potentially large number of passengers involved in each incident. Reports on the July 24, 2013 Spanish Santiago rail-crash, which killed 79 passengers and crew members and injured 170 indicated that the operator was allegedly talking over the phone prior to the accident [1]. Investigation of ex post facto data from these transit accidents point out to driver distraction as the cause in many accidents. Regardless of these efforts, data on the types and causes of transit operator distractions have limited literature coverage.

To mitigate transit operator distractions, regulators have enforced rules and guidelines that prohibit the use of personal electronic devices while driving. These rules and guidelines often overlook other causes of operator distractions such as interaction with passengers, other road users, pedestrians, etc. in spite of the evidence that many of these factors similarly contribute to operator distractions [2]. Guidelines on how to conduct a transit operator distraction and the type of data to be collected are not documented well. Without such guidelines or well tested methodology, transit agencies face difficulties and delays while conducting distraction studies to improve the safety of their operations. Transit agencies would greatly benefit if a methodology is available to analyze the distraction data and implement the results.

This research project [2] attempted to develop a detailed methodology for conducting a public bus transit driver distraction study. Although the main focus of the research was on the bus transit mode, the methodology could be applied to the other transit modes after making essential modifications in the types and causes of distractions. The objective of this paper is to present the modular research framework that was used in the project to study bus transit driver distractions and related field tests at two transit agencies in the Commonwealth of Virginia, USA [2].

This paper is the first of its kind to present transit agencies with a tested methodology to conduct a driver distraction study thereby saving time and cost. The framework consists of a detailed set of standardized modules for studying bus transit driver distraction over a wide range of cost and time intervals. An agency may choose one or more modules to suit their study requirements and have the flexibility of modifying, adding, or deleting modules as and when needed. The preliminary research framework presented earlier by D'Souza and Maheshwari [3] is expanded to provide detailed methodologies that offer an agency the option of choosing from a set of modules for conducting a driver distraction study.

2 LITERATURE REVIEW

Driver distraction is the cause of a large proportion of traffic accidents, which have been studied by researchers [4] and government agencies [5,6] around the world. McEvoy *et al*. [7] identified the types of distracting activities and factors associated with serious crashes on Australian highways. Commercial truck and bus driver distractions were studied by Hickman *et al*. [8] by analyzing 1 year of naturalistic data. The National Safety Council provides a detailed analysis of driver distraction due to cell phone usage including hands-free devices [9]. Public bus transit driver distraction study was initiated by Salmon *et al*. [10,11] and further expanded by D'Souza and Maheshwari [12] through the application of multivariate statistical models.

Due to limited number of distraction studies that have been conducted for bus transit drivers, there is a lack of an established framework for conducting such a study at a transit agency. Each bus driver distraction study is planned and conducted independently using different methodologies that take additional time and resources. The recent bus transit driver distraction studies conducted in the Commonwealth of Virginia has demonstrated that the methodology could be standardized for use by other transit agencies [2].

The framework proposed in this paper builds upon the works of Salmon *et al*. [10,11] that identified the sources and duration of distraction for bus drivers at the State Transit Authority, New South Wales (STA, NSW), Australia. In these studies, a taxonomy of the sources of distraction was developed for bus transit driver and a descriptive statistical analysis was conducted. The limited sample size of 18 drivers comprising 16 males and 2 females provided insufficient data for an inferential statistical analysis. D'Souza and Maheshwari [2] expanded the exploratory work of Salmon *et al*. [11] using multivariate statistical models and simulation to confirm the impact of driver attributes, driving pattern, service location, and type and age of buses on the distracting activities.

Driver distraction has been identified as a leading cause of traffic accidents prompting regulators to introduce policies to control distraction while driving. Out of the myriad causes of distraction, the use of cellular phones has generated a growing concern of distraction prompting several states and the District of Columbia to ban its use during driving. Most transit agencies allow bus drivers to use a two-way radio or hand-free wireless phone to communicate with the operations centers. Studies show that cellular phones represent a smaller part of the distraction problem [13] in comparison to the growing use of advanced in-vehicle information systems [14], which are installed in current transit buses.

The multivariate statistical models used in the analysis module of the research framework have been widely used in transportation to study the relationship between a categorical response variable consisting of more than two levels and a set of continuous and categorical predictor variables. Washington *et al*. [15] developed a multinomial logit (MNL) model consisting of 18 independent variables covering driver factors, traffic flow, distance, number of signals, etc. in a study of factors that influence drivers' selection of route on their morning commute to work. The nominal outcome variable represented the modes of travel (an arterial, a two-lane road, or a freeway) and the covariates consisted of categorical and continuous variables like gender, number of signals, age of vehicle, commute distance, etc. [15]. A multinomial logistic regression (MLR) model was developed by Morfoulaki *et al*. [16] 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 MNL models to investigate the socio-economic and demographic factors that significantly affect passenger satisfaction with airport security screening process. The odds ratios for the tasks/variables, along with 95% confidence intervals (CIs) were utilized to identify the high-risk tasks/variables and the strength of association between the categorical-dependent variable and independent variables [18]. Following the approach of Washington *et al*. [15] and Morfoulaki *et al*. [16], D'Souza and Maheshwari [12] proposed a MLR model to analyze public bus transit driver distraction that included five predictor variables linked to a categorical response variable having four levels of distraction.

