Final Report

A RESEARCH FRAMEWORK FOR STUDYING TRANSIT BUS DRIVER DISTRACTION

Kelwyn A. D'Souza
Professor
School of Business
Hampton University
Hampton, VA 23668
757-727-5037
kelwyn.dsouza@hamptonu.edu

And

Sharad K. Maheshwari
Associate Professor
School of Business
Hampton University
Hampton, VA 23668
757-727-5605
sharad.maheshwari@hamptonu.edu

Eastern Seaboard Intermodal Transportation Applications Center (ESITAC)
Hampton University
Hampton, VA 23668

September, 2013

DISCLAIMER

TABLE OF CONTENTS

ACKNOWLEDGEMENT

Hampton University's Eastern Seaboard Intermodal Transportation Applications Center (ESITAC) thanks the Management of Hampton Roads Transit (HRT) and Potomac and Rappahannock Transportation Commission (PRTC) for permission to conduct the driver distraction study. We are grateful for efforts of HRT's Office of Risk Management and PRTC's Office of Planning and Operations for organizing and conducting the bus operator survey, and in particular, the dispatch staff and bus drivers that participated in the survey. We also owe thanks to our student research interns, Ms. Alexa Hollinshead for conducting the route observations and creating the databases, and Ms. Lexis Phillips for assistance in analysis of accidents and preparing this final report.

A RESEARCH FRAMEWORK FOR STUDYING TRANSIT BUS DRIVER DISTRACTION

Executive Summary

The increase in transit bus ridership nationally during the past ten years, along with the proliferation of personal electronic control and communication gadgets is causing more distractions for the drivers. Earlier research studies have found driver distraction to be a significant cause of accidents on the highway. The transit bus driver distraction has received limited attention in the literature even though transit bus accidents may cause more injuries due to larger number of passengers. Unlike personal vehicles, where most of the distraction is within the control of the driver, 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. Due to lack of reporting of distractions by transit drivers and very limited research studies, transit bus driver distraction is not clearly understood and no established research framework is available to conduct a detailed study of transit bus drivers' distraction. Every study at a transit agency is conducted independently from start to finish at the cost of additional time and resources.

The objective of this project was to develop a modular research framework for studying transit bus driver distractions. The framework would provide a transit agency with a set of standardized modular methodologies for studying distraction over a wide range of cost and time intervals. An agency could choose one or more modules to suit their study requirements. These 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 results are reproduced in this report for purposes of illustrating the type of outputs obtainable from the framework and are not meant to reflect the accuracy of the data and model results.

The major deliverables from this project are as follows:

- The research framework structure consisting of four modules.
- Standardized processes for data collection and analysis.
- Identifying sources, duration, and driver's perception of distraction.
- Method of classifying distracting activities into risk zones.
- Appropriate statistical models to determine the significant factors that impact the levels of distraction.
- Methods to access model fit and validate the model results.
- Guidelines on interpreting and using the results.
- Recommendations for improving the model results.

A RESEARCH FRAMEWORK FOR STUDYING TRANSIT BUS DRIVER DISTRACTION

1. INTRODUCTION

Driver distraction is the cause of a large proportion of traffic accidents and has been studied by researchers and government agencies in the U. S. and other countries around the world [1, 4, 6, 25, 34, 35, 49, 50]. However, transit bus driver distraction is not as well studied and lacks 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, thus costing additional time and resources. The distraction studies conducted for regional and urban transit bus drivers [13, 31] by the Eastern Seaboard Intermodal Transportation Applications Center (ESITAC) in the Commonwealth of Virginia using modules for data collection, analysis, validation, and results interpretation, has demonstrated that the methodology could be standardized for other transit agencies.

The objective of this project was to develop a modular research framework for studying transit bus driver distractions at any agency. The framework would provide a transit agency with a set of standardized modular methodologies for studying distraction over a wide range of cost and time intervals whereby an agency may choose one or more modules to suit their study requirements. These 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 results are reproduced in this report for purposes of illustrating the type of outputs obtainable from the framework and are not meant to reflect the accuracy of the data and model results.

The primary research questions addressed in the study are:

- What are the common sources and durations of transit bus driver distraction?
- Can the transit bus driver distraction activities be classified into risk zones?
- Are bus driver attributes such as age, educational level, marital status, experience, gender, and driving hours/week together with driving pattern, and type of bus related to driving distraction?
- Can the risk of distraction for existing and new drivers be predicted?

The research framework presented in Section 4 offers an agency the option of choosing one or more modules for conducting a driver distraction study. The tools necessary for studying the sources and durations of driver distractions, the risks associated while engaging in potential distracting activities, and visual, manual, and cognitive factors that are responsible for distraction are combined together to form the structure of the research framework.

The distracting activities can be classified into different risk zones in the exploratory data module. The confirmatory data analysis module can be used to establish statistical relationships between level of distraction and factors causing distractions. The agencies will have the option of validating the results using simple approaches such as expert opinion and route observations to advanced simulation models. Guidelines for interpretation of distraction results are proposed along with recommendations on how to improve the accuracy of the model results.

This research framework evolved from studies conducted at two transit agencies that provide service to over 25 million riders per year at eight cities and surrounding suburbs in the Commonwealth of Virginia. One agency is a regional transit agency employing around 460 non-seasonal drivers that provides service to 22 million riders per year in six cities and surrounding suburbs [13]. The other is an urban transit agency employing around 150 drivers through a contract service provider [31]. It provides service to around 3.5 million riders per year through commuter bus service along the busy I-95, I-495, and I-66 corridors to points north, local bus services in Prince William County and the cities of Manassas and Manassas Park, and a free ridesharing service. Both these agencies are a good representation of the transit agencies located in other parts of the Commonwealth. Hence, the proposed research framework could be used by other transit agencies in the Commonwealth of Virginia that are planning a driver distraction study but could expect different results.

2. BACKGROUND

Driver distraction represents a significant problem in the personal and public transport sector and has been studied by several national and international researchers, and governmental agencies [1, 4, 6, 25, 34, 35, 49, 50]. A study funded by the AAA Foundation [1] identified the major sources of distraction that causes crashes in personal vehicles, developed a taxonomy of driver distractions in the U.S., and examined the potential consequences of these distractions on driving performance.

The source of bus driver distractions at a major Australian public transit 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 [34]. In a ground-breaking study, Salmon et al. [35] developed a taxonomy of distraction sources and duration for bus drivers at the State Transit Authority, New South Wales (STA, NSW), Australia. In this study, a taxonomy of the sources of distraction was developed for transit bus drivers and a descriptive statistical analysis was conducted. The limited sample size of 18 drivers comprising of 16 males and two females provided insufficient data for an inferential statistical analysis. D'Souza and Maheshwari [6] expanded the exploratory work of Salmon et al. [35] using multivariate statistical models and simulation to confirm the impact of driver attributes, driving pattern, and location on the distracting activities.

Studies on the impact of driver attributes and driving patterns on driving performance have produced mixed results. 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 [28]. Blower et al. [4] reported that age, sex, hours driving, trip type, method of compensation, and previous driving records are related to driver errors. A significant difference in reaction times between the age groups supports the hypothesis that difficulty of processing multitasks increases with age but no significant difference in reaction time was found between males and females, and no interaction was reported between gender and age [28]. Factors such as location, driving hours/week; and driver age, gender, and experience have an impact on public bus driver distraction [28, 46]. A driving route running through a densely populated area would generally service a larger number of passengers and experience higher distraction due to external sources like more frequent stops, higher traffic, other road users, and/or pedestrians [1]. A driver less familiar with the driving routes is more likely to be involved in rear-end accidents at signalized intersections [55].

Multivariate statistical models are widely used in transportation to study the relationship between a categorical response/dependent variable (DV) consisting of two or more levels and a set of continuous and categorical predictors/independent variables (IVs). The multivariate model applied by Yan et al. [56] utilized multinomial logistic regression (MLR) to study the impact of potential factors such as driver characteristics, road layout, and environmental conditions on rear-end truck to car, car to truck, and car to car crashes. Washington et al. [53] developed a multinomial logit (MNL) model consisting of 18 independent variables covering driver factors, traffic flow, distance, and 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 mode 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. [53]. A MLR model was developed by Morfoulaki et al. [23] 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. [9] 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 [30] computed the odds ratios for the tasks/variables, along with 95 % confidence intervals (CIs) to identify the high risk tasks/variables and the strength of association between the categorical dependent variable and independent variables. Following the approach of Washington et al. [53] and Morfoulaki et al. [23], D'Souza and Maheshwari [6, 7] proposed an MLR model to analyze public transit bus driver distraction that included five IVs linked to a categorical DV with four levels of distraction.

The Monte Carlo simulation method was applied to validate empirical results obtained from the conceptual models. 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 [36, 37]. A simulation approach was developed by Smith et al. [39] to evaluate the impacts on safety that occur when drivers become distracted by secondary tasks, and the approach was tested using data collected from test tracks and on-the-road trials. These simulation results were used to compute a *Hazard Index* that measured the potential for a collision to occur due to a driver's being distracted.

Researchers have developed frameworks for different transportation applications. Preliminary work [5] on the research framework has been conducted in the U. S. utilizing results from this project that outlines methods for data collection, analysis, and interpretation of results that a transit agency could readily use to conduct a bus driver distraction study. Salmon et al. [34] proposed a framework of ergonomic methods for assessment of transit bus driver distraction which includes the analysis of tasks, identification of distraction sources, and risk assessment. Wong and Huang, [54] have proposed a research framework for studying driver's mental process to determine how accidents occur which includes a conceptual framework of the driving mental process; that is a step towards development of a workable model to study accident causality. Trick et al. [45] have provided a conceptual framework that combines the two fundamental dimensions of attention selection in order to have a more comprehensive driving theory. Although these studies [34, 45, 54] are not directly related to the research framework for conducting a driver distraction study, their methodologies provide useful inputs for developing the framework in this project.

3. METHODOLOGY

Several transit agencies serving cities, counties, and surrounding areas in the Commonwealth of Virginia were invited to participate in this bus driver distraction project. Two agencies servicing eight cities, counties, and surrounding suburbs accepted the invitation and participated in the study and three agencies declined (Table 1). Driver distraction information was collected from the participating transit agencies' past two to three years accident databases, self-administered surveys, and discussions with agency staff. The accident database format generated from police reports differed slightly in each agency though the basic information on accident type, date and time, and driving experience remained the same.

Table 1. Selected Transit Agencies and Cities in the Commonwealth of Virginia.

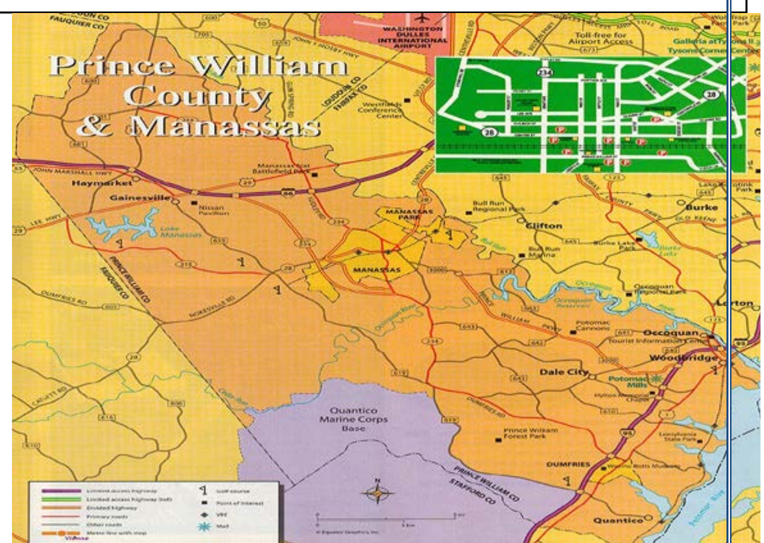
City	Transit Agency	Participated In The Study
Hampton	Hampton Roads Transit (HRT)	YES
Chesapeake	Hampton Roads Transit (HRT)	YES
Newport News	Hampton Roads Transit (HRT)	YES
Norfolk	Hampton Roads Transit (HRT)	YES
Virginia Beach	Hampton Roads Transit (HRT)	YES
Richmond	Greater Richmond Transit Co.	NO
Fredericksburg	Fredericksburg Regional Transit	NO
Harrisonburg	Harrisonburg Transit	NO
Prince William County	Potomac and Rappahannock Transportation Commission (PRTC)	YES
City of Manassas	Potomac and Rappahannock Transportation Commission (PRTC)	YES
Manassas Park City	Potomac and Rappahannock Transportation Commission (PRTC)	YES

The Hampton University Transit Bus Driver Distraction Survey (Appendix 1) was adapted from Salmon et al. [35] to suit local conditions and it was approved by the participating transit agencies and the Hampton University Institutional Review Board (IRB). The survey was administered anonymously and voluntarily with full disclosure during drivers' breaks, before their shifts began, and after their shifts were completed to avoid any disruption of their normal routine. At each transit agency, a representative was assigned to distribute the surveys, deliver the introduction, answer questions, and assist in the survey process.

The study was conducted in two parts: The first part conducted at a regional transit agency in June – August 2012 covered two locations of Hampton Roads: the Northside (Peninsula) that included the cities of Hampton and Newport News; and the Southside that included the cities of Norfolk, Virginia Beach, and Chesapeake. (Figure 1). Both locations included counties and surrounding areas. These locations differ in population density, street layouts, and accident rates. The Southside is more commercialized and densely populated with an accident rate of 62 accidents/million miles compared to the Northside where the accident rate is 54 accidents/million miles [13]. The second part conducted in September – December 2012 at an urban transit agency covered commuter service (I-95, I-66, and I-495 corridors), metro feeders, local services (Prince William County, City of Manassas, and Manassas Park City), and cross county connector (Figure 2). These locations differ in population density, street layouts, and accident rates. According to the 2011 Virginia Traffic Crash Facts [52], Prince William County had around 280,605 licensed drivers and totally 5,221 crashes while the Cities of Manassas and Manassas Park had 27,737 and 8,084 licensed drivers with totally 594 and 73 crashes indicating a positive relationship between number of licensed driver and total crashes.

Figure 1. Hampton Roads Region *(Wikipedia, July 2011)*

Figure 2. Potomac and Rappahannock *www.princewilliamcountywebsite.com/maps.htm*



The survey instrument was distributed to drivers in the regional and urban transit agencies. A sample of 77 regional drivers out of the 250 drivers surveyed responded resulting in a response rate of 31%. And a sample of 53 urban drivers out of 150 drivers surveyed responded resulting in a response rate of 27%. The survey responses reflected the perceptions of the drivers who were the primary sources for distraction-related information. Their responses were fairly consistent and comparable with responses from other transit bus driver distraction studies [35].

The data collected from the accident databases and surveys were analyzed for eliciting accident and distraction patterns. The accident data were classified into non-preventable and preventable accidents which included the accidents caused by distractions. Various descriptive statistics were calculated such as accidents occurring during hours of the day, day of the week, and the relationship between accidents and driving experience together with accidents caused by driver distractions. The distracting activities that were rated by the drivers on the survey instrument were ranked [35] and graded relative to the maximum values. The grades were used to compute the Distraction Risk Index (DRI) which is conceptually similar to the *Hazard Index* developed by Smith et al. [39]. The DRIs were used to classify the distracting activities 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*. Furthermore, the eight distracting activities in Risk Zone I and II were statistically analyzed to elicit the factor(s) that impact the distraction levels.

A variety of tools are included in the validation module for confirming the model results. Finally, guidelines for interpreting and using the results are provided for implementation and improving the driver's performance. 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 that will be discussed later in Section 4.

Lastly, Section 5.2 recommends approaches for enhancing the quality of results from existing data and models by pre-analysis screening methods to check accuracy of the data, missing data, extreme values or outliers, and fulfillment of necessary assumptions [22]. The inaccurate data, missing values, outliers, and uneven splits of 90% – 10% or worse reported in the Case Processing Summary were corrected by methods recommended by researchers [2, 3, 11, 15, 22, 42]. The number of levels for the outcome variable was collapsed from four to three by combining *Distracted* and *Very Distracted* levels into a single level: *Distracted/Very Distracted*. Almost all tests indicated a good fit for the model with transformed outcome variable (having three levels). A total of 10 variables are significant in the transformed model as against 4 variables in the original model (a model with four levels in the outcome variable).

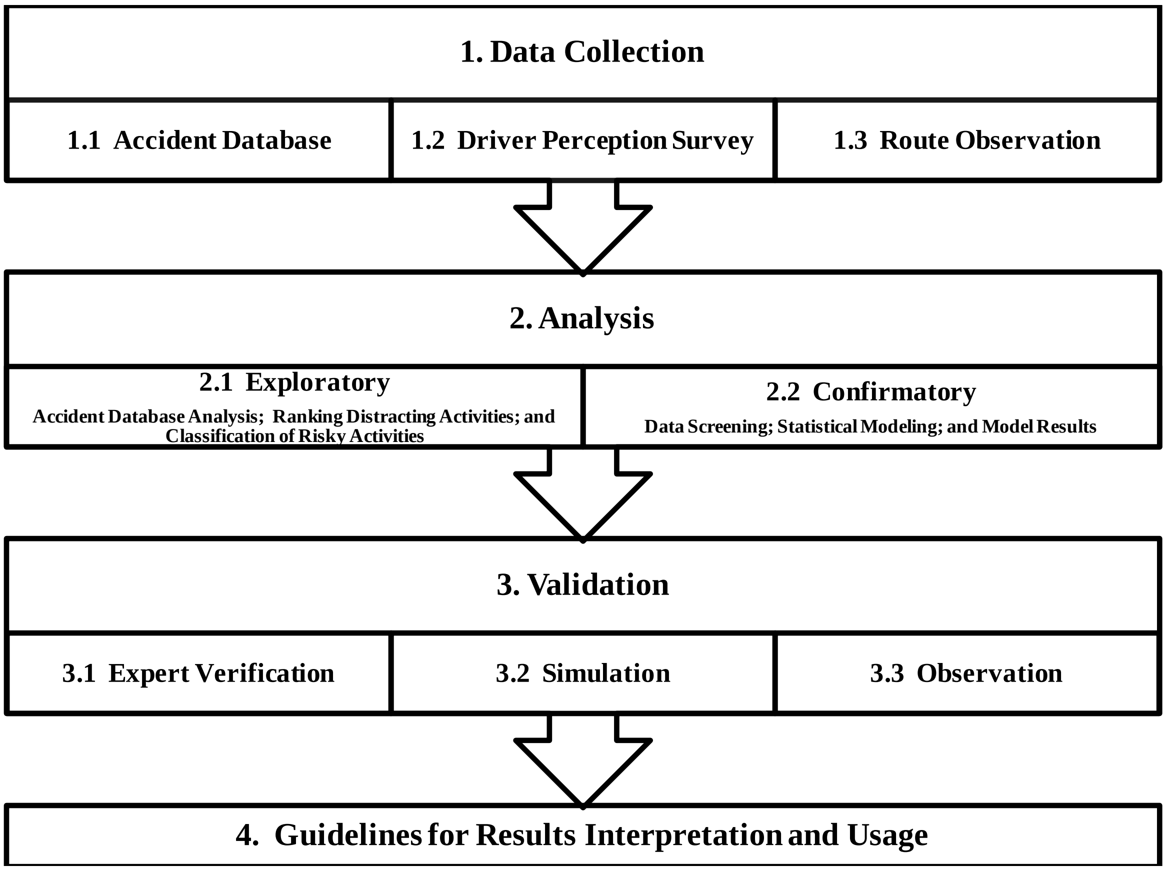
4. THE RESEARCH FRAMEWORK

An updated schematic outline of the preliminary research framework proposed earlier by D'Souza and Maheshwari [5] for a conducting transit bus driver distraction study is presented in Figure 3. It has four modules, each containing ready-to-use standardized methods that were tested at the regional and urban transit agencies. Each module is discussed in the following sub-sections:

4.1. 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 each of these sources.

Figure 3. Research Framework Outline



4.1.1. Accident Database

The transit agencies generally collect accident data from police reports. The past two to three years accident databases for the regional and urban transit agencies 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. 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 are caused by driver distraction but the proportion is unknown as drivers generally do not report distraction as a cause of their accident.

The accidents resulting from distractions are not normally recorded but reported estimate of 17% [48] of the total accidents may be applied to compute the number of accidents due to driver distraction. According to the USDOT [50], *distraction-affected crashes are preventable*. Hence, some of the agencies' preventable accidents could have been caused by driver distraction but the proportion is unknown. In this study, an estimate of 17% [48] of the total accidents was applied to compute the number of accidents due to distracted driving (Table 2).

Table 2. Estimated Accidents due to Driver Distraction.

Location Of Accident	Non-Preventable	Preventable	Driver Distraction*	Total
NORTHSIDE	553	84	131	768
SOUTHSIDE	1124	261	284	1669
TOTAL	1677	346	414	2437
% OF TOTAL ACCIDENTS	69%	14%	17%	100%

* 17% of Total (DRIVER DISTRACTION related accidents are part of PREVENTABLE accidents).

4.1.2. Driver Perception Survey

The HU Transit Bus Driver Distraction Survey (Appendix 1) was used to collect driver attributes, driving patterns, and type of bus driven along with the driver's perception of distraction. Data collected using a survey instrument is more extensive for the analysis of driver distraction factors as well as distraction prediction. The survey instrument could be modified by the transit agency for the purpose of collecting data tailor-made for its distraction related factors. The self-administered survey instrument containing 70 – 80 items were grouped under the following sections:

I. Driver Attributes, Driving Pattern, and Type of Bus: Driver's age, gender, education level, driving experience at the agency, driving hours/week, service location, schedule, and type of bus driven.
II. Source and extent of distraction.
III. Duration of distraction.
IV. Perceived effect of distraction.

For illustration purposes, Table 3 provides a summary of driver attributes, driving location and driving pattern for the transit agency.

4.1.3. Route Observations

Data on driver distraction can also be collected via route observations. A Route Observation Form shown in Appendix 2 could be used to collect route data that will help rapid determination of some distraction factors. Observers can ride the bus on selected routes having relatively high accident rates and record any type of distraction along with possible causes. 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. Some training of the observer maybe necessary to ensure accuracy and consistency of the observations.

Table 3 Summarized Data Collected by the Survey Instrument.

Characteristic — **Statistics**

1. Gender: Males = 74%, Females = 26%
2. Age (Years): Mean = 47, Std. Dev = 9.69
3. Driving Exp (Years): Mean = 8, Std. Dev = 8.35
4. Marital Status:

MARRIED	65%
SEPARATED	2%
DIVORCED	13%
NEVER MARRIED	13%
NOT REVEALED	7%

5. Education Level:

< HIGH SCHOOL	2%
HIGH SCHOOL	44%
SOME COLLEGE	20%
2 YR COLLEGE	20%
4 YR COLLEGE OR HIGHER	14%

6. Driving Hrs/Wk: Mean = 37.14, Std Dev = 15.2

7. Driving Service Location.

COMMUTER	64%
LOCAL	19%
METRO FEEDER	8%
OTHERS	6%
NO RESPONSE	3%

8. Driving Schedule:

DAY	62%
NIGHT	9%
PEAK	22%
NON-PEAK	7%
OTHERS	0%

9. 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%

10. Avg. Age of Bus = 9.5 years, Std. Dev. = 3.7 years.

The age distribution is as follows:

LESS THAN 5 YEARS	18%
6 YEARS – 10 YEARS	39%
11 YEARS – 15 YEARS	38%
MORE THAN 15 YEARS	5%

4.2. Analysis

The data analysis has been categorized into exploratory data analysis (EDA), and confirmatory data analysis (CDA). In the application of EDA, the data is first screened for accuracy, missing values, extreme values (or outliers), and violation of assumptions. The resulting high quality data is transformed largely into graphical or tabular models. EDA focuses on discovering characteristics and patterns in the data through a wide range of techniques ranging from simple graphs, classifications, and tables to advance techniques such as decision trees and neural networks, The EDA offers the flexibility of choosing one or more models for conducting a driver distraction study. The EDA models can provide ample information to a transit agency to identify the driver attributes and external factors that contribute to distraction. This knowledge can help to develop effective policies to mitigate risk of accidents. The EDA would also help in establishing various hypotheses which can be formalized and tested if the CDA models are constructed. At the CDA, models and hypotheses are statistically tested and validated so that further steps can be taken to interpret and apply them within the transit agency.

4.2.1. Exploratory Data Analysis (EDA)

The exploratory methods analyze the data collected to identify patterns that may provide preliminary results on the distraction activities. It develops a Distraction Risk Index (DRI) which is used to classify each distracting activity into risk zones.

4.2.1.1 Accident Database Analysis

The accident data can be very useful in conducting EDA 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 privacy or other reasons). An analysis of historical accident data for the past two to three years is recommended to identify causes of accidents. The city could be divided into different locations (for example Northside and Southside) 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.

The Two-Way Contingency Table 4 illustrates estimation of the number of accidents due to driver distraction and other causes for each location.

Table 4. Estimated Number of Accidents Due to Distracted Driving.

Location Of Accident	Driver Distraction	Other Causes	Total
NORTHSIDE	131	637	768
SOUTHSIDE	284	1385	1669
TOTAL	415	2022	2437
% OF TOTAL ACCIDENTS	17%	83%	100%

The following approach of Agresti [3] is applied to the accident data in Table 4 to predict the probability of accidents due to driver distraction.

Joint, Marginal, and Conditional Probabilities (Refer to Table 5)

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 a tabular from which the predictive probabilites can be computed. Suppose a driver is selected at random and then classified on the basis of X and Y.

Table 5. Distracted Driving Events

Location Of Accident	Driver Distraction (Event B_1)	Other Causes (Event B_2)	Total
NORTHSIDE (Event A_1)	$n_{11} = 131$	$n_{12} = 637$	$n_{1+} = 768$
SOUTHSIDE (Event A_2)	$n_{21} = 284$	$n_{22} = 1385$	$n_{2+} = 1669$
TOTAL	$n_{+1} = 415$	$n_{+2} = 2022$	$n = 2437$
%	17%	83%	100%

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

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 $n = \Sigma_{i,j}\, n_{i,j}$

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

P(Accident in Northside) = $(n_{1+})/(n)$ = 768/2437 = 0.32
P(Accident in Southside) = $(n_{2+})/(n)$ = 1669/2437 = 0.67

What is the probability of a driver from the Northside (Event A_1) and will have an accident due to distraction (Event B_1)?

APPROACH 1: Difference of Proportions [3] can be used when the number of accidents due to distraction is available at the transit agency. Let P_1 and P_2 denote the conditional probabilities of an accident (success) in the Northside or Southside. The *difference of proportions* $P_1 - P_2$ compares the probabilities of an accident occurring (success) in the Northside and Southside.

$P_1 - P_2$ is estimated from the sample difference found in the Contingency Table 5.
The 95% Confidence Interval for $P_1 - P_2 = \{P(B_1 \mid A_1) - P(B_1 \mid A_2)\} \pm Z_{\alpha/2}(SE)$ ……. (1)

If the interval contains only positive values, it can be concluded that $P_1 - P_2 > 0$ or $P_1 > P_2$. Therefore, probability of accidents due to distraction is higher in the Northside.

This approach cannot be used on Table 5 since the number of accidents caused by driver distraction was unknown and had to be estimated as 17% of the total accidents. In this case, since $P_1 = P_2 = 0.17$, it is recommended to use the following Approach 2.

APPROACH 2:
Using the general rule of multiplication $P(A_1 \text{ and } B_1) = P(A_1)P(B_1 \mid A_1)$ = (768/2437)(131/768) = 0.055

What is the probability of a driver from the Southside (Event A_2) and will have an accident due to distraction (Event B_1)?

Using the general rule of multiplication $P(A_2 \text{ and } B_1) = P(A_2)P(B_1 \mid A_2)$ = (1669/2437)(284/1669) = 0.114

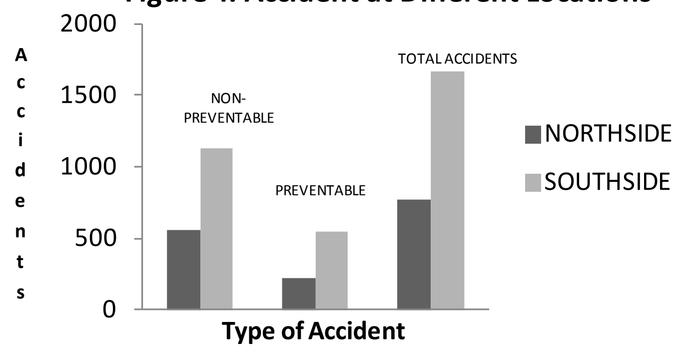
It is clear from the Table 5 data, that the overall probability of the accidents as well as the joint probability of accidents with distractions is higher (two times) in the Southside compared to Northside.

Analysis of Accident Data Using Graphical Models

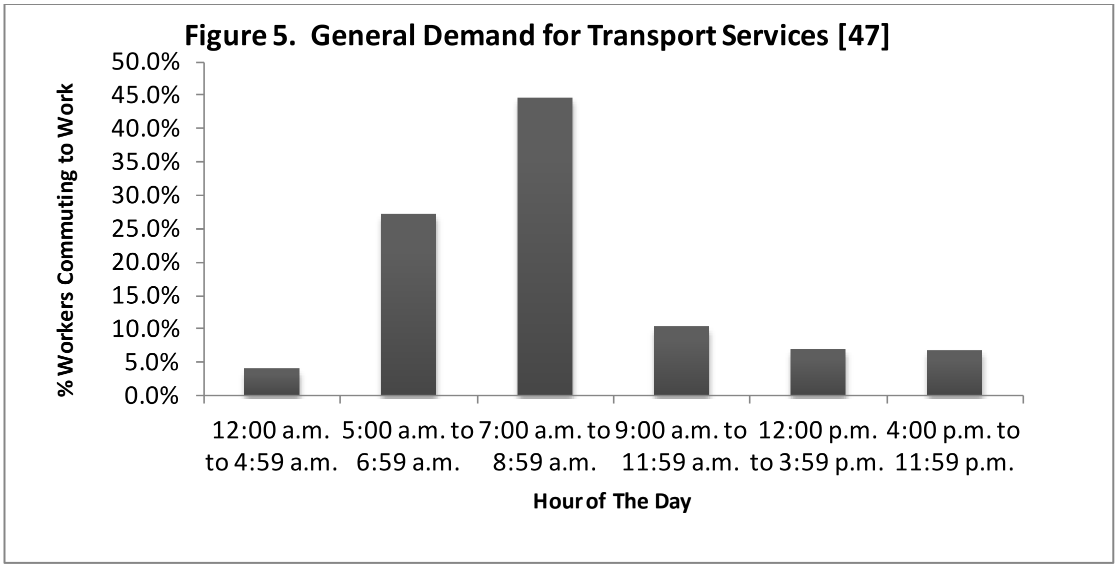
This accident related data can be utilized to get some estimate of the distracted driver activities. Uniform method of data collection and extraction will provide an overall view of accident data analysis as illustrated in Figure 4 which summarizes preventable and non-preventable accident data at two distinct locations in a city. 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.

According to Title 7 of the Transportation Code [44], "Daytime" means the period beginning one-half hour before sunrise and ending one-half hour after sunset. "Nighttime" means the period beginning one-half hour after sunset and ending one-half hour before sunrise. The agency's drivers reported (Table 3) that they drive mostly during the day (62%) and peak times (22%). Only 7% drive during non-peak hours and 9% drive at night. This schedule coincides with the general demand for transport services shown in Figure 5 which is higher during the daytime and peaks between 7:00 AM and 9:00 AM [47].