Monte Carlo simulation has been widely used to validate statistical models' empirical results. For example, the impact of age and cognitive functions on driving performance has been studied extensively to predict cognitive distraction using a computational cognitive model and validating the results through simulation [19]. A simulation approach was developed by Smith *et al*. [20] to evaluate the impacts on safety that occur when drivers become distracted by secondary tasks, and to compute a *Hazard Index* that measured the potential for a collision to occur due to a driver's being distracted.

3 PROPOSED RESEARCH FRAMEWORK AND ITS APPLICATION

This section presents the research framework comprising a detailed methodology for conducting a public bus transit driver distraction study. The research framework was constructed by combining modules for data collection, analysis, validation, and guidelines for results interpretation and usage. A schematic outline of the entire framework is shown in Fig. 1. It is a modular framework for distraction data collection, analysis, and implementation. These modules are created so that transit agencies have multiple options while conducting a driver distraction study. The tools within each module are necessary for studying the sources and durations of driver distractions, and the risks associated while engaging in potential distracting activities. The visual, manual, and cognitive factors that are responsible for distraction were selected independently from earlier research [10,11]. The rationale and application for options within each module is discussed with examples from bus transit driver distraction studies conducted in Common Wealth of Virginia [2].

3.1 Data collection

At the data collection stage, three different data sources were identified: accident database, driver perception survey, and route observation. Data collection methods were developed and tested for
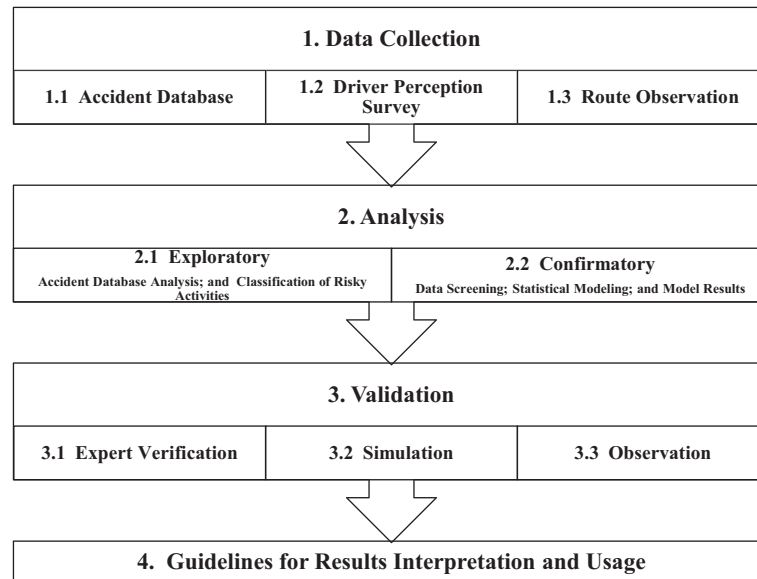
1. Data Collection		
1.1 Accident Database	**1.2 Driver Perception Survey**	**1.3 Route Observation**

2. Analysis	
2.1 Exploratory Accident Database Analysis; and Classification of Risky Activities	**2.2 Confirmatory** Data Screening; Statistical Modeling; and Model Results

3. Validation		
3.1 Expert Verification	**3.2 Simulation**	**3.3 Observation**

4. Guidelines for Results Interpretation and Usage

Figure 1: Research framework outline [3].

each of these sources. In view of space limitations, all descriptive statistical results are summarized in Table 1.

3.1.1 Accident database

The past 2–3 years accident databases were utilized to get a quick analysis of accidents and an estimate of the distracted driving activities. The accidents are classified as being either preventable or non-preventable [3]. Some of the preventable accidents are caused by driver distraction but the proportion is unknown as drivers generally do not report distraction as a cause of their accident. In this study, the reported estimate of 17% [21] of the total accidents is applied to compute the number of accidents due to driver distraction (Table 1).

3.1.2 Driver perception survey

The Transit Bus Driver Distraction Survey [2] was used to collect driver attributes, driving patterns, and type of bus driven along with the drivers' perception of distraction. The driver attributes are summarized in Table 1 and the drivers' perception of distraction is further discussed in Section 3.2.1.

3.1.3 Route observations

Data on driver distraction can also be collected via route observations. Observers can ride the bus on selected routes having relatively high accident rates and record any type of distraction along with possible causes on the Route Observation form [2].

3.2 Analysis

The data analysis has been categorized into exploratory data analysis (EDA), and confirmatory data analysis (CDA). In the application of EDA, the data are transformed largely into graphical or tabular

Table 1: Summarized data output from data collection module.

Gender   Males = 74%
                                 Females = 26%

Age (years):   Mean = 47, Std. dev = 9.69
Driving exp. (years): Mean = 8, Std. dev = 8.35

Marital Status:

Married	65%
Separated	2%
Divorced	13%
Never married	13%
Not revealed	7%

Education Level:

<High school	2%
High school	44%
Some college	20%
2-year college	20%
4-year college or higher	14%

Driving hours/week: Mean = 37.14, Std dev. = 15.2.