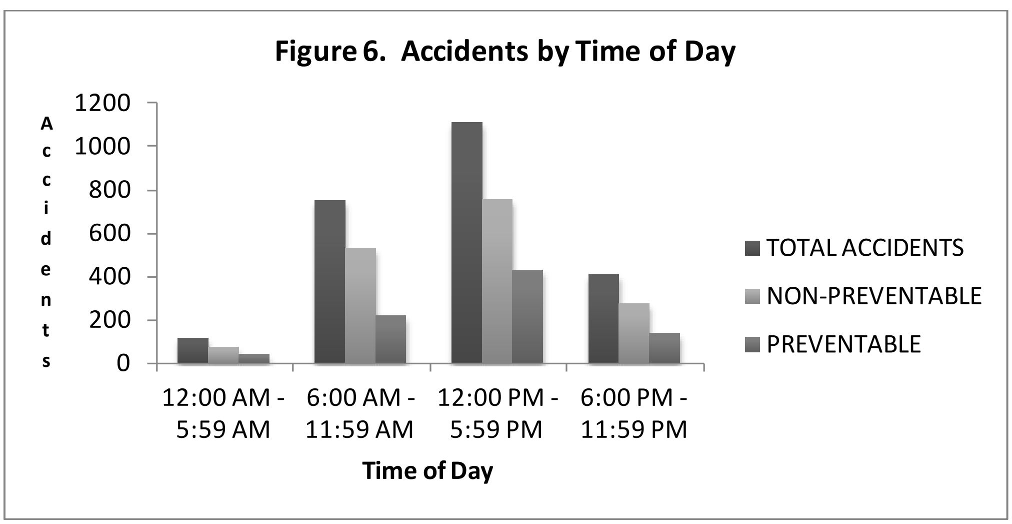
Figure 4. Accident at Different Locations



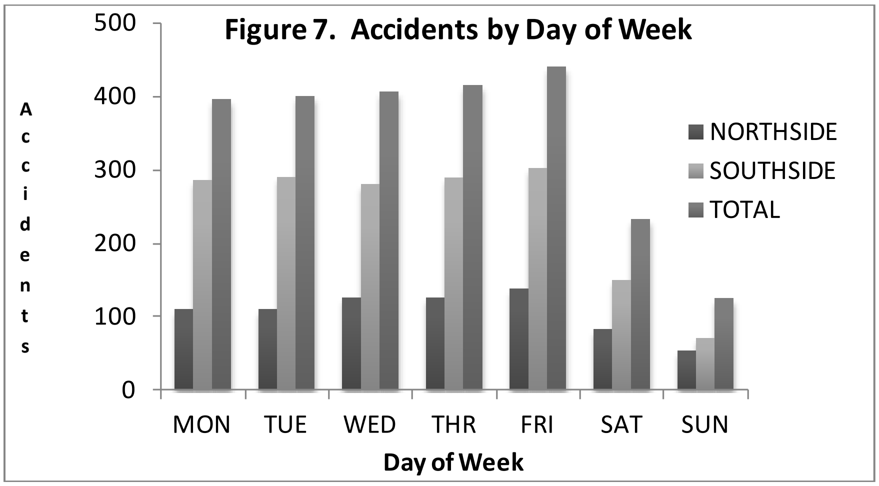
Studies by the USDOT [48] report that most accidents occur between 3:00 PM and 9:00 PM. However, the highest rate of accidents as a percentage of cars on the road occurs between Midnight and 3:00 AM. In general, the risk of accident is higher at night than during the day. Per mile driven, the nighttime fatal involvement rate for drivers of all ages was 4.6 times the daytime rate [19]. The difference varied with age of the driver. However, among drivers 20-24, the nighttime fatal rate was 6.1 times the daytime rate, but among drivers 75 and over, the nighttime rate was only 1.1 times the daytime rate [19]. Figure 6 shows the highest number of accidents for the agency occurs between 12:00 PM to 6:00 PM.



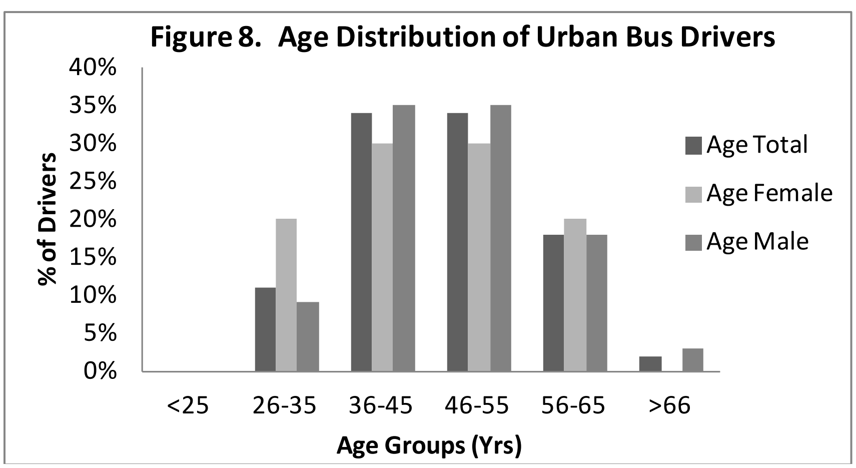
The number of accidents is dependent to the days of the week with Fridays having the highest number of accidents in the Southside compared to Northside. Being the end of the week, it is expected that Fridays would have a lot more distraction due to fatigue than other days. Therefore, the highest number of accidents due to driver distraction is on Fridays. 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. The number of accidents at the agency gradually rises between Monday and Friday and then decreases (Figure 7) with the highest number of accidents occurring on Friday. Early Friday evening appears as the worst day of the week for fatalities and serious injuries on the road.

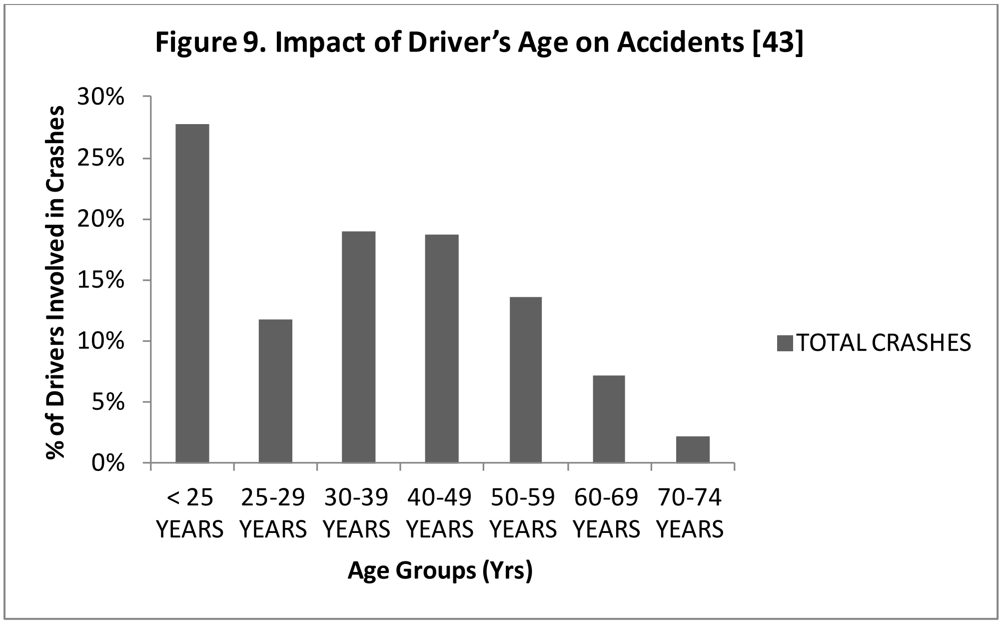


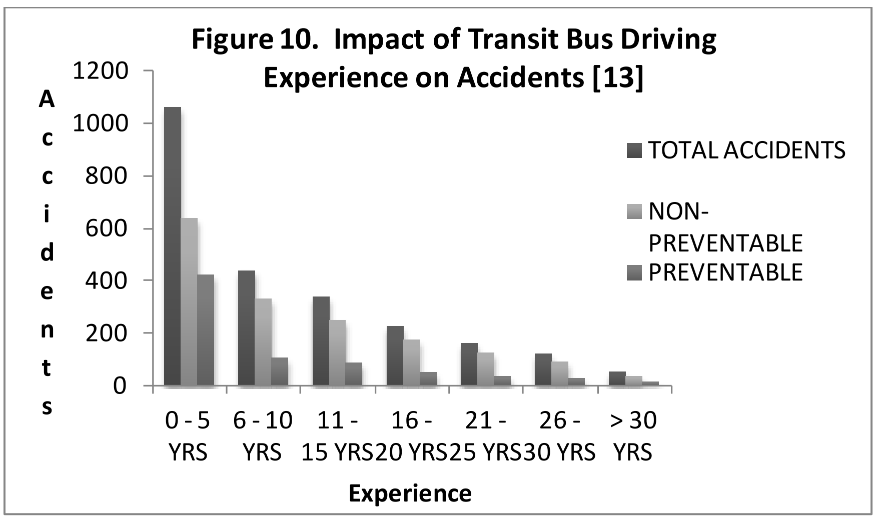
The average age of a bus driver at the urban agency is 47 years (male = 47.4 years, female = 45.1 years). Most of the male and female drivers fall in the 36-55 year age group (Figure 8). A study by Tefft [43] shows age as a significant factor related to accidents with younger drivers (< 25 Years) who are more prone to accidents and distracted driving (Refer to Figure 9). The 36 - 55 year age group accounts for around 30% of the crashes in the U. S. Hence, the agency may not face this problem since around 68% of its drivers are between 36 - 55 years with zero drivers below 25 years.



A driver's experience in driving transit buses influences her/his driving performance. Less experienced drivers perform common driving tasks without thinking (for example slowing down before making a turn). Younger drivers believe that they have the cognitive capability under all driving conditions until an accident proves them wrong [25]. Results obtained from the regional transit agency study [13], reveals that less experienced drivers have higher accidents than the more experienced drivers (Figure 10). Since, less experienced drivers are generally young, it is clear that young, inexperienced transit bus drivers are at an increased risk to themselves are also a major hazard for other road users.







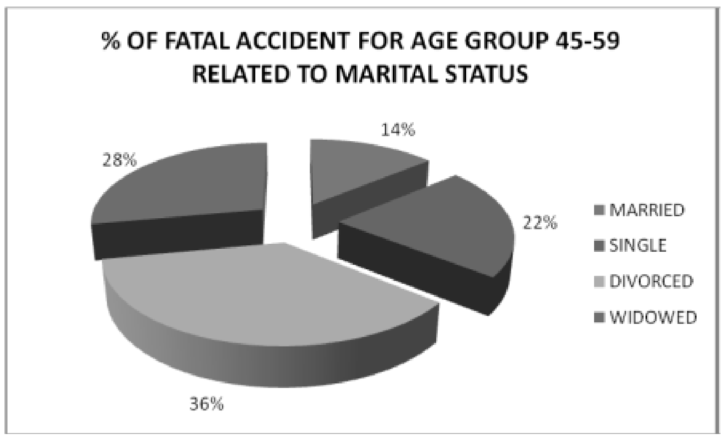
The urban transit bus drivers had an average of eight years experience driving buses. It is believed that experienced drivers would get less distracted by the distracting activities due to the fact that they would be older and have been driving long enough to be affected by distracting activities. Studies on the impact of age on driving performance suggests that younger (below 25 years) tend to be more vulnerable to the effects of distraction than middle-aged drivers [43].

Personal and emotional behavior of drivers was found to influence accident rates. Personal responses to stressful life events are suspected of increasing the risk of serious traffic accidents. A study by Lagarde et al. [17] showed that stressful events in a driver's personal life increase the risk of serious traffic accidents including at-fault accidents. Marital separation or divorce was associated with an increased risk of a serious accident and recent separation and divorce are associated with an increase in serious traffic accidents [17]. Furthermore, insurance premiums are lower for married people compared to single people

[12] because married people on an average have lower accidents and hence, get less distracted. This means that being unmarried has been associated with increased risk of accidents.

Figure 11 developed from a study by Norman [26] shows variation in fatal accidents according to marital status for the 45 – 59 year driver age group (similar to the urban study group with mean age as 47 years). For drivers in this age group, married men and women have the lowest accident rate (14%). The rates increase for single (22%) and widowed (28%) with the divorced group having the highest rate: over two and half times (36%) the married rate [26]. Hence, marital separation or divorce was associated with an increased risk of fatal accidents. Around 65% of the urban transit bus drivers are married which could lower the risk of distraction related accidents.

Figure 11. Impact of Marital Status on Accidents [26]



Higher education does not necessarily result in fewer distractions and accidents. Well educated drivers have much more activities going on in their professional lives which requires more reading, using cell phones, and texting, some of which takes place while driving. Studies show that well educated drivers are more prone to be in an accident or near-miss accident due to being distracted while driving [16]. A study by Powell et al. [32] reported that college graduates have a 27% chance of being involved in a “sleepy accident” compared to a 12% chance for a high school graduate. The national average educational level of a transit bus driver is a High School Diploma or equivalent. At the urban agency, around 44% of the drivers are HS graduates and 54% have reported having some college or a college degree. Hence, education level of drivers at the agency has to be analyzed further to study its impact on distraction.

The Gillig model bus is mostly driven by regional drivers is the Gillig model and the MCI bus model is mostly driven by urban drivers. The proportion of accidents due to mechanical failure is small since the buses undergo periodic maintenance and state mandated inspections. The bus design should ensure safety and comfort of the driver hence reducing the risk of distraction due to fatigue which could lead to accidents. This must include wide windshields for better road visibility, driver seating capacity to accommodate a wide range of body sizes, and ergonomically designed dashboards for easy reach and operations.

The age of the transit bus plays a major role in safety. Older buses develop more maintenance problems and on-board rattles that can be highly distracting. In 2010, the average age of a full size transit bus in the U. S. was 7.8 years [41]. The average age of the urban agency's bus is 9.5 years (Table 3) but over 40% of the buses are more than 10 years old which is above the national age. The urban agency undertakes an exhaustive mid-life maintenance on the buses which restores the older buses to an equivalent average age of 4.1 years.

4.2.1.2. Classification of Risky Activities

Driving tasks can be defined as everything that is needed to operate the transit vehicle. These driving tasks are divided into primary driving tasks and secondary driving tasks. Common examples of primary driving tasks for a transit driver are steering, using the accelerator, applying the brakes, changing lanes, determining what speed to use, and communicating with other drivers by using the turn signal and turning on the headlights, etc. In contrast, secondary driving tasks that cause distraction are non-driving activities estimated at approximately one third of moving time [33]. The internal secondary driving tasks generally include conversing with passengers, tending to passengers with infants, collecting fares, making announcements, using a navigation system or other wireless device, managing climate control, etc. The transit drivers are also distracted by external events such as other road users, pedestrians, etc. When transit drivers focus their attention on secondary driving tasks, their attention is diverted from the primary driving tasks causing distractions that may lead to an accident.

From prior studies and discussions with agency's staff, and the self-administered survey, the study team identified between 20 to 23 distracting activities at the transit agencies. These activities were classified into risk zones according to relative grades that were estimated for rating, duration, and perception of distraction. The classification of all the distracting activities into risk zones would reveal the (few) high risk activities that 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. The classification was carried out in the following steps:

<u>Classification Step 1</u>. *Ranking*: The driver's rating of each distracting activity and the estimated duration that they experienced these activities in a typical eight-hour shift were averaged for all drivers and ranked from highest to lowest [35] as shown in Appendices 3 and 4. The ratings and durations for the top five distracting activities are shown in Tables 6 and 7. Four of the top five distracting activities in Table 6 were passenger related and three out of five distracting activities in Table 7 were related to internal secondary activities and condition of the bus.

The average distraction ratings and durations were compared for males and females (Appendices 5 and 6). In these comparisons, many of the top distracting activities for the average rating were passenger-related activities while for the average duration it was internal secondary activities.

Table 6. Top Five Ratings of Distracting Activities

Rank	Activity	Average Distraction Rating	Related Category
1	Unruly Kids	2.85	Passenger
2	Passenger Using Mobile Phone	2.70	Passenger
3	Driver's Mobile Phone	2.60	Technology
4	Passengers Not Following Etiquette (eating, drinking, smoking, noisy)	2.37	Passenger
5	Passengers	2.36	Passenger

Table 7. Top Five Durations of Distracting Activities.

Rank	Activity	Average Distraction Duration (Hrs)	Related Category
1	Pedestrians	2.45	External Infrastructure
2	Other Road Users	2.25	External Infrastructure
3	Announcing Bus Stops	2.00	Operational
4	Ticket Machine/ Farebox	1.98	Operational
5	On-board Rattles	1.89	Bus Cabin

Classification Step 2. *Driver's Perception*: The USDOT [49] has categorized driver distractions as visual, manual, and cognitive and reported that the severity of distraction increases as it involves more than one category. The survey collected information on these three categories of distraction along with the number of bus drivers that categorized each of the perceived impact of distracting activities (Appendix 7). The activities were sorted by number of drivers and the top five for each category of effects are summarized in Table 8. For example, in the case of Visual Effects of Distraction, the distracting activities that were rated by the highest number of drivers (13) were ranked 1. Once again, the most common distracting activities were passenger-related.

Classification Step 3. *Distraction Risk Index (DRI)*
Each distracting activity's rating and duration was graded as a percentage (%) relative to the highest rating (2.85) and highest duration (2.45 hours). The number of driver responses for distracting activities in each category was graded as a percentage (%) relative to the highest eyes off the road (13 driver responses), mind/attention off the road (28 driver responses), and physical interference (17 driver responses).

The graded scores for rating and duration of distraction, eyes off the road, mind/attention off the road, and physical interference distractions are summarized in Table 9. The graded scores of each distracting activity were averaged to produce the Distraction Risk Index (DRI) that measures the potential risk associated with each distracting activity.

Table 8. Top Five Ranking of Distraction Categories as Perceived by the Driver

Activity	Distraction Category	Eyes Off the Road (Visual Effects of Distraction) 1 - 5	Mind/Attention Off the Road (Cognitive Effects of Distraction) 1 – 5	Physical Interference (Manual Effects of Distraction) 1 - 5
Passengers using a mobile phone	Passenger	-	4	-
Unruly Kids	Passenger	4	-	4
Looking at Advertisement	Infrastructure	-	-	-
Passengers with Infants	Passenger	-	-	4
Other Road Users	Infrastructure	-	-	-
Reading (eg Route Sheet)	Operational	1	-	-
Ticket Machine/ Farebox	Technology	3	-	-
Climate Control	Technology	3	-	4
Passengers	Passenger	1	-	-
Disabled Passengers	Passenger	-	-	1
Fatigue/Sickness	Personal	-	-	2
Pedestrians	Infrastructure	2	-	-
Passengers not following etiquette (eating, drinking, smoking, noisy)	Passenger	5	-	-
Passengers trying to talk to you	Passenger	-	5	-
General Broadcast	Operational	-	2	-
Personal Broadcast	Operational	-	-	-
Looking At Advertisements	External	1	-	-
Food and Other Smells	Passenger	-	-	-
Audible Alerts	Operational	-	3	-
Onboard Rattles	Operational	5	-	-
Other Road Users	External	1	-	-
Dispatch Broadcast	Operational	-	1	4
Communication with Dispatch	Operational	5	-	3
Mobile Data Terminals	Operational	4	-	5
Driver's Mobile Phone		5	-	-
Announcing Bus Stops	Operational	-	-	4

The average DRIs of the 23 distracting activities listed in Table 9 ranged from 50–72 percent with a mean of 60 percent and standard deviation of 5.8 percent. Following the approach of Peng and Nichols [29], distracting activities scoring a DRI of at least one standard deviation above the mean, i.e., higher than 66 percent were identified as Risk Zone I (very high risk) activities. Those scoring higher than the mean 60 percent but up to 66 percent were identified as Risk Zone II (high risk) activities. Similarly, the range for Risk Zone III (moderate risk) activities was set at DRI scores higher than at least one standard deviation below the mean i.e. more than 54 percent and up to 60 percent, and the range for Risk Zone IV (low risk) was set at DRI scores less than or equal to 54 percent.

Table 9. Graded Scores and Distraction Risk Index for Each Distracting Activity

Distracting Activities	Distraction Rating (% OF HIGHEST)	Distraction Duration (% OF HIGHEST)	Eyes off the road (% OF HIGHEST)	Mind/Attent. off the road (% OF HIGHEST)	Physical Interference (% OF HIGHEST)	Average (DISTRACTION RISK INDEX)	RISK ZONE
Pedestrians	74%	100%	92%	68%	24%	72%	I
Passengers (moving around, standing and talking next to driver's cabin)	83%	64%	100%	75%	24%	69%	I
Other Road Users	73%	92%	100%	57%	18%	68%	I
Unruly Kids	100%	45%	77%	71%	35%	67%	I
Passengers Using Mobile Phone	95%	75%	54%	89%	18%	66%	II
Mobile Data Terminals	80%	71%	77%	75%	29%	66%	II
Passengers not following etiquette (eating, drinking, smoking, noisy)	83%	72%	62%	75%	24%	63%	II
Ticket Machine/ Farebox	47%	81%	85%	75%	18%	61%	II
On-board Rattles	68%	77%	62%	71%	24%	60%	III
Communication with Dispatch	70%	46%	62%	82%	41%	60%	III
Looking at Advertisements	79%	48%	100%	57%	18%	60%	III
Passengers Trying to Talk to Driver	81%	57%	54%	86%	24%	60%	III
Fatigue and Sickness	82%	42%	38%	54%	82%	60%	III
Climate Control	59%	63%	85%	50%	35%	58%	III
Driver's Mobile Phone	91%	49%	62%	64%	24%	58%	III
Disabled Passengers	53%	49%	54%	32%	100%	58%	III
Announcing Bus Stops	52%	82%	31%	79%	35%	56%	III
Reading (e.g. Route Sheet)	64%	44%	100%	50%	24%	56%	III
Dispatch Broadcasts	60%	50%	23%	100%	35%	54%	IV
Food and Other Smells	67%	50%	54%	75%	24%	54%	IV
Passengers with Infants	68%	59%	38%	71%	35%	54%	IV
General Broadcasts/ Other	68%	49%	23%	96%	24%	52%	IV
Audible Alerts	54%	66%	23%	93%	12%	50%	IV

The graded scores of all distracting activities with the DRIs are classified into risk zones according to the DRI range shown in Table 10. Four out of the 23 distracting activities were classified into Risk Zone I, four into Risk Zone II, ten into Risk Zone III, and the remaining five into Risk Zone IV.

Table 10. Classification of Distracting Activities into Risk Zones

DRI RANGE	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

4.2.2. Confirmatory Data Analysis (CDA)

In Section 4.2.1.1, EDA was conducted using contingency tables to estimate the impact of a predictor variable (location) on response variable (driver distraction). The 2x2 contingency table (Table 4) can only analyze a single variable at a time. In order to analyze several variables simultaneously, it is necessary to utilize models [3].

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 method like driver survey could be more detailed as well as would reflect existing conditions and perception of drivers.

Multivariate statistical models are suitable to analyze the high risk distracting activities using levels of distraction as the response/dependent variable and correlating it with predictor/independent variables. For example, the categorical dependent variable (driver distraction) had four levels: *Not Distracted*, *Slightly Distracted*, *Distracted*, and *Very Distracted*. The independent variables included categorical variables: gender, marital status, educational level, driving schedule, and location, as well as continuous variables: age, driving experience, and driving hours per week. The research hypotheses were to *determine the likelihood that the transit bus driver getting Slightly Distracted, Distracted, and Very Distracted with respect to Not Distracted is related to her/his pertinent demographical characteristics, driving pattern, location, and type of bus.*

4.2.2.1. Statistical Modeling

The survey collected nominal (discrete) and scalar (continuous) data about the drivers including demographical details and information about their driving pattern, service location, and type of bus commonly driven. Furthermore, the survey also collected data on the source of distraction, duration, and the driver's perception of the type of distraction (visual, physical, and cognitive) caused by the activities.

Each distracting activity listed in the survey captured four responses from the drivers: *Not Distracted, Slightly Distracted, Distracted, and Very Distracted.* The possible factors causing the distractions are a combination of discrete categorical variables such as gender, location, marital status etc., and continuous variables such as age, driving experience, and driving hours per week. The categorical nature of the response (dependent variable) and predictors (independent variables) violate the linearity, normality, and continuity assumptions required for linear regression models. Therefore, a multiple linear regression model was not suitable for studying the relationship between the distracting activity and the variables that were causing the distractions.

Generalized Linear Model

The *generalized linear models* (GLMs) broadly refer to a wide range of statistical models that include continuous DVs such as regression as well as models for discrete or categorical DVs [14]. GLMs extend the use of regression and analysis of variance to discrete or categorical DVs which are non-linearly related to the IVs through the use of three components [3]: The *random component* which is the response or DV having a probability distribution. The *systematic component* enumerates the explanatory or IVs. And the *link function* connects the *random component* to the *systematic component* and indicates the relationship between both components. This structure of the GLM was utilized to develop a multivariate model for studying the impact of driver attributes, driving pattern, and type of bus on the distracting activities.

Multinomial Logistic Regression

The (GLM) technique of multinomial logistic regression (MLR) was developed to determine the variables that have an explanatory impact on the level of risk zone distracting activities. The estimated coefficients of the IVs allow determination of the factors responsible for increasing or decreasing the risk of distractions. The advantage of using MLR is that it can handle discrete, dichotomous, and continuous predictors or independent variables (IVs), and nonlinear categorical outcome or dependent variables (DVs) with less stringent requirements as compared to linear regression models [42].

The outcome is the distraction activity having risk-levels 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. An activity that causes the driver to get *Very Distracted* is more likely to lead to an accident.

The increasing risk of the four categories would suggest an ordered discrete probability model. According to Washington et al. [53], ordered models may not be suitable for such applications since they can restrict the impact of predictor variables on the response variables. An ordinal logistic regression test model developed for the regional transit agency [13] exhibited a poor fit ($p = 0.381$) with no significant IVs. Due to these issues and the reported problems with the restriction imposed by the ordered logit model [9, 53], an unordered discrete outcome model (MLR) was used in this study even though the response variable appeared to be ordered.

MLR Model Formulation

The MLR model was developed as an extension of the logistic regression [8, 15, 24] 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. When comparing with the multiple linear regression, the logistic regression predicts the *probability of the outcomes* occurring given known values of the predictors while the multiple linear regression model predicts the *value of the outcomes* from known predictors but with more stringent requirements.

Theoretical Framework of the Logistic Regression Model

In this study, the logistic regression model was developed as a type of GLM comprising of the three components discussed earlier [3].

The *random component* Y is the random DV of the logistic regression having an independent set of observations (Y_1, ……Y_n). In a logistic regression, Y is binary (*Distracted* or *Not Distracted*) and assumed to follow a *binomial distribution* [3]. Y can represent the occurrence (success) of a distraction activity level (for example, *Slightly Distracted*, *Distracted*, and *Very Distracted*) with reference to *Not Distracted*..

The *systematic component* represents right hand side of the GLM called *linear predictor* [3]. It is constructed by combining the explanatory variables as follows:

$$\beta_0 + \beta_1 x_1 + \beta_1 x_1 + \ldots \ldots\ldots\ldots + \beta_k x_k.$$

Where,

$\{x_k\}$ is a set of non-random fixed explanatory variables (x_1, x_2, ……..x_k).

β_0 is the model constant (Y intercepts) and β_1 ……, β_k are the regression coefficients corresponding to the $\{x_k\}$ which may be continuous or categorical. The coefficients β_k represents the change in the probability per unit change in x [3]. When β_k are positive values, increasing $\{x_k\}$will increase the logit of Y and conversely, when β_k are negative values, increasing $\{x_k\}$will decrease the logit of Y. For both cases, the reverse is true when $\{x_k\}$ is decreasing.

The third component called the *link component* is necessary to connect the *random component* to the *systematic component* [3].

The logistic regression has a binary response variable Y having two possible outcomes: Y = 1 (*success*) and Y = 0 (*failure*).

Let the probability of success P (Y = 1) = π. The probability of failure P (Y = 0) = (1 - π)

Since, Y follows the binomial distribution, the mean E (Y) = $\mu = n\pi$ and standard deviation

$$\sigma = \sqrt{n\pi(1 - \pi)}.$$

If n = number of independent observations. Then, n = 1 for each binary observation (distraction level).

Since, the value of π can vary as the value of x changes, π is replaced by π (x) [3].

For k observations having a binary response parameter π (x), the *linear probability model* is defined as:

$\pi(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \ldots\ldots + \beta_k x_k$ ………………………………….…. (2)

The principle mathematical theory behind logistic regression is the logit which is the natural logarithm (ln) of the odds of Y. The odds are defined as the ratio of the probability π (x) that event Y occurs (for example a driver gets distracted by Passengers) divided by the probability (1 - π (x)) that event Y will not occur (driver is not distracted by Passengers).

Or,

Odds $=\left[\frac{\pi(x)}{(1-\pi(x))}\right]$ ………………………………………………………………………….. (3)

Therefore,

logit (Y) = natural log [Odds] = $ln\left[\frac{\pi(x)}{(1-\pi(x))}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_k x_k$ ………. (4)

Taking antilog of Equation 4 on both sides, we derive the Equation 5 which predicts the probability of outcome of an event (for example, distraction level of an activity)

$\pi(x) = \left[\frac{e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k}}{1+e^{\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k}}\right]$ ……………………………………………………………. 5)

Equation 5 can be simplified as:

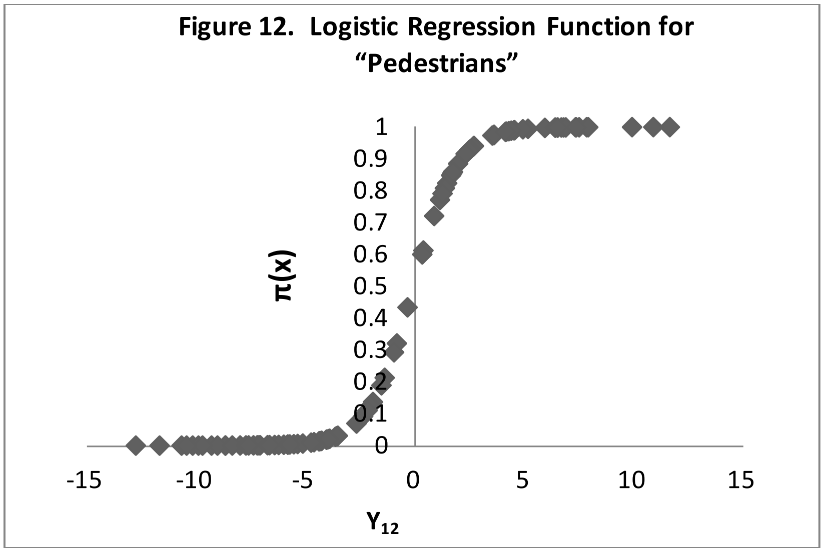
$\pi(x) = \left[\frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_k x_k)}}\right]$ ……………………………………………………….. (6)

Where, π(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.

π(x) increases continuously as x increases, taking the shape of an S-shaped graph [3]. As an illustration, the fitted relationship between π(x) and the *linear predictor* of distracting activity "Pedestrians" was:

Y_{12} = -13.47 +0.178*Drive Hrs/Wk + 0.264*Age – 5.937*Peak=0 ……………………………. (7)

Simulating the predictor variables, values obtained for Equation 7 were substituted into Equation 6. The S-shaped curve plotted in Figure 12, shows π(x) increases continuously from 0 to 1 as the value of the *linear predictor* Y_{12} increases from $-\alpha$ to $+\alpha$.



Applying Equation 3, the general MLR model can be expressed in logistic regression form. But since the *random component* cannot be linked directly to the *systematic component*, a non-linear link function called the *logit* must be used [24] as follows:

$$ln\left[\frac{\pi(Y=j)}{\pi(Y=j')}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \ldots\ldots + \beta_k x_k \text{ .... } (8)$$

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

Logit model (equation 9) comparing *Slightly Distracted* with *Not Distracted* could be s stated as:

$$ln\left[\frac{\pi(Y=Slightly\ Distracted)}{\pi(Y=Not\ Distracted)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \ldots\ldots + \beta_k x_k \text{ .... } (9)$$

Logit model (equation 10) comparing *Distracted* with *Not Distracted* is stated as:

$$ln\left[\frac{\pi(Y=Distracted)}{\pi(Y=Not\ Distracted)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \ldots\ldots + \beta_k x_k \text{ .... } (10)$$

Logit model (equation 11) comparing *Very Distracted* with *Not Distracted* is stated as:

$$ln\left[\frac{\pi(Y=Very\ Distracted)}{\pi(Y=Not\ Distracted)}\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \ldots\ldots + \beta_k x_k \text{ .... } (11)$$

The logit models 9, 10, and 11 provide three estimates for the impact each independent variable has on the dependent variable, allowing the impact of independent variable x_k to be computed for each logit model and for the whole model [53].