Two-way contingency table for distraction accidents:

Location of accident	Driver distraction (B_1)	Other causes	Total
Northside (A_1)	131	637	768
Southside (A_2)	284	1385	1669
Total	415	2022	2437

Driving service location:

Commuter	64%
Local	19%
Metro feeder	8%
Others	6%
No response	3%

Driving schedule:

Day	62%
Night	9%
Peak	22%
Non-peak	7%
Others	0%

Type of bus commonly driven

MCI	45%
Gillig 30′ Low Floor	15%
Gillig 40′ Low Floor	12%
Gillig 40′ High Floor	8%
Orion V 40′	9%
No response	11%

Avg. age of bus = 9.5 years
Std. dev. = 3.7 years

The bus age distribution is as follows:

Less than 5 years	18%
6–10 years	39%
11– 15 years	38%
More than 15 years	5%

models for identifying the characteristics and patterns in the data. The EDA would also help in establishing various hypotheses, which can be formalized and tested by the CDA models. A variety of tools can be used for corroborating the research hypotheses, for example, to assess the likelihood of degree of driver distraction based on his/her demographical characteristics, driving pattern, location, and type of bus.

3.2.1 Exploratory data analysis (EDA)

3.2.1.1 Accident database analysis

The accident data can be very useful in conducting EDA to determine the impact of driver distraction. However, the quality and extent of analysis will depend upon the type of data collected and available for analysis (not all collected data are always available due to privacy or other reasons). An analysis of historical accident data for the past 2–3 years is recommended to get a fair estimate of the distracted driver activities. The city could be divided into different locations. For example, in the case of the regional bus transit agency study, the city was divided into Northside and Southside locations based on population density characteristics and layout of the streets, accident frequency, etc. In such cases, the accident data could be categorized for each location (Table 1). There was a statistically significant difference ($p < 0.05$) in total number of accidents in Northside and Southside as well as preventable and non-preventable accidents with the Southside having a higher accident rate. Since preventable accidents are related to driver distraction, accidents due to distraction can be assumed to be higher in the Southside as compared with Northside. To prove this, the following approach of Agresti [22] is applied to the two-way contingency data in Table 1 to predict the probability of accidents due to driver distraction. The detail analyss is published in [3] and only the final results is reproduced in the following paragraphs.

The probability that a driver from the Northside (Event A_1) will have an accident due to distraction (Event B_1), or the probability of a driver from the Southside (Event A_2) will have an accident due to distraction (Event B_1) is analyzed as follows:

Applying the general rule of multiplication for the Northside: $P(A_1 \text{ and } B_1) = P(A_1)P(B_1 \mid A_1) = (768/2437)(131/768) = 0.055$.

Applying the general rule of multiplication for the Southside: $P(A_2 \text{ and } B_1) = P(A_2)P(B_1 \mid A_2) = (1669/2437)(284/1669) = 0.114$.

It is clear from the above analysis that the overall probability of the accidents as well as the joint probability of accidents with distractions is higher (two times) in the Southside compared with Northside.

A driver's experience in driving transit buses influences her/his driving performance. Results obtained from the regional transit agency study [2] reveal that less experienced drivers have higher accidents (preventable and non-preventable) than the more experienced drivers (Fig. 2). Since less experienced drivers are generally young, it is clear that young, inexperienced bus transit drivers are at an increased risk of distraction and a major hazard for other road users.
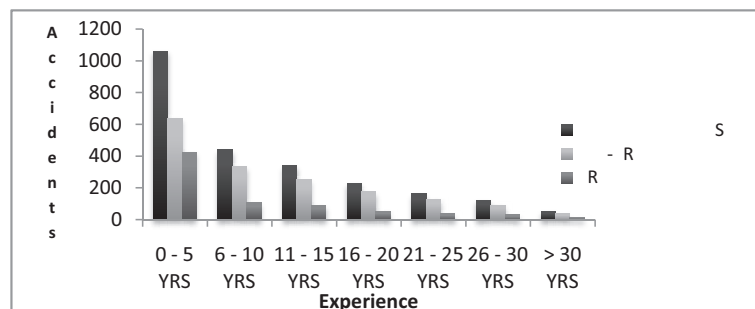

Figure 2: Impact of bus transit driving experience on accidents [2].

3.2.1.2 Classification of risky activities

The 20–23 distraction activities collected from the surveys were classified into *risk zones* according to relative grades that were estimated for rating, duration, and perception of distraction [3]. A Distraction Risk Index (DRI) computed for each distracting activity was used to classify the activities into risk zones according to the ranges shown in Table 2 [2]. All distracting activities were classified into four risk zones: Risk Zone I *Very High Risk*, Risk Zone II *High Risk,* Risk Zone III *Moderate Risk*, and Risk Zone IV *Low Risk*. The classification reveals the (few) high-risk activities contributed largely to distraction. The agency could prioritize plans for mitigating these high-risk distracting activities in order to improve safety and bus driver performance.

3.2.2 Confirmatory data analysis (CDA)

The confirmatory methods propose appropriate multivariate statistical models for confirming results from the exploratory methods as well as providing an agency with conclusive results. Analysis techniques depend upon the type of data collection method used. The quality and detail of the data extracted from the accident database will depend upon each agency's guidelines for recording accident data. Direct data collection via methods like driver survey could be more detailed as well as would reflect existing conditions and perception of drivers.

Valid conclusions can only be drawn from data that correctly represent the problem being studied and with an adequate number of cases (samples). Hence, a pre-analysis data screening [23] is recommended prior to the actual statistical analysis to detect the accuracy of the data, missing data, extreme values or outliers, and fulfillment of necessary assumptions. Hosmer *et al*. [24] have reported that including several predictor variables with less number of cases could result in *numeric instability* and recommends the *Rule of 10* for deciding on the number of cases, i.e. the sample must contain (at least) 10 cases for each predictor variable. The resulting model with all 10 covariates had 17 variables (including the pairs (1, 0) for the dichotomous variables).