The multinomial *linear predictor* which measures the total contribution of the 10 factors (independent variables) for the transit agency is expressed as:

$Y_{ij} = \beta_0 + \beta_1 SEX + \beta_2 AGE + \beta_3 EXP + \beta_4 MARITAL + \beta_5 EDU + \beta_6 DRIVING/WK + \beta_7 LOCAT + \beta_8 DAY + \beta_9 PEAK + \beta_{10} EQUIP$ …………………………………………… (12)

Where,
SEX: Gender of driver, 1 = Male, 0 = Female.
AGE: Reported age of driver in years.
EXP: Number of years of experience driving a bus.
MARITAL: Marital Status*, 1 = Married, 0 = Others (Separated, Divorced, Never Married, etc.).
EDU: Educational Level*, 1 = HS or Equivalent, 0 = Others (Some College, 2,4 year degree, etc.).
DRIVING/WK: Weekly driving hours.
LOC: Location* of transit agency service area, 1 = Commuter, 0 = Others (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 = Others (Gillig, Orion etc.).

(*) Original multi level predictor variables were collapsed to dichotomous variables.

The model constant and coefficients β_0, β_1, β_2, ……. β_k are estimated by the *maximum-likelihood* method that estimate coefficients that make the observed values most likely to have occurred [8]. The coefficients computed by the MLR models are relative to the reference category and are utilized to predict the probability of the extent that a driver finds an activity distracting versus the reference category (*Not Distracted*). The Equations 9, 10, and 11 are solved by fitting the model with the observed data such that the values of Y_{ij} are close to the observed values. The least square method is used in linear regression while the logistic regression uses the *maximum-likelihood estimation* which selects the coefficients that makes the predicted value of Y_{ij} as close as possible to the observed values. This requires the development of the *likelihood* function which expresses the probability of the observed values as a function of the unknown coefficients [3]. By maximizing the function, the values of these coefficients can be computed. The SPSS 17.0 [40] software computes the estimates of each coefficient along with the Wald statistics, standard errors and significance levels, ORs, and 95 percent confidence intervals.

Pre-Analysis Data Screening

Valid conclusions can only be drawn from data that correctly represents the problem being studied. A few random errors in the input data values will not impact the accuracy of the results and most statistical packages delete such data by default. But a large number of non-random errors could impact the results drawn from the sample. This limits generalization of driver distraction for a wider population of bus drivers from the sample study of drivers. Hence a pre-analysis data screening [22] is recommended prior to the actual statistical analysis to detect accuracy of the data, missing data, extreme values or outliers, and fulfillment of necessary assumptions.

Data Accuracy

Inaccuracies in data collection and recording can occur due to human errors. The drivers can make wrong entries in the survey forms or the analyst could incorrect code the data during entry into the database. Such errors can be mitigated by including clear instructions for filling the survey forms and assigning a trained transit representative to distribute the surveys, deliver the introduction, answer questions, and assist in the survey process.

Prior to the data analysis, a careful inspection of the raw data is recommended. In the case of small data file, a comparison of the printed input data with the actual data may point out incorrect data entries, and for large data sets descriptive statistics such as mean, range, and standard deviations can be examined for discrepancies [22, 42]. Categorical data can be checked for any incorrect coding or if the coded values are out of range [42].

Missing Data from the Self Administered Surveys

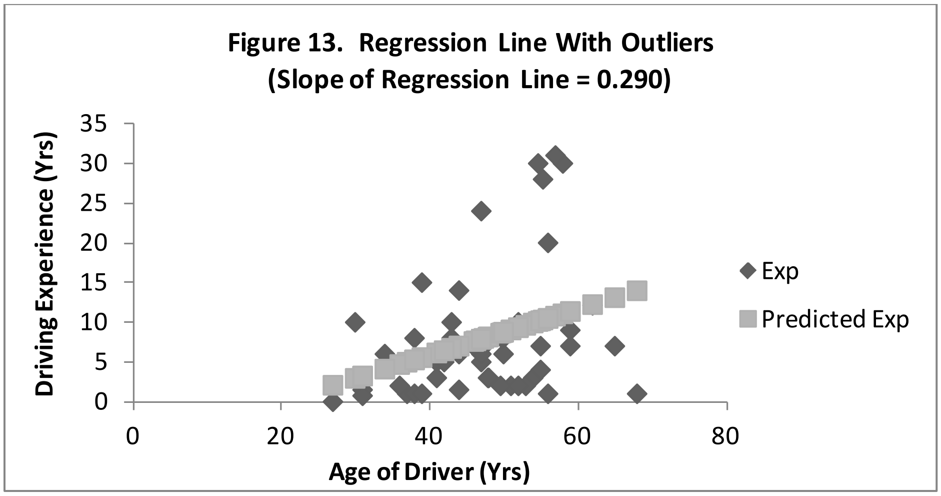
A major problem with the self administered survey conducted at the transit agency was the data that was *missing not at random (MNAR)* [42]. For example, some drivers were reluctant to provide their age, marital status, and educational level. A few (less than 5%) randomly missing data values will not cause substantial inaccuracy in the output results and can be dropped from the data set, but a larger number of non-random missing data values can affect the inferences drawn from the sample and generalization of results to a population of drivers [22, 42]. For small data set, researchers have recommended to repeat the analysis with and without the estimated missing values and choose the data set that better represents the true population [15].

Different approaches listed in Appendix 8 were used to treat the data missing in the survey. A few (three) surveys left completely unfilled were deleted from the database with a cautionary note attached in the report as suggested by Afifi et al. [2]. The *mean substitution* and *regression approach* [2, 22] were applied for imputation of missing distraction ratings of the response variable, and driver attributes, driving patterns and type of bus (predictor variables). The missing ratings in the survey for each response variable were filled using the method of *mean substitution* where the missing ratings were replaced by the mean rating. This conservative approach resulted in no change in the overall mean distraction rating but the variance of the variables is reduced along with reduction of the correlation with other variables [42]. The *regression* approach was used to estimate the missing data for age, experience, driving hours/week, marital status, education level, and type of equipment. *Prior knowledge* [42] was applied to estimate the missing data for day/night, peak/non-peak, and location.

Extreme Values or Outliers

Extreme values or outliers can occur at the upper or lower end of the data range. For example, a driver's age of 16 years or 75 years. The slope of a regression model is greatly influenced due to outliers [15]. As an illustration, the driving experience was plotted for different ages of the driver resulting in the slope of the regression line = 0.290 (Figure 13). Using the statistical method of *standardizing all raw scores* the age and experience data was transformed into z-scores (number of standard deviations away from the mean). In this study, a raw score in excess of $\pm$ 2.0 standard deviations was considered as a possible outlier and deleted from the data set. This modification of the input data with reduced number of outliers changed the slope of the regression line substantially to 0.119 (Figure 14).

The continuous variables in this study were examined for possible outliers. In the data set for driving hours per week, entries of 2, 70, and 80 hours appeared to be wrongly entered and were replaced by the mean hours per week = 37.14.



Ratio of Cases to Variables

Very little work has been reported on the sample size of cases required for logistic regression. Hosmer et al. [11] have reported that including several predictor or independent variables (IVs) could result in *multicollinearity* and recommends the *Rule of 10* for deciding on the number of cases, i.e. the sample must contain ten cases for each IV. While simulation studies show 5 – 9 events per parameter was acceptable and contributed around 10% to the mean squared error [51]. An insufficient number of cases relative to the number of IVs could result in large coefficient estimates and high standard errors for the IVs, and very large or very small odds ratios as seen for some of the IVs reported in Appendix 10.

Around 200 surveys were distributed to the urban bus drivers and out of 150 drivers 53 returned the completed surveys resulting in a response rate of 35%. Three surveys were unfilled and were deleted from the database. The *linear predictor* in Equation 12 has ten variables (covariates) some of which are dichotomous. The categorical variables having more than two levels in the survey were collapsed to dichotomous variables during trial statistical runs. For example, marital status had: *married, never married, divorced, separated, not disclosed*. This was collapsed to: *married* and *others* (*never married, separated, divorced, not disclosed*). Hence, the resulting MLR model with all ten covariates had 17 variables (including the pairs (1, 0) for the dichotomous variables). These 17 variables required at least 170 survey cases as per the *Rule of 10* [11], but collected only 50 survey cases for the urban study thus resulting in a low case to variables ratio of 3:1 which could cause *numeric instability* [11].

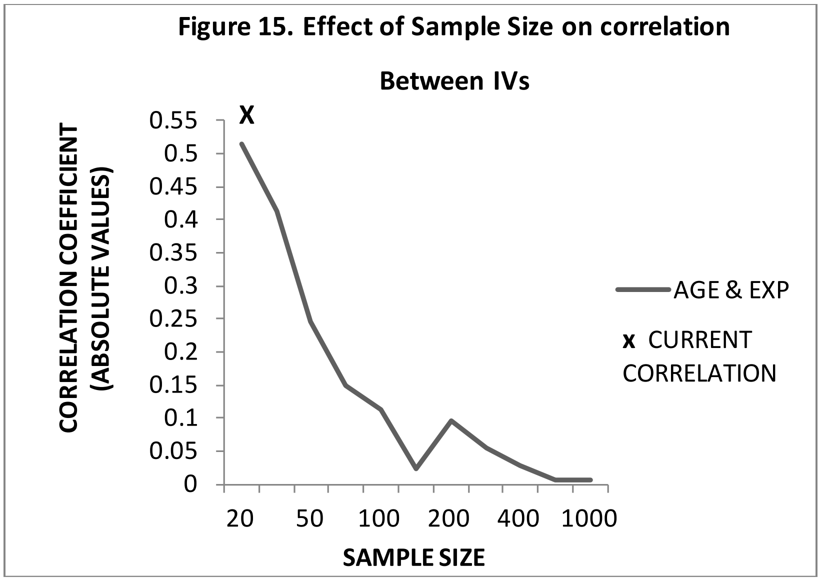
Sample Size

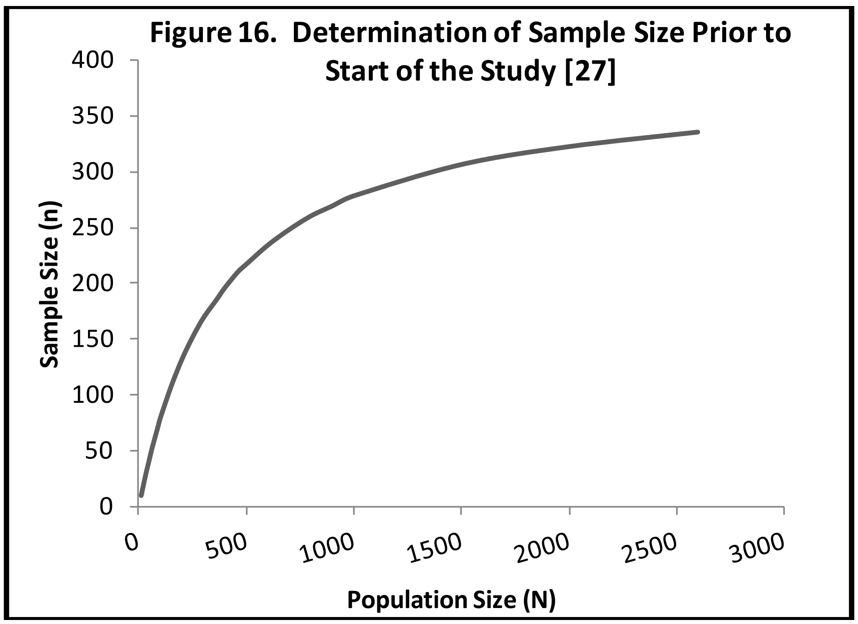
Several approaches are recommended for computing an appropriate sample size. In an earlier study [13], simulation replications were conducted for 1,000 drivers and correlation coefficient for the independent variables was computed for various sample sizes. The sample size of 48 cases returned by the drivers resulted in a correlation coefficient of 0.54 between variables Age and Experience [13]. The correlation coefficient between Age and Experience computed for varying sample sizes showed a reduction at larger sample sizes (Figure 15).

Using Figure 15 as a guideline, a sample size of 150 drivers for the urban transit agency would lower the correlation coefficient to less than 0.05 and would increase the case-to-variable ratio to 9:1 which is close to satisfying the *Rule of 10* [11]. With only 150 drivers it is not possible to reach the ratio of 10:1 unless the number of covariates is reduced in the MLR model.

The appropriate sample size could also be computed prior to the start of the study from Figure 16 created from data provided by Orcher, 2007 [27]. The number of bus drivers at the regional transit agency was around 460. Referring to Figure 16 for a population size of 460, the suggested sample size is around 220 surveys. This would reduce the correlation coefficient to less than 0.1 and have a case-to-variable ratio of 13:1.

In the absence of specific guidelines on the sample size needed for fitting logistic and multinomial logistic regression models, the number of survey cases that are being tested could be linked to the number of IVs (covariates) using the *Rule of 10* formulated by Hosmer et al. [11]. With 17 covariates in this study, a minimum of 170 cases (survey) responses are needed. This sample size would lower the correlation coefficient between IVs to less than 0.05 (Figure 15) thus reducing or eliminating multicollinearity and standard errors.





Using Stepwise Method to Select Independent Variables and Interactions

The stepwise method available with the SPSS 17.0 [40] software builds the model sequentially by inclusion or exclusion of IVs. As a start, all the independent variables are included as direct entry. A stepwise (MLR) procedure then eliminates non-significant factors until a good fit is achieved with the significant factors producing three output tables for each high risk distracting activity. According to Hosmer et al. [11] stepwise is an effective way of *screening a large number of variables* and providing a good fit, but it relies on the computer to select variables instead of the judgment of researchers who are finally responsible for the model outcomes.

Multicollinearity
The MLR model is influenced by high correlation among the IVs which could result in multicollinearity. Multicollinearity causes a high standard error and inaccurate coefficients for the IVs. A test was conducted to determine if there was any correlation among the IVs and the results are presented in Table 11. The highest correlation coefficient was 0.61 (Equipment and Location). All the coefficients were between –0.70 and 0.70 and should not cause a correlation problem according to Lind et. al. [18]. Hence, none of the IVs were excluded from the MLR model formulation

4.2.2.2. Model Results

Fitting the MLR Model
Equations 9, 10, and 11 were fitted to the survey data by SPSS 17.0 [40] to test the research hypotheses: *to determine the likelihood that the transit bus driver getting Slightly Distracted, Distracted, and Very Distracted with respect to Not Distracted is related to his/her attributes, driving pattern, service location, and type of bus driven*. The output is split into three tables since the continuous and categorical IVs are compared in pairs. The method followed by SPSS 17.0 [40] for including variables in the MLR model is direct entry of all variables. It is not necessary to create dummy variables for categorical variables LOCAT, SEX, MARITAL etc since the software does this automatically when we input these variables as "factors" at the input stage [40].

Table 11. Correlation Between the IVs

	Loc	*Age*	*Sex*	*Edu*	*Marital*	*Exp*	*Drive/Wk*	*Day*	*Peak*	*Equip*
Loc	1									
Age	0.32	1.00								
Sex	0.06	0.05	1.00							
Edu	-0.36	-0.42	0.08	1.00						
Marital	0.03	0.13	0.45	-0.08	1.00					
Exp	-0.12	0.33	-0.07	0.20	-0.01	1.00				
Drive/Wk	0.28	0.02	0.05	-0.31	-0.04	-0.19	1.00			
Day	0.06	-0.15	-0.01	0.19	0.06	-0.08	-0.05	1.00		
Peak	0.20	0.23	0.01	-0.22	-0.03	0.04	0.09	-0.62	1.00	
Equip	0.61	0.28	0.19	-0.37	0.24	-0.17	0.14	0.07	0.13	1.00

Goodness of Fit Test
As an illustration, the results for *Pedestrian* (the highest risk distracting activity) are presented in this Section. The model is evaluated for goodness-of-fit using the Step Summary where the -2 log-likelihood computes the unexplained variability in the data (Table 12). The Model 0 enters the main effects followed by Model 1 which enters the Sex*Edu interaction generating a chi-square (27.812, 3) which is highly significant (< 0.001). The Table 12 presents the model fitting criteria for the full model. The reduction of -2 log-likelihood from 70.165 to 42.314 indicates that the variance has been explained by the model

The -2 log-likelihood of the baseline model with only the intercept had a Chi-square of 122.773 (Table 13) which drops to a Chi-square of 42.354 which is a reduction of 80.420. This change is significant ($p< 0.001$). Hence, the final model explains a significant amount of the initial variability meaning that the model is a better fit than the original model [8].
.

Table 12. 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, Exp	70.165	.		
1	Entered	Sex * Edu	42.354	27.812	3	.000

Stepwise Method: Forward Entry. a. The chi-square for entry is based on the likelihood ratio test.

The Model Fitting Information such as the Pearson and Deviance statistics, and Pseudo R-Square describes how well the model fits the data (Tables 13, 14, and 15), and whether the model's predicted values differ significantly from their observed data. Since both, the Pearson and Deviance statistics are not significant ($p = 1.000$) we 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); a lack of over dispersion indicates a good fit. The over dispersion can be calculated as follows:

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

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

Both ratios are less than the ideal value 1 hence the data is not over dispersed. Furthermore, the Cox and Snell, Nagelkeke, and McFadden statistics of 0.793, 0.872, and 0.655 are reasonably high indicating a good fit (Table 15).

Table 16 shows the results of the *likelihood ratio* tests that can be used to ascertain the significance of the IVs to the model. The IVs sex and education have no significant values (sig column has blank spaces) since they are involved in higher order interactions. Significant variables ($p < 0.1$) for this study are shown in the last column.

Table 13. 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	.000

Table 14. **Goodness-of-Fit**

	Chi-Square	df	Sig.
Pearson	51.814	117	1.000
Deviance	42.354	117	1.000

Table 15. Pseudo R-Square

Cox and Snell	.793
Nagelkerke	.872
McFadden	.655

Table 16. Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	42.354[a]	.000	0	.
DriveWk	50.196	7.842	3	.049
Age	47.365	5.012	3	.171
Peak	47.018	4.664	3	.198
Marital	44.439[b]	2.085	3	.555
Sex	42.354[a]	.000	0	.
Loc	70.394	28.040	3	.000
Edu	42.354[a]	.000	0	.
Day	44.886	2.533	3	.469
Equip	68.565	26.212	3	.000
Exp	58.244	15.890	3	.001
Sex * Edu	70.165	27.812	3	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

b. Unexpected singularities in the Hessian matrix are encountered. This indicates that either some predictor variables should be excluded or some categories should be merged.

The *likelihood ratio* tests provides a general idea of which IVs (DriveWk, Loc, Equip, Exp)have a significant impact on the levels of distraction but they do not tell us what is the magnitude and direction. Appendix 9 summarizes the model fitting characteristics of the eight Very High and High Risk Zone distracting activities. Except for the non-availability (N/A) of Step Summary for four of the distracting activities, all the other tests show that the models have a good fit. The distracting activities statistical outputs for the significant predictor variables have been summarized in Appendix 10.

Classification of Dependent Variable Levels

The Classification Table 17 is a contingency table that compares the observed versus predicted responses for all combinations of the dependent variable and is an indicator of the usefulness of the model. In the case of the distracting activity *Pedestrian*, from an overall performance, the model could predict 82.4% of the responses, 84.6% of those are *Not Distracting*, 80.0% are *Slightly Distracting*, 77.8% are *Distracting*, and 100% are *Very Distracting* to the bus driver. Correctly predicted responses are shown on the diagonal of the table (11 responses in the *Not Distracting* group, 20 in the *Slightly Distracting* group, in the *Distracting* group, and 4 in the *Vey Distracting* group. A summary of correctly predicted responses for the Very High Risk and High Risk distracting activities is shown in Appendix 11. Five out of the eight highly risky distracting activities had correct overall prediction of above 70%

Table 17. Classification for Pedestrian

Observed	Predicted				
	NOT DISTRACTING	SLIGHTLY DISTRACTING	DISTRACTING	VERY DISTRACTING	Percent Correct
NOT DISTRACTING	11	2	0	0	84.6%
SLIGHTLY DISTRACTING	2	20	3	0	80.0%
DISTRACTING	0	2	7	0	77.8%
VERY DISTRACTING	0	0	0	4	100.0%
Overall Percentage	25.5%	47.1%	19.6%	7.8%	82.4%

Out of the eight MLR models for Risk Zones I and II distracting activities (Appendix 8), five were found to be highly significant and exhibited a good fit ($p \leq 0.100$). The model fitting significance level for PASSENGERS, FAREBOX, and UNRULY KIDS had a p value = 0.242, 0.255, and 0.230 respectively. In addition, the over dispersion ratios ($\Phi_{Pearson}$, $\Phi_{Deviance}$) were greater than or close to 1, and their Pseudo R-Square ratios were comparatively lower than the corresponding ratios of the significant distracting activities. It was decided to include these three distracting activities in the analysis of the significant IVs since these activities were classified as Very High Risk (Zone I) and High Risk (Zone II) distracting activities. The multinomial linear predictor function Y_{ij} for the distracting activities shown in Appendix 12 includes estimated coefficients that had significance levels of ≤ 0.10. Since the Wald statistic is quite conservative, Tabachnick and Fidell [42] have suggested a higher level of significance of $p < 0.05$ or $p < 0.1$ may be used during interpretation of the variables.

Interpretation of MLR Constant Terms and Continuous/Categorical Variables

The MLR models for the eight distracting activities in Risk Zones I and II were fitted and a summary of the significant ($p \leq 0.10$) IVs and coefficient estimate (B), standard error, Wald Statistic, and OR along with 95% confidence Intervals for each urban transit distracting activity is provided in Appendix 10. The significance of each variable is tested by the Wald Statistic and the corresponding significance (p) value. The multinomial coefficient estimates were interpreted using their magnitude and direction [8, 24, 29, 53] together with the OR guidelines [10, 30].

Some of the MLR linear predictors do not have a constant because the estimated constants do not vary across the levels of distraction and can be considered as a zero (0) baseline [53]. Each coefficient of the linear predictor in Appendix 12 is interpreted on the basis of its magnitude and sign. In the case of Pedestrian distraction in Appendix 12, keeping everything else fixed, a driver is less likely to get Distracted (with its negative constant = -13.47) relative to Slightly Distracted. 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. The positive coefficient of Drive Hrs/Week (0.178) indicates that keeping everything else fixed, additional driving hours per week increases the likelihood of a driver getting Distracted by Pedestrians. Older drivers are more likely (positive coefficient = 0.264) to get Distracted by Pedestrians. The negative coefficient (-5.937) on the Peak variable (Peak = 1, Non-Peak = 0) indicates that drivers who drive during peak hours are less likely to get Distracted by Pedestrians.

The "Exp (B)" column which contains the odds ratios (ORs) for each estimated coefficients is used to interpret the independent variables for the eight distracting activities. The ORs are used to compare the relative odds of the occurrence of a type of driver distraction (for example Pedestrian) for a given dichotomous predictor variable (for example gender), and to compare the magnitude (for example, the odds of a male driver getting distracted by Pedestrians is 1.5 times that of a female driver). The OR greater than 1.0 where the 95% CI does not include 1.0 indicates the odds of the outcome (distraction) are greater in the first group (male drivers) than the second group (female drivers). And ORs less than 1.0 where the 95% CI do not include 1.0 indicate the odds of the outcome (distraction) are smaller in the first group (male drivers) than the second group (female drivers) [10]. According to McHugh [20], an OR less than 1.0 *is not directly interpretable* since the OR does not provide the extent to which the first group is *less likely* to experience the activity. It is suggested to reverse the OR such that the first group becomes the second group and the second group becomes the first group [20]. If the OR = 1 or the 95% CI include 1 (overlaps the null value), then it indicates that both groups are equally likely to experience the event (get distracted) [20]. It would be inappropriate to interpret an OR = 1 or having the 95% CI overlapping the null value [20].

Some of the coefficient estimates had extremely large or small ORs. According to Hickman et al. [10], an OR is a measure of association (not unlike a correlation), which can be used under the correct circumstances as an estimate of the rate ratio and hence, it is difficult to report the OR in any meaningful sense other than to report there was a very strong or weak relationship between the variable and the distracting outcome. In this research, ORs are interpreted according to the approach of by McHugh [20] discussed earlier. Furthermore, the ORs were used as *broad stroke estimate of effect* [30] instead of an accurate estimate of the likelihood of a driver getting distracted. Hence, the phrases such as *more* or *less* likely are acceptable according to Petrucci [30].

The Table 18 summarizes in a descriptive form the impact of all the significant variables on the distracting activities classified in Risk Zones I and II. The coefficients of these significant variables along with standard errors, wald statistics, and ORs are listed in Appendix 10. The Table 19 presents sample outputs from the three binary logistic regression models for “Pedestrians“.

Table 19. MLR Model Outputs for Pedestrians.

Model Chi-Square (χ^2) = 80.420 (33)**** Pearson Stat (NS) Deviance Stat(NS)	R^2 = 0.79 (Cox & Snell); 0.87 (Nagelkerke); 0.66 (McFadden)	AIC initial/final values: N/A BIC initial/final values: N/A		
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 HRS/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; ****p < 0.001. N/S = Not Significant.

4.3. Validation of Results

The MLR models developed for Risk Zones I and II distracting activities have been statistically assessed in Section 4.2.2.2 using Goodness of Fit tests. The assessment tests presented for Pedestrian showed that the MLR model developed from a localized sample of transit bus drivers exhibited a good fit and the independent variables were contributing significantly to the outcome of the distraction level. These mathematical models are estimates of the current levels of distraction at the agency. The question is: Are the results generated by the MLR model linear predictors for Risk Zones I and II distracting activities in Appendix 12 also valid for a large random population of transit bus drivers? Three methods of validating the results are presented in the following sections: Expert Verification, Simulation, and Route Observations.

Table 18: Impact of Independent Variable Coefficients on Risk Zone I Activities.

	DISTRACTING ACTIVITIES			
INDEPENDENT VARIABLES COEFFICIENT	**PEDESTRIAN**	**PASSENGERS**	**OTHER ROAD USERS**	**UNRULY KIDS**
Age	Older drivers are more likely to get Distracted			
Experience	Drivers with more number of driving experience years are less likely to get Slightly Distracted		Drivers with more number of driving experience years are less likely to get Slightly Distracted followed by Distracted	
Drive Hrs/Week	Drivers with more number of driving hours per week are more likely to get Distracted			
Gender	N/S			
Marital Status	N/S		Married Drivers are more likely to get Slightly Distracted followed by Distracted	
Educational Level	N/S			
Day Driving	N/S	Night shift Drivers are more likely to get Distracted		Night shift Drivers are more likely to get Slightly Distracted
Peak Driving	Peak drivers are less likely to get Distracted	Peak drivers are less likely to get Distracted followed by Slightly Distracted		
Type of Equipment	N/S	MCI Drivers are more likely to get Distracted		

Table 18 (Continued). Impact of Independent Variable Coefficients on Risk Zone II Activities.

	DISTRACTING ACTIVITIES			
INDEPENDENT VARIABLES COEFFICIENT	**PASS MOBILE PHONE**	**MOBILE DATA TERMINAL**	**PASS NOT ETIQUETTE**	**FARE BOX**
Age			Older Drivers are less likely to get Very Distracted due to Age	
Experience			Drivers with more number of driving experience years are less likely to get Distracted followed by Slightly Distracted	
Drive Hrs/Week				
Gender		Male Drivers are more likely to get Very Distracted		Male Drivers are more likely to get Slightly Distracted
Marital Status			Married Drivers are more likely to get Very Distracted	
Educational Level			HS level Drivers are more likely to get Very Distracted	HS level Drivers are more likely to get Slightly Distracted
Day Driving				Day shift Drivers are less likely to get Slightly Distracted
Peak Driving	Peak drivers are less likely to get Slightly Distracted followed by Distracted	Peak drivers are less likely to get Very Distracted	Peak drivers are less likely to get Distracted	
Type of Equipment		MCI Drivers are more likely to get Very Distracted	MCI Drivers are more likely to get Slightly Distracted	

4.3.1 Expert Verification

Expert verification by safety managers in the participating agencies is the starting point for validation. Standardized Expert Verification forms as per the sample shown in Table 20 will be needed for general verification of results.

Table 20. Expert Verification Form for Validating MLR model's Results.

DISTRACTING ACTIVITY	MLR MODEL RESULTS	EXPERT FEEDBACK
PEDESTRIAN	This distraction was impacted by Age, Experience, Driving Hours/Week, and Peak Driving	
PASSENGERS	This distraction was impacted by Driving Shift, Peak Driving, and Type of Equipment.	
OTHER ROAD USERS	This distraction was impacted by Driving Experience and Marital Status.	
UNRULY KIDS	This distraction was impacted by Driving Shift.	
PASS MOBILE PHONE	This distraction was impacted by Peak Driving.	
MOB DATA TERMINAL	This distraction was impacted by Gender, Peak Driving, and Type of Equipment.	
PASS NOT ETIQUET	This distraction was impacted by Age, Driving Experience, Marital Status, Education Level, Peak Driving, and Type of Equipment.	
FARE BOX	This distraction was impacted by Gender, Educational Level, and Driving Shift.	

4.3.2. Simulation

Computer simulation is commonly used by transportation researchers to validate output results from a model. A simulation tool was developed with replicating features of urban traffic flow and was used to validate and calibrate an urban traffic modeling tool [36, 37, 38]. Secondary driving tasks play a major role in driving performance and its impact on distraction has been studied extensively using models and simulation [39]. The MLR linear predictors (Appendix 12) 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 Equation (6) for a range of 100 drivers getting *Slightly Distracted*, *Distracted*, and *Very Distracted*. The $\pi(x)$ values were plotted graphically and then compared to the results from the estimated coefficients of MLR linear predictors. The results for a sample of distracting activities and factors are illustrated in the following Section.

4.3.2.1. MLR Model and Simulation Outputs

Monte Carlo simulation used discrete and continuous probability distributions that incorporated random variability into the model to validate a model's output results. The MLR models' *linear predictors* for Risk Zones I and II distracting activities presented in Appendix 12 were repeatedly simulated by a different random set of values (inputs) drawn from the probability distribution of the predictor/independent variables (IVs) producing a set of probability values for each distraction outcomes (outputs).

The Simulation Model

The *linear predictors* for all MLR models listed in Appendix 12 were substituted into Equation 6 developed in Section 4.2.2.1. The probability of the extent that a driver finds a source distracting is computed from the logistic regression's probability function π (x) via Equation 6. The probability values from the function π(x) will vary between 0 and 1. The event x is very unlikely to occur if π(x) is close to 0 and very likely to occur if it is close to 1.

Simulation was replicated for 100 (bus drivers) by varying each predictor variable in the MLR *linear predictor* one at a time while keeping the rest fixed. The discrete probability distribution was applied to generate a set of random categorical variables for 100 bus drivers, while the normal and uniform probability distributions were applied to generate a set of random continuous variables. The probability function π(x) computed from Equation 6 was plotted for each variable (x). The graphical outputs from the simulation models for the distracting activity "Pedestrian" are presented as an illustration.

1. <u>Reported Age of Driver</u>. Age is an important factor related to accidents with younger drivers more prone to distracted driving and accidents. Earlier studies conducted on personal vehicle concluded that driver age had a significant impact on distraction, with younger and older drivers more prone to distraction [46]. The average age of an urban driver is 47 years and the MLR model reveals positive impact of age on distraction. The coefficient of the variable Age for external distracting activities such as Pedestrians (Y_{12}) is a positive value (0.264) indicating that as age of the driver increases, they are more likely to get Distracted by the Pedestrians. Figures 17 from the simulation output confirms the MLR results for Pedestrians. Older drivers get more distracted by external activities such as Pedestrians compared to younger drivers although an earlier study by Tefft [43] found the accident rates were higher for younger drivers and personal vehicle drivers [46].

2. <u>Number of Years of Experience Driving a Bus</u>. The urban transit bus drivers had an average of eight years experience driving buses. It is believed that experienced drivers would get less distracted by the distracting activities due to the fact that they would be older and have been driving long enough to be affected by distracting activities. Studies on the impact of age on driving performance suggests that younger (below 25 years) tend to be more vulnerable to the effects of distraction than middle-aged drivers [46]. The coefficients of the Experience variable are consistently negative (Y_{11}, Y_{31}, Y_{32}, Y_{71}, Y_{72}) indicating that added experience decreases the likelihood of getting Slightly Distracted followed by Distracted due to Pedestrians, Other Road Users, Mobile Data Terminals, and Passengers Not Following Etiquette. This matches popular belief, where experience made a driver better at handling distraction and possibly has less accidents as reported in the preliminary data analysis (Figure 10). Simulation output in Figure 18 validates the MLR model results.