3.2.2.1 Statistical modeling

The generalized linear model (GLM) technique of MLR was applied to determine the variables that have an explanatory impact on the level of risk zone distracting activities. The estimated coefficients

Table 2: Classification of distracting activities into risk zones [2].

DRI range [10,24]	Risk zone	Type of risk	Distracting activities
More than 66%	I	Very high	Pedestrians, passengers (moving around, standing next to driver's cabin, talking next to driver's cabin), other road users, unruly kids
More than 60% and up to 66%	II	High	Passengers using mobile phone, mobile data terminals, passengers not following etiquette (eating, drinking, smoking, noisy), ticket machine/farebox
More than 54% and up to 60%	III	Moderate	On-board rattles, communication with dispatch, looking at advertisements, passengers trying to talk to driver, fatigue/ sickness, climate controls, driver's mobile phone, disabled passengers, announcing bus stops, reading (e.g. route sheet)
Less than or equal to 54%	IV	Low	Dispatch broadcasts, food and other smells, passengers with infants, general broadcasts/ other, audible alerts

of the predictors allow determination of the factors responsible for increasing or decreasing the risk of distractions. The outcome is the distraction activity having multiple risk levels was experienced differently by the drivers with specific attributes, driving pattern, service location, and type of bus driven. The distracting activities had four categories of risk levels: *Not Distracted, Slightly Distracted, Distracted*, and *Very Distracted*. The higher is the risk level, the greater is the chance of an accident.

The MLR model was developed as an extension of the logistic regression [25–27] that generates a relationship between dichotomous outcomes and one or more continuous or categorical predictors. The polytomous outcome of the MLR model is converted into dichotomous outcomes using one of the outcomes (not distracted) as a reference level. Hence, the four-outcome level MLR model is converted into three logistic regression models.

The principle mathematical theory behind logistic regression is the logit, which is the natural logarithm (ln) of the odds of event Y (for example, a driver gets distracted by passenger) that is related to one or more predictor variable. The odds are defined as the ratio of the probability $\pi(x)$ that event Y occurs (for example, a driver gets distracted by Passengers) divided by the probability $(1 - \pi(x))$ that event Y will not occur (driver is not distracted by passengers).

Or

$$\text{Odds} = \left[\frac{\pi(x)}{(1 - \pi(x))} \right] \tag{1}$$

Therefore,

$$\text{Logit } (Y) = \text{natural log [Odds]} = ln\left[\frac{\pi(x)}{(1 - \pi(x))} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{2}$$

Taking antilog of eqn (2) on both sides, we derive eqn (3) which predicts the probability of an outcome of an event (for example, distraction level of an activity)

$$\pi(x) = \left[\frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k}} \right] \tag{3}$$

eqn (3) can be simplified as

$$\pi(x) = \left[\frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}} \right] \tag{4}$$

where $\pi(x)$ is the probability of a driver getting *Slightly Distracted, Distracted*, or *Very Distracted* with reference to *Not Distracted*. e = 2.71828 is the base of the natural logarithms. $\pi(x)$ increases continuously as x increases, taking the shape of an S-shaped graph [2].

Applying eqn (1), the general MLR model can be expressed in logistic regression form. The *random component* is linked to the *systematic component*, using a nonlinear link function called the *logit* [25]:

$$ln\left[\frac{\pi(Y = j)}{\pi(Y = j')} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{5}$$

where j is the identified distraction level (*Slightly Distracting, Distracted,* and *Very Distracted*) and j' is the reference distraction level (*Not Distracted*).

Logit model (eqn (6)) comparing *Slightly Distracted* with *Not Distracted* could be stated as

$$ln\left[\frac{\pi(Y = \text{Slightly distracted})}{\pi(Y = \text{Not distracted})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{6}$$

Logit model (eqn (7)) comparing *Distracted* with *Not Distracted* is stated as

$$ln\left[\frac{\pi(Y = \text{Distracted})}{\pi(Y = \text{Not distracted})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{7}$$

Logit model (eqn (8)) comparing *Very Distracted* with *Not Distracted* is stated as

$$ln\left[\frac{\pi(Y = \text{Very distracted})}{\pi(Y = \text{Not distracted})}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{8}$$

The logit models (eqns (6)–(8)) provide three estimates for the impact each predictor variable has on the response variable, allowing the impact of predictor variable x_k to be computed for each logit model and for the whole model [15]. The multinomial *linear predictor* that measures the total contribution of the 10 factors (predictor variables) for the transit agency is expressed as

$$
\begin{aligned}
Y_{ij} = \beta_0 &+ \beta_1 \text{SEX} + \beta_2 \text{AGE} + \beta_3 \text{EXP} + \beta_4 \text{MARITAL} + \beta_5 \text{EDU} + \beta_6 \text{DRIVING/WK} \\
&+ \beta_7 \text{LOCAT} + \beta_8 \text{DAY} + \beta_9 \text{PEAK} + \beta_{10} \text{EQUIP}
\end{aligned}
\tag{9}
$$

where SEX: gender of driver, 1 = male, 0 = female
   AGE: reported age of driver in years
   EXP: number of years of experience at the transit agency
   MARITAL: marital status, 1 = married, 0 = other (separated, divorced, never married, etc.)
   EDU: educational level, 1 = HS or equivalent, 0 = other (some college, 2- or 4-year degree, etc.)
   DRIVING/WK: weekly driving hours
   LOC: location of transit agency service area, 1 = commuter, 0 = other (local, metro, etc.).
   DAY: driving schedule, 1 = day, 0 = night
   PEAK: driving time, 1 = peak, 0 = non-peak
   EQUIP: type of equipment driven, 1 = MCI, 0 = other (Gillig, Orion, etc.)

3.2.2.2 Model results
Unruly passengers and children, and passengers not following etiquette appear to be common distracting activities in transit agencies. Passengers using mobile phones, unruly kids, and passengers were classified under Zone I (very high risk). This coincides with earlier studies [10,11] that identified passenger-related activities as the most common form of distraction. Distracting activities such as a driver carrying on a conversation with a passenger or listening to a passenger's mobile cell phone conversation leads to multitasking while driving. A threshold is reached, particularly as additional tasks are added, which increases mental inattention towards the primary driving tasks producing a crash risk [9].

3.2.2.3 Fitting the MLR model
Eqns (6)–(8) were fitted to the survey data by SPSS 17.0 [28]. As an illustration, the summarized results from Appendix 10 [2] for *Pedestrian* (the highest risk distracting activity) are presented in Table 3. The model is evaluated for goodness of fit from the Step Summary (Table 4) and Model Fitting Information (Table 5), where the 2 log-likelihood computes the unexplained variability in the data. Hence, the differences or change in log-likelihood is the new variance explained by the model [27]. The chi-square test computes the difference in variances from the baseline model to the final

Table 3: MLR model outputs for pedestrians [2].

Model chi-square (χ^2) = 80.420 (33)****	$R^2 = 0.79$ (Cox and Snell); 0.87 (Nagelkerke); 0.66 (McFadden)	AIC initial/final values: N/A		
Pearson stat (NS)		BIC initial/final values: N/A		
Deviance stat (NS)				
Independent variables and interactions	Coeff β (SE)	Wald statistic	Odds ratio Exp. (B)	95% CI
Slightly distracted vs. not distracted				
Exp	−0.487*** (0.183)	7.101	0.615	0.43–0.88
Distracted vs. not distracted				
Intercept	−13.47 (8.332)	2.614	N/A	N/A
AGE	0.264* (0.147)	3.211	1.302	0.97–1.74
PEAK = 0	−5.937* (3.458)	2.942	0.003	0–2.33
Drive (h/wk)	0.178** (0.082)	4.654	1.194	1.02–1.40
Very distracted vs. not distracted				
N/S	N/S	N/S	N/S	N/S

*$p < 0.10$; **$p < 0.05$; ***$p < 0.01$, N/S, not significant.

Table 4: Step summary.

Model	Action	Effect(s)	Model fitting criteria	Effect selection tests		
			−2 Log likelihood	Chi-square[a]	df	Sig.
0	Entered	Intercept, DriveWk, Age, Peak, Marital, Sex, Loc, Edu, Day, Equip, E	70.165			
1	Entered	Sex * Edu	42.354	27.812	3	0.000

Stepwise method: forward entry.
[a]The chi-square for entry is based on the likelihood ratio test.

Table 5: Model fitting information.

Model	Model fitting criteria	Likelihood ratio tests		
	−2 log likelihood	Chi-square	df	Sig.
Intercept Only	122.773			
Final	42.354	80.420	33	0.000

model for Table 4 (27.812) and Table 5 (80.420). These differences are significant ($p < 0.001$), which signifies the final MLR model explains a significant amount of original variability; this has a better fit than the original model [27].

The Model Fitting Information such as the Pearson and deviance statistics, and pseudo R^2 describe how well the model fits the data and whether the models' predicted values differ significantly from their observed data. Since both the Pearson and deviance statistics are not significant ($p = 1.000$), it

can conclude that the predicted and observed values are not significantly different. Hence, the model is a good fit.

Similarly, a model with a good fit can be shown by measuring over dispersion (difference in the distribution of predicted and actual data). The over dispersion values that were computed by eqns (10) and (11) are not high indicating a good fit [27].

$$\Phi_{\text{Pearson}} = \frac{\chi^2_{\text{Pearson}}}{\text{df}} = \frac{51.814}{117} = 0.443 \tag{10}$$

$$\Phi_{\text{Pearson}} = \frac{\chi^2_{\text{Deviance}}}{\text{df}} = \frac{42.354}{117} = 0.362. \tag{11}$$

3.2.2.4 Interpretation of MLR continuous/categorical variables

The MLR models for the eight distracting activities in risk zones I and II were fitted and a higher significant level ($p \leq 0.10$) recommended by researchers [15,24] for predictors and coefficient estimate (*B*), standard error, Wald statistic, and odds ratio (OR) along with 95% confidence intervals (CIs) for each urban transit distracting activity were computed by SPSS 17.0 [28]. The significance of each variable was tested by the Wald Statistic and the corresponding significance (*p*) value. The multinomial coefficient estimates were interpreted using their magnitude and direction proposed by Washington *et al.* [15] and Hosmer *et al.* [24] together with the OR guidelines [29,30].

3.3 Validation of results

The MLR models developed for risk zones I and II distracting activities have been statistically assessed in Section 3.2.2 using goodness of fit tests. Are the sample results generated by the MLR model linear predictors for risk zones I and II distracting activities also valid for a large random population of bus transit drivers? Three methods of validating the results are presented in the following subsections: expert verification, simulation, and route observations.

3.3.1 Expert verification

Expert verification by safety managers in the transit agencies is the starting point for validation. Standardized expert verification forms have been designed [2] for quick, low-cost verification of results.