3. <u>Driving Time</u>. PEAK = 1, NON-PEAK = 0. The negative coefficients (-5.937) associated with PEAK = 0 (Y_{12}) implies that 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 19 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 than Peak drivers (0.557).

4. Driving Hours/Week. The positive coefficient (0.178) associated with DRIVING/WK (Y_{12}) implies that holding all other IVs fixed, the higher the driving hours/week, the more likely the a driver would get Distracted by Pedestrians. The OR is more than 1 and the 95% CI does not include 1 for Distracted. If a driver increases her/his driving hours/week by one hour, the odds of getting Distracted to Not Distracted would increase by 1.194 times given the other independent variables are held fixed. Simulation results shown in Figure 20 validate this result.

Figure 17. Probability Function for Pedestrians (D)

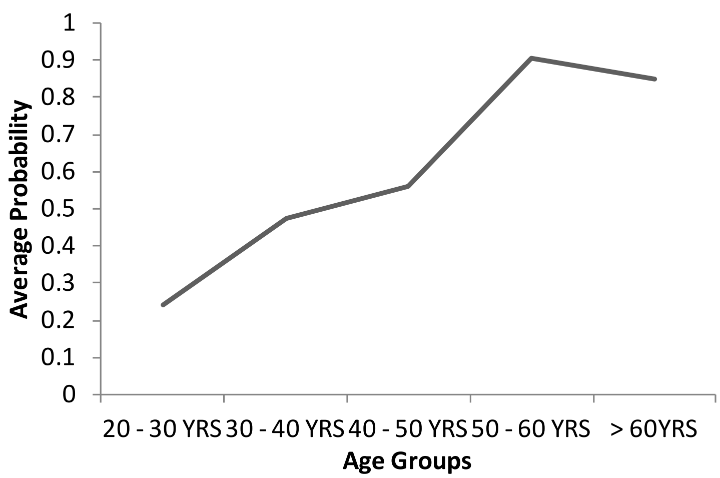


Figure 18. PROBABILITY FUNCTION FOR PEDESTRIAN (SD)

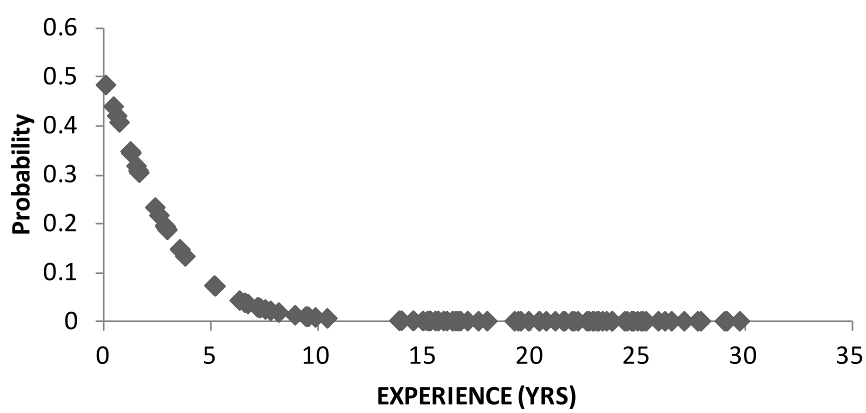


Figure 19. PROBABILITY FUNCTION FOR PEDESTRIAN (D)

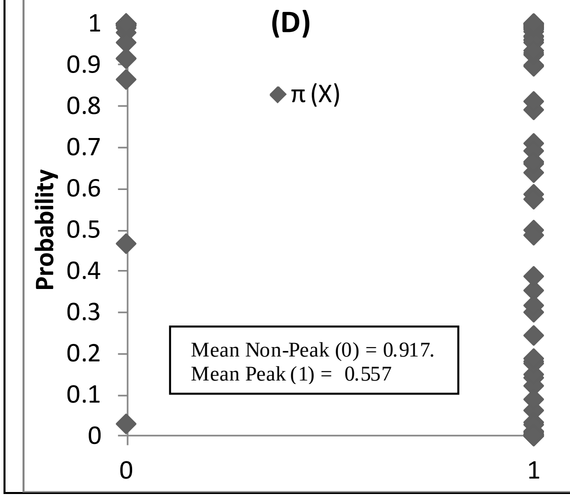
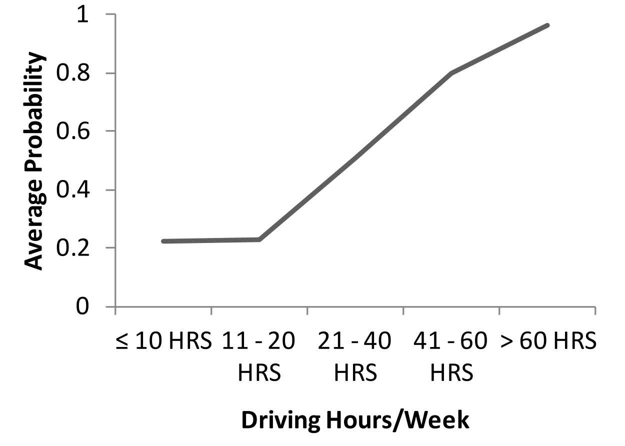


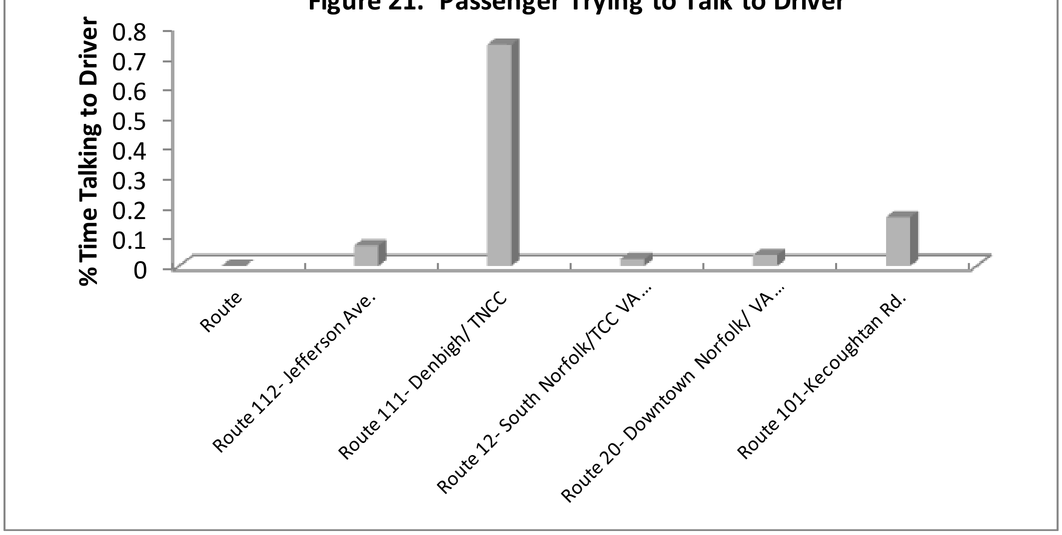
Figure 20. PROBABILITY FUNCTION FOR PEDESTRIAN (D)



4.3.3. Route Observations

Route observation is very useful for a quick distraction study as well as for validation of statistical models. A standardized form illustrated in Appendix 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. But this type of distraction is commonly observed in some route such as Route 111 (Figure 21). 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.

Figure 21. Passenger Trying to Talk to Driver



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 require some special training to record distraction. Also, observers maybe 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.

4.4. Guidelines for Results Interpretation and Usage

In the final module, guidelines are created for the agencies to interpret the results and apply them to predict driver distraction, develop policies, determine training needs, redesign of driver cabin, adopt new technology, etc.

The interpretation of descriptive statistics related to the urban drivers' attributes is summarized in Table 3. From the survey data, it appears that generally, male drivers (75%) outnumber female drivers (25%) at the urban agency. At the regional agency, the number of female bus drivers (54%) was slightly higher than the male drivers (46%) [13]. The average age of an urban bus driver is 47 years with an average of 8 years of driving experience. Most of the regional and urban bus drivers fall into the 36-55 age group which has a lower risk of accidents. Around 65% of the urban drivers were married and 34% had either a two-year or four-year college degree. A few (2%) had less than a high school diploma. The fact that large proportion of drivers are married and hold either a HS diploma (or some college) are useful attributes in accident and distraction control [16]. Drivers have reported an average of 37.14 driving hours per week which come close to the normal 40 hours work week. Hence, distraction due to fatigue/sickness which is often linked to overwork in other studies has been classified as a moderate risk activity under Risk Zone III.

A study found that transit bus drivers involved in one or more collisions are 2.0 times more likely to be regularly distracted by a handheld cell phone [21]. Surprisingly, drivers ranked the rating for *Driver Mobile Phone Usage* 3rd (Appendix 3) and duration of distraction 17th (Appendix 4). Both agencies have banned the use of personal cell phone while driving. However, drivers communicate with the operations center through Citizen Band (CB) radio, The survey participants have possibly confused the CB with cell phone and that explains why driver distraction due to *Driver Mobile Phone usage* is so high. In this study, the top five distracting activities reported by the agency's bus drivers were mostly passenger-related (Tables 6, 7, and 8). These closely resemble the top five distracting activities rated by local or international bus drivers (Table 21).

Table 21. Comparison of Top Five Distracting Rating Activities at Transit Agencies.

Highest Rated Distracting Activities	PRTC Ranking	HRT Ranking	STA NSW Ranking 35]
Unruly Kids/School Children	1	2	5
Passenger Using Mobile Phone	2	1	-
Driver's Mobile Phone	3	-	-
Passengers Not Following Etiquette (eating, drinking, smoking, noisy) or Noisy Passengers	4	3	4
Passengers	5	5	-
Passengers Trying to Talk to Driver	-	4	-
Sickness	-	-	2
Fatigue	-	-	3
Unruly Passengers	-	-	1
AVERAGE RATING	**2.05**	**2.17**	**2.16**

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 [5, 6, 7] 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. The transit driver attempts to distribute his or her attention to both the secondary driving tasks as well as the primary tasks associated with operating the vehicle thereby increasing cognitive distraction. Mental inattention [25] begins to take place, particularly when additional secondary driving tasks are factored in such as attending to unruly kids. This mental inattention increases the amount of time that it takes for the driver to fully process information and to formulate and act upon the decisions made, based on such information [25]. A threshold is reached, particularly as additional tasks are added which increases mental inattention due to being overtaxed by a heavy mental workload. At this point, it becomes impossible to multitask and mental inattention towards the primary driving tasks produces a crash risk [25].

4.4.1. Key Model Results

According to the literature review, distraction is one of the major causes of accidents [10, 34, 48, 49]. Distraction occurs when a driver's attention is diverted away from driving by a secondary task that occurs approximately 30% of the vehicle movement time [33]. Preliminary analysis of the 10 driver attributes indicate the possibility of some/all of them being causes of distraction.

Mobile phone usage in public transit systems is an annoyance and distraction to other passengers and the driver. To avoid such situations, a growing number of cities and states have banned the use of personal mobile phones by drivers and passengers in the transit system. It is a challenge for the transit agency to develop effective policies for handling passenger behavior so that they are less likely to stand next to the driver's cab, talk to the driver, engage in using cell phones, non-etiquette and noisy conversation etc. Providing route maps and other pertinent information in the bus and at the stops would reduce talk between passenger and driver.

Personal use of electronic devices by passengers may be permitted beyond the middle section of the bus to avoid distracting the driver. The front section of the bus could be designated as *cell phone free* not enforceable through legislation but by posting friendly sign boards. Drivers must not permit any passengers to stand next to the driver's cab. In order to control unnecessary communications between driver and passenger, appropriate sign boards could be posted on the side of the driver's cabin [6]. If conversation cannot be avoided, it must be done cautiously while driving or when the bus is stopped.

The design of fare boxes, control panels, and other devices must be user-friendly, and not require long glances away from the roadway. 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 drivers who are likely to be distracted by these devices. MLR models and simulation results support the hypotheses stated in Section 4.2.2 and Table 18 summarizes the impact of each significant variable on the level of Zone I and II distracting activities.

4.4.2. Interpretation of Key Results

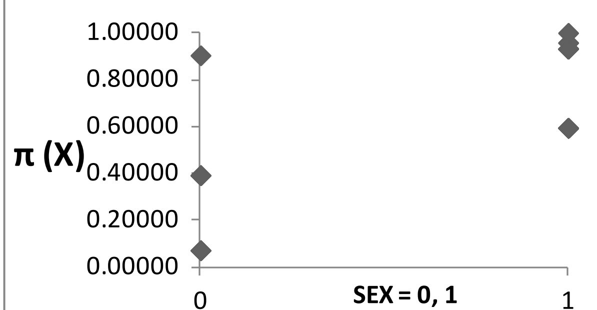
The interpretation for each significant predictor variable follows from the magnitude and direction (sign) of the estimated coefficients of each MLR linear predictor in Appendix 12. The interpretation of ORs (Appendix 10) follow the approach of McHugh [20] and are used as a *broad stroke estimate* [30] 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 [29, 53].

Gender of driver

There is no conclusive evidence from the literature review showing if male or female drivers are more or less prone to distraction related accidents. The positive coefficients associated with SEX (Male = 1, Female = 0) in Equations Y_{63} (3.57) and Y_{81} (3.00) implies that when gender changes from 0 to 1, the probability of getting distracted will increase. Therefore, male drivers are more likely than female drivers to get Very Distracted by Mobile Data Terminals (MDTs) and Slightly Distracted by Fare Boxes. The ORs are > 1 and the 95% CI does not includes 1 for the Fare Box (Appendix 10) indicates the odds of male drivers getting Very Distracted by MDT is greater than female drivers. Figures 22 and 23 shows the simulated results for the impact of Gender on MDT and Fare Box. For males, the probability values are 1 for MDT and between 0.6 and 1 for Fare Box. For females, it is between 0.7 and 1 for MDT and between 0.05 to 0.9 for Fare Box. Hence, the simulated results match the MLR model results.

Figure 22. PROBABILITY FUNCTION FOR MDT (VD)

Figure 23. PROBABILITY FUNCTION FOR FARE BOX (SD)



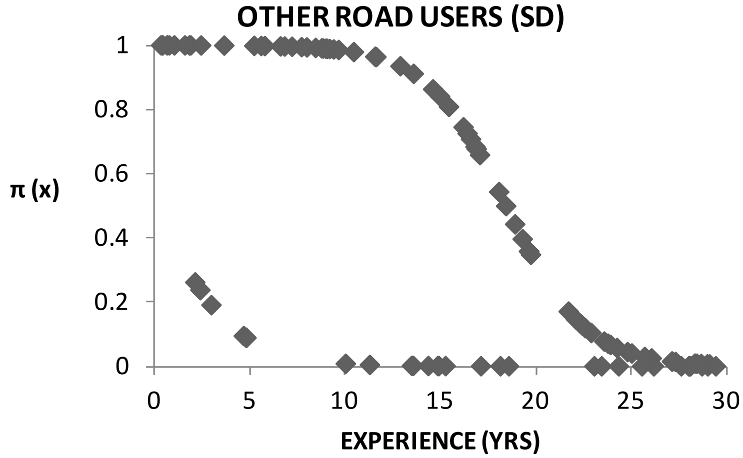
Age of driver

According to USDOT [48], 40 percent of the drivers who died in crashes were in one of these two age groups: 23 percent in the 16 - 24-year age group and 17 percent were in the 65 and older age group. The traffic crash fatality rate per 100,000 populations is the highest in the 16 -24 year age group, followed by those over age over 74 [19]. According to the USDOT [48], 28% of drivers involved in fatal crashes were under 30 years while only 10% of drivers in the

40 - 49 year age groups were involved in fatal crashes. While a study of truck-involved in rear-end crashes found that the incidents of such accidents increases with age groups. Younger car or truck drivers (< 25 years) are less likely to get involved in a truck crash compared to middle age (26-55 years), and older drivers (> 56 years), who are more likely to be involved in such crashes [56]. The impact of age has been discussed in details for *Pedestrians* in Section 4.3.2.1.