3.3.2 Simulation

Computer simulation is commonly used by transportation researchers to validate output results from a model. The MLR linear predictors for pedestrian were simulated using probabilistic distributions to generate driver attributes, driving pattern, type of bus, and distraction events that would occur in practice over a range of random factors. Monte Carlo simulation was applied to generate the probability value $\pi(x)$ from eqn (4) for a range of 100 drivers getting *Slightly Distracted, Distracted,* and *Very Distracted.* The $\pi(x)$ values were plotted graphically and then compared with the results from the estimated coefficients of MLR linear predictors.

3.3.3 Route observations

Route observation is very useful for a quick distraction study as well as for validation of statistical models. A standardized form [2] can be used to collect route data for rapid determination of causes of some distraction activities. For example, Passengers Trying to Talk to the Driver is a high-risk

distracting activity in the transit agency. This type of distraction was commonly observed in some routes [2].

3.4 Guidelines for results interpretation and usage

The interpretation of descriptive statistics related to the urban drivers' attributes is summarized in Table 1 and results for each significant predictor variable follows from the magnitude and direction (sign) of the estimated coefficients of each MLR linear predictor. The interpretation of ORs follow the approach of McHugh [29] and are used as a *broad stroke estimate* [18] of the impact of the predictor variable on the response variable. One variable is interpreted at a time while the other variables are kept fixed. The interpretation of the results for distracting activity 'Pedestrian' is presented as an illustration of the guidelines (refer to Table 3).

3.4.1 Reported age of driver
Earlier studies concluded that driver age had a significant impact on distraction, with younger and older drivers more prone to distraction [30]. The MLR model reveals positive impact of age on distraction. The coefficient of the variable Age for external distracting activities such as Pedestrians is a positive value (0.264), which indicates that keeping everything else fixed, as age of the driver increases, they are more likely to get distracted by the Pedestrians. Figure 3 from the simulation output confirm the MLR results for pedestrians. Older drivers get more distracted by external activities such as pedestrians compared with younger drivers although an earlier study by Tefft [31] found the accident rates were higher for younger non-transit bus drivers.

3.4.2 Number of years of experience driving a bus
The negative sign for the coefficient of Experience (−0.487) indicates that keeping everything else fixed, increasing the years of experience reduces the likelihood of the driver getting Slightly Distracted by Pedestrians. This matches the graphical analysis in Fig. 2 presented earlier in Section 3.2.1, which shows that the less experienced drivers have a higher number of accidents. Simulation output in Fig. 4 validates the MLR model results.

3.4.3 Driving time
Peak = 1, non-peak = 0. The negative coefficients (−5.937) associated with Peak = 0 implies that when driving time changes from non-peak (0) to peak (1), the probability of getting distracted decreases. Therefore, non-peak drivers were more likely than peak drivers to get distracted by pedestrians. Figure 5 shows the simulated results of the impact of non-peak (0) driving on pedestrian distraction. The mean probability values for non-peak drivers (0.917) getting distracted is higher
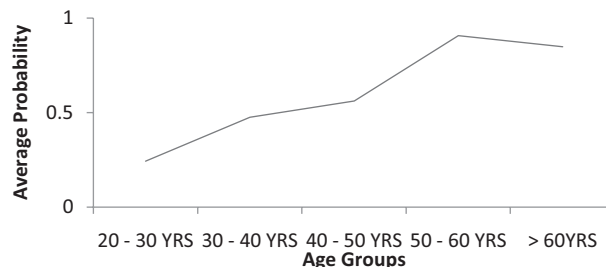

Figure 3: Driver distraction probability function for pedestrians (distracted) [2].
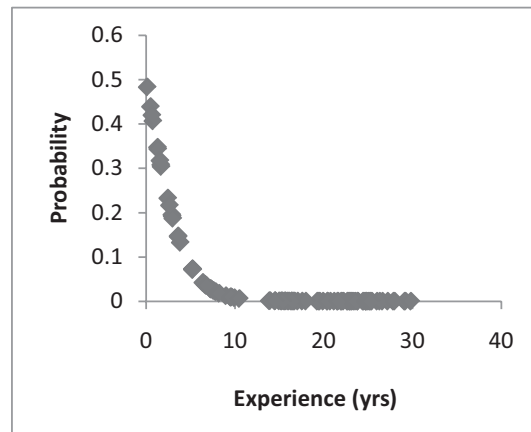
Figure 4: Driver distraction probability function for pedestrian (slightly distracted) [2].
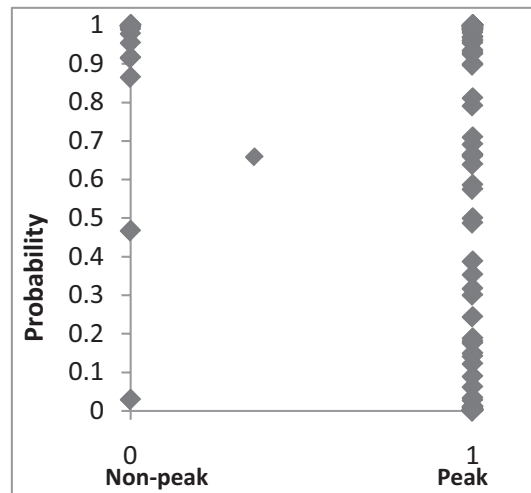

Figure 5: Driver distraction probability function for pedestrian (distracted) [2].

than peak drivers (0.55). The number of accidents is the highest on Fridays between the peak hours of 12:00 to 6:00 PM (preventable and non-preventable) [2].

3.4.4 Weekly driving hours

More driving hours per week would result in higher levels of fatigue that may cause higher distraction. Fatigue is a contributory factor in a large number of accidents. As a driver becomes more fatigued, she/he has a higher chance of getting distracted that may result in an accident. The urban drivers reported that they drive a bus for an average of 37.14 h per week and that they typically drive the buses mostly during the day (62%) and peak times (22%). The driver drives an average close to the normal 40 h per week; hence, fatigue may not be a significant cause of distraction but requires further analysis. The positive coefficients associated with DRIVING/WK (0.178) implies that holding all other independent variables constant, the higher the driving hours/week, the more likely the

driver would get Distracted by pedestrians. The odds ratio (1.194) is >1 and the 95% CI does not include 1 for the Distracted response variable. If a driver increases her/his driving hours/week by 1 h, the odds of getting Distracted to Not Distracted would increase by 1.194 times given the other independent variables are held fixed.

The above results demonstrate the impact of risk factors related to driver attributes such as age (Fig. 3), experience (Fig. 4), and driving pattern (Fig. 5) on the level of distraction. The other factors related to service location and types of bus driven are discussed in D'Souza and Maheshwari [2]. These factors have implications in the day-to-day operations of transit buses to provide safe transportation. Using the above results to develop related training for drivers could improve performance and safety of operations of the transit agencies.

4 CONCLUSIONS

This paper attempted to consolidate independent procedures for studying bus transit driver distraction into a modular research framework. It is the first of its kind to standardize methodologies for data collection, analysis, validation, and interpretation of results into a workable framework that could be used by transit agencies contemplating a driver distraction study. A transit agency in the Commonwealth of Virginia planning to conduct a bus driver distraction study could choose relevant tools from the modules according to the time available and budgetary limits such as a quick, low-cost study like analysis of existing accident databases maintained by the agencies or route observations, to a relatively longer duration, higher cost study involving field data collection, statistical modeling, analysis, and simulation.

Among the useful findings was similarity in the highest risk distracting activities for regional and urban bus transit drivers, which were mainly due to passengers, pedestrians and other road users. While many of the distraction related factors such as the service area (regional/urban), driver attributes (age, gender, driving experience, educational level, marital status, etc.), driving pattern (driving schedule, driving hours per week, service location, etc.), and type and age of the buses were significant in either regional and urban models, there were few that impacted both the transit agencies concurrently. The reasons for the resulting differences could be due to significant variations in driver attributes, driving pattern, type, and age of buses between the transit agencies. Hence, training needs and policies to curb distracted driving may differ at both agencies although the same modules of the research framework were used in both agencies.

As additional information becomes available from studies conducted at other agencies, the modules can be updated accordingly. The expanded data set can be used for validation as well as further refinement of the proposed framework. The proposed modular structure of the framework developed in this research permits updating and adding tools in each module as and when required without affecting the other modules. The four modules outlined in this framework is only a start and is expected to get updated and grow as more studies are conducted at other transit agencies and fresh results are acquired.

The MLR model presented in this paper analyzes the common risk factors due to driver attributes and external factors. It is recommended that future research focuses on longitudinal observational studies that examine drivers' psycho-physiological signal to evaluate road safety level [32], psycho-physiological analysis of mental load [33], and psychomotor performance during shift timings [34] under various road and traffic conditions. The static DRI developed in Section 3 could be linked to the driver's dynamic psycho-physiological parameters that changes with occurrence of distracting events in the traffic environment and the current static DRI could possibly be replaced with a *dynamic driving safety index* (DDSI) that could be possibly integrated with existing in-vehicle information systems (IVIS).

REFERENCES

[1] BBC, Spain Santiago rail-crash driver in new recording. *BBC News*, Europe, 5 September, 2013.

[2] D'Souza, K.A. & Maheshwari, S.K., A research framework for studying transit bus driver distraction. Conducted by Eastern Seaboard Intermodal Transportation Applications Center (ESITAC). USDOT, RITA TRID Database, Accession Number 01495385. www.trid.trb.org, pp. 79, 2013.

[3] D'Souza, K.A. & Maheshwari, S.K., Research framework for studying transit bus driver distraction. *Proc. of the Urban Transport 2013 Conf.*, Kos, Greece, ed. C.A. Brebbia. Wessex Inst. of Tech. Press: Southampton, UK, pp. 137–148, 2013. doi: http://dx.doi.org/10.2495/ut130111

[4] Ranney, T.A., Driver distraction: a review of the current state-of-knowledge. *DOT HS 810 787*. National Highway Traffic Safety Administration Final Report, 2008.

[5] U.S. Department of Transportation, An examination of driver distraction as recorded in NHSTA databases. *DOT HS 811 216*, 2009.

[6] U.S. Department of Transportation, Statistics and facts about distracted driving, available at www.distraction.gov/stats-and-facts, 2010.

[7] McEvoy, S.P., Stevenson, M.R. & Woodward, M., The prevalence of, and factors associated with serious crashes involving a distracted activity. *Accident Analysis and Prevention*, **39**, pp. 475–482, 2007. doi: http://dx.doi.org/10.1016/j.aap.2006.09.005

[8] Hickman, J.S., Hanowski, R.J. & Bocanegra, J., Distraction in commercial trucks and buses: assessing prevalence and risk in conjunction with crashes and near-crashes. *FMCSA-RRR-10-049*, U.S. DOT, Washington, DC, 2010.

[9] National Safety Council, Understanding the distracted brain: why driving while using a hand-free cell phones is risky behavior. *National Safety Council, White Paper*, April 2012.

[10] 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,** pp. 602–610, 2011. doi: http://dx.doi.org/10.1016/j.apergo.2010.07.007

[11] Salmon, P.M., Young, K.L. & Regan, M.A., *Analysis of risk for State Transit Authority New South Wales bus drivers. Bus Driver Distraction Stage 1: Final Report*, Monash University Accident Research Centre: Clayton, Victoria,Australia, 2006.

[12] D'Souza, K.A. & Maheshwari, S.K., Multivariate statistical analysis of public transit Bus Driver Distraction. *Journal of Public Transport, Rural and Intercity Bus,***15(3)**, pp. 1–23, 2012. doi: http://dx.doi.org/10.2495/ut130111

[13] McCartt, A.T., Hellinga, L.A. & Braitman, K.A., Cell phones and driving: review of research. *Traffic Injury Prevention*, **7**, pp. 89–106, 2006. doi: http://dx.doi.org/10.1080/15389580600651103

[14] Stuffs, J.C., Reinfurt, D.Wm., Staplin, L. & Rodgman, E.A., *The Role of Driver Distraction in Traffic Crashes*. AAA Foundation for Traffic Safety: Washington, DC, 2001. doi: http://dx.doi.org/10.1037/e363942004-001

[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*, 2nd edn., Chapman & Hall/CRC: pp. 303–359, 2011.

[16] 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)**, pp. 63–72, 2007. doi: http://dx.doi.org/10.5399/osu/jtrf.46.1.981

[17] Gkritza, K., Niemeier, D. & Mannering, F., Airport security screening and changing passenger satisfaction: an exploratory assessment. *Journal of Air Transport Management*, **12,** pp. 213–219, 2006. doi: http://dx.doi.org/10.1016/j.jairtraman.2006.03.001

[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)**, pp. 193–205, 2009. doi: http://dx.doi.org/10.1080/01488370802678983

[19] Salvucci, D.D., Chavez, A.K. & Lee, F.J., Modeling effects of age in complex tasks: a case study in driving. *Proc. of the 26th Annual Conf. of the Cognitive Science Society*, pp. 1197–1202, 2004.

[20] 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, Paper 4340, November 2005.

[21] U. S. Department of Transportation, *Distracted Driving*, DOT HS 811 650, NHTSA, 2010.

[22] Agresti, A., *An Introduction to Categorical Data Analysis*, 2nd edn., John Wiley and Sons, Inc.: New York, NY, 2007. doi: http://dx.doi.org/10.1002/sim.3564

[23] Tabachnick, B.G. & Fidell, L.S., *Using Multivariate Statistics*, 6th edn., Pearson Education, Inc.: NJ, 2013.

[24] Hosmer, D.W., Lameshow, S. & Sturdivant, R.X., *Applied Logistic Regression*, 3rd edn., Wiley Series in Probability and Statistics, Wiley and Sons, Inc.: New York, NY, 2013. doi: http://dx.doi.org/10.1002/9781118548387

[25] Peng, C.Y. & Nichols, R.N., Using multinomial logistic models to predict adolescent behavioral risk. *Journal of Modern Applied Statistical Methods*, **2(1)**, pp. 1–13, 2003.

[26] Moutinho, L. & Hutcheson, G., Store choice and patronage: a predictive modelling approach. *Int J. Business Innovation and Research*, **1(3),** pp. 233–252, 2007. doi: http://dx.doi.org/10.1504/ijbir.2007.012109

[27] Field, A., *Discovering Statistics Using IBM SPSS Statistics*, 4th edn., Sage Pub. Ltd : Thousand Oaks, 2013.

[28] SPSS, Inc., *SPSS 17.0*, SPSS, Inc., Chicago, IL, 2008. doi: http://dx.doi.org/10.1177/1558689808317830

[29] McHugh, M.L., The odds ratio: calculation, usage, and interpretation. *Biochemia Medica*, **19(2)**, pp. 120–126, 2010. http://www.biochemia-medica.com/content doi: http://dx.doi.org/10.11613/bm.2009.011

[30] Olson, R.L., Hanowski, R.J., Hickman, J.S. & Bocanegra, R.J., Driver distraction in commercial operations. *FMCSA-RRR-09-042*, U.S. Department of Transportation Federal Motor Carrier Safety Administration, Washington, DC, 2009.

[31] Tefft, B.C., Motor vehicle crashes, injuries, and deaths in relation to driver age: United States, 1995–2010. *AAA Found. for Traffic Safety*, 2012, available at www.aaafoundation.org.

[32] Chung, B.-J., Park, J.-B., Kim, J.-Y. & Kim, J.-Y., A study on analysis methodology of driver's psycho-physiological signal to evaluate road safety level. *Journal of the Eastern Asia Society for Transportation Studies*, **5,** pp. 2607–2617, 2003.

[33] Richter, P., Wagner, T., Heger, R. & Weise, G., Psychophysiological analysis of mental load during driving on rural roads – a quasi-experimental field study. *Ergonomics*, **41(5)**, pp. 593–609, 1998. doi: http://dx.doi.org/10.1080/001401398186775

[34] Vivoli, R, Rovesti, S., Bussetti, P. & Bergomi, M., Psychomotor performance of truck drivers before and after day shifts. *Traffic Injury Prevention*, **14(8)**, pp. 791–796, 2013. doi: http://dx.doi.org/10.1080/15389588.2013.779686