Number of years of experience driving a bus.
The coefficients of the Experience variable are consistently negative (Y_{11}, Y_{31}, Y_{32}, Y_{71}, Y_{72}) and the ORs are less than one and the 95% CI does not include one, indicating that added experience decreases the likelihood of getting Slightly Distracted followed by Distracted due to Pedestrians, Other Road Users, and Passengers Not Following Etiquette. This matches popular belief, where experience made a driver better at handling distraction and possibly have less accidents as reported in the preliminary data analysis (Figure 10). Simulation output in Figure 24 validate the MLR model results.

Figure 24. PROBABILITY FUNCTION FOR OTHER ROAD USERS (SD)



Marital Status
The positive coefficients associated with MARITAL (Y_{31}, Y_{32}, Y_{73}) implies that when marital status (Married = 1, Others = 0 (Separated, Divorced, Never Married, etc)) changes from 0 to 1, the probability of getting distracted increases. Married drivers are more likely to get Slightly Distracted to Distracted by Other Road Users and Very Distracted by Passengers Not Following Etiquette. Simulation outputs shown in Figures 25 and 26 validate the MLR model results.

Educational Level
The positive coefficients associated with EDU (Y_{73}, Y_{81}) implies that when educational level (HS or Equivalent = 1, Others = 0 (Some College, 2,4 year degree, etc.)) changes from 0 to 1, the probability of getting distracted increases. Therefore, drivers with HS or equivalent were more likely than college educated drivers to get Very Distracted by Passengers Not Following Etiquette and Slightly Distracted by Fare Boxes. The ORs are > 1 and the 95% CI does not include 1 for the Passengers Not Following Etiquette. Therefore, the odds of HS or equivalent drivers getting Very Distracted by Passengers Not Following Etiquette are more

likely as compared to Other drivers. Figures 27 and 28 shows the simulated results for the impact of Education on Passengers Not Following Etiquette and Fare Box.

Figure 25. PROBABILITY FUNCTION FOR OTHER ROAD USERS

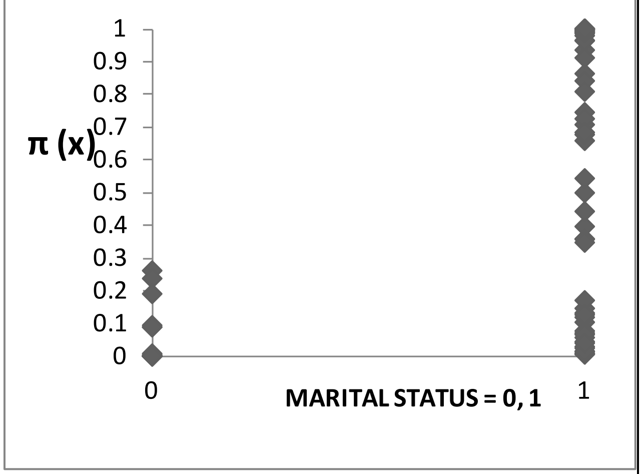


Figure 26. PROBABILITY FUNCTION FOR PASSENGER NOT FOLLOWING ETIQUETTE (VD)

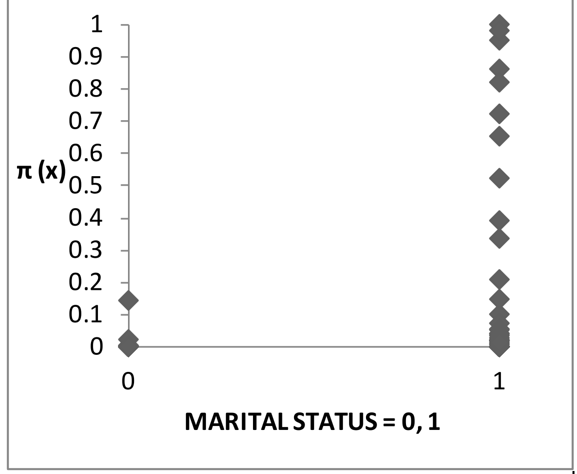


Figure 27. PROBABILITY FUNCTION FOR PASSENGER NOT FOLLOWING ETIQUETTE (VD)

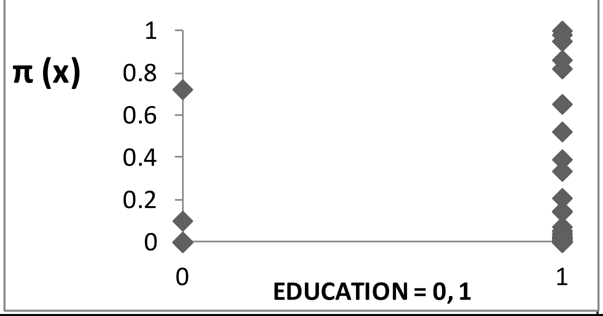
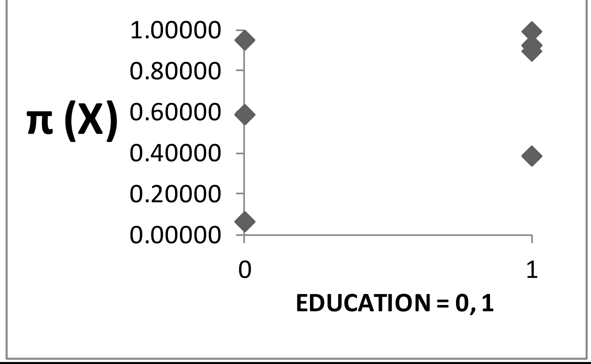


Figure 28 PROBABILITY FUNCTION FOR FARE BOX



For HS level drivers (EDU = 1), the maximum probability value is 1 for both activities and a maximum of 0.75 for drivers with education beyond HS or some college education, and a probability range of 0.6 and 1 for Fare Box. Hence, the simulated results match the MLR model results.

Driving Hours/Week

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. It is difficult to measure fatigue which is often due to traffic conditions, personal life style, and health of the driver. Powell et al. [32] reported that around 19% of drivers have at least one sleep disorder. Hence, driver fatigue may be attributed to certain sleep disorders (sleep apnea, insomnia, and narcolepsy) or just lack of sufficient rest. USDOT [48], reports that driver fatigue and drowsiness may have been a factor in 56,000 crashes annually, resulting in 1,550 fatalities and 40,000 injuries a year. The urban drivers reported that they drive a bus for an average of 37.14 hours 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 hours per week; hence fatigue may not be a significant cause of distraction but requires further analysis.

The positive coefficients associated with DRIVING/WK (Y_{12}) implies that the higher the driving hours/week, the more likely the a driver would get Distracted by pedestrians (Figure 20). The OR is > 1 and the 95% CI does not include 1 for Distracted. If a driver increases her/his driving hours/week by one hour, the odds of getting Slightly Distracted to Not Distracted would increase by 1.194 times given the other independent variables are held fixed.

Location of service area

The positive coefficient associated with LOC (Y_{41}) implies that when driving location changes (Commuter = 1, Others = 0 (Local, Metro etc)) from 0 to 1, the probability of commuter drivers getting distracted increases. Therefore, the commuter drivers are more likely than their other counterparts to get Slightly Distracted by Unruly Kids. The commuter routes had a larger number of young passengers than the local and metro routes hence, could possibly lead to more passenger distraction. The simulation output shown in Figure 29 shows the opposite i.e. Commuter drivers have a lower likelihood of distraction compared to other drivers. This may be true because the commuter routes are separated by distinct lanes and there are no signals on the interstate highways.

Driving Schedule (Day/Night)

The negative coefficients associated with DAY = 0 (Y_{22}, Y_{41}, Y_{81}) implies that when driving schedule (Day = 1, Night = 0) changes from 0 to 1, the probability of Day drivers getting distracted decreases. Therefore, night shift drivers were more likely than day shift drivers to get Distracted by Passengers and Slightly Distracted by Unruly Kids and Fare Boxes. Figures 30 and 31 shows the simulated results for the impact of DAY on Unruly Kids and Fare Boxes. In both cases, the probability getting distracted is higher for night shift drivers.

Driving Schedule (Peak/Non-Peak)

Peak = 1, Non-Peak = 0. The negative coefficients associated with PEAK = 0 (Y_{12}, Y_{21}, Y_{22}, Y_{51}, Y_{52}, Y_{63}, Y_{72} implies that that when driving time changes from 0 to 1, the probability of getting distracted decreases. Therefore, non-peak drivers were more likely than peak drivers to get Distracted by Pedestrians, Slightly Distracted to Distracted by Passengers and Passengers Using Mobile Phones, Very Distracted by MDT, and Distracted by Passengers Not Following Etiquette. Figure 32 shows the simulated results for the impact of Peak on MDT. The probability getting distracted is higher for non-peak drivers.

Type of Equipment Driven

The age of the transit bus plays a major role in safety. Older buses develop maintenance problems and on-board rattles that can be highly distracting. In 2010, the average age of a full size transit bus in the U. S. was 7.8 years [41]. The average age of a bus is 9.5 years and

43% of the buses are more than ten years old which is above the national age (refer to Table 3). An exhaustive mid-life maintenance is done on the urban buses which restores the older buses to an average age of 4.1 years.

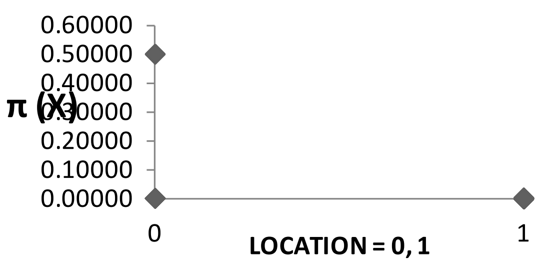


Figure 29. PROBABILITY FUNCTION FOR UNRULY KIDS

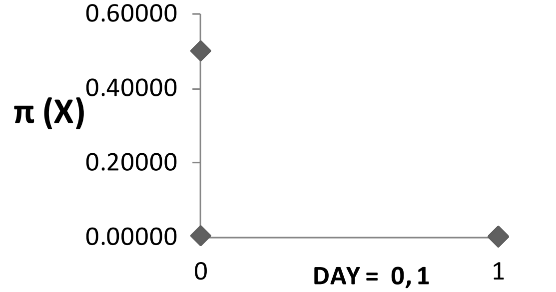


Figure 30. PROBABILITY PLOT OF UNRULY KIDS

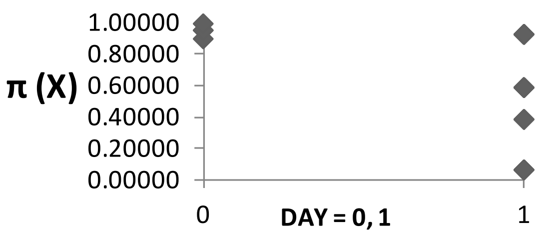


Figure 31. PROBABILITY FUNCTION FOR FARE BOX

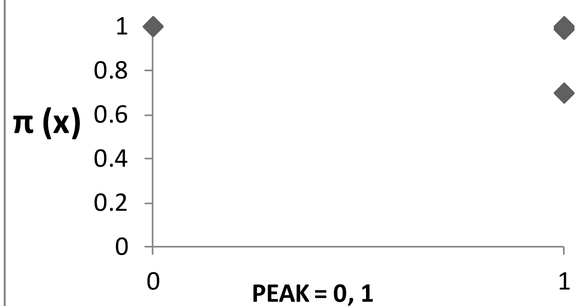


Figure 32. PROBABIITY FUNCTION FOR MDT (VD)

The positive coefficient associated with EQUIP = 0 (Y_{22}, Y_{63}, Y_{71}) implies that when type of equipment (MCI = 1, Others = 0 (Gillig, Orion etc.)) driven changes from 0 to 1, the probability of getting distracted increases. Therefore, drivers operating the MCI equipment are more likely than their other counterparts to get Distracted by Passengers, Very Distracted by MDT, and Slightly Distracted by Passengers Not Following Etiquette. Figures 33 and 34 shows the simulated results for the impact of Equipment on MDT and Passenger Not Following Etiquette. For the MCI, the probability values are 1 for both distracting activities and lower (between 0.7 and 1) for Other equipment. Hence, the simulated results match the MLR model results.

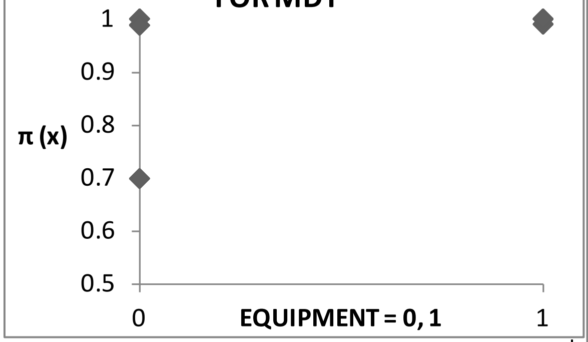


Figure 33. PROBABILITY PLOT FOR MDT

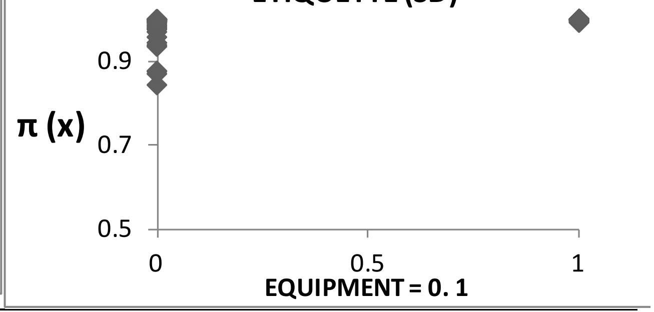


Figure 34. PROBABILITY FUNCTION FOR PASSENGER NOT FOLLOWING ETIQUETTE (SD)

4.4.3. Applications of the Model Results

The results of the MLR models, simulation, and route observations highlight the factors that significantly impact the distracting activities classified under Zone I and II that are most distracting to the driver. Taking the necessary steps to mitigate the impact would improve safety and driver performance. It is therefore a challenge for the cities to develop effective policies and driver training for handling the distracting activities. Training should focus on drivers who are more likely to get distracted by the activities in Zones I and II. 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 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 transit agency utilize the MLR models developed in this study? They could be applied to predict distraction for varying driver attributes, driving patterns, service location, and type of bus driven. Furthermore, other transit agencies could use this study as a framework for conducting similar distraction analysis of their drivers [5, 6, 7]. It is observed that drivers are affected differently by distracting activities which could be possibly corrected through proper training. Transit agencies could develop driver-based 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 high risk activities. If the probability is high, the new driver could be scheduled for related training. For example, Table 22 contains the driver attributes, location, and driving pattern for a sample drivers of 21 bus drivers.

Using the MLR linear predictors from Appendix 12, the probability of getting distracted was computed using Equation 6 and presented in Table 23. The agency may specify a cut-off probability of 0.75. Hence, drivers scoring a probability of greater than or equal 0.75 will be scheduled for the appropriate training (Table 24). Based on the overall average probability values computed in Table 23, the agency needs frequent training in dealing with *Pedestrians*, *Pass Using Mob Phone*, and *Other Road Users* [24].

Table 22. Sample Driver Attributes and Driving Pattern

DRIVER	Loc	Age	Sex	Edu	Marital	Exp	Drive/Wk	Day	Peak	Equip
P01	1	59	1	0	0	7	45	0	1	1
P02	1	47	1	1	0	5	44	1	0	0
P03	1	42	1	0	1	6	44	0	1	1
P05	1	51	1	0	0	9	40	1	0	0
P07	0	37	1	1	1	1	37	0	0	0
P09	0	39	1	1	1	15	40	1	0	0
P11	0	46	1	1	0	8	43	1	0	0
P13	1	56	1	0	1	20	40	0	1	1
P15	0	43	1	0	1	8	40	0	1	0
P16	0	47	1	1	1	7	8	1	0	0
P17	1	54	1	0	1	3	50	1	1	1
P18	1	65	1	0	1	7	45	1	0	1
P19	1	38	1	1	1	1	40	1	0	1
P20	1	55	1	1	0	30	40	1	0	1
P21	0	62	1	0	1	12	37	1	0	1

Table 23. Probability of Getting *Slightly Distracted (SD)* and *Distracted (D)*.

DRIVER	PEDESTRIAN (SD)	PEDESTRIAN (D)	PASSENGERS (SD)	PASSENGERS (D)	OTHER ROAD USERS (SD)	OTHER ROAD USERS (D)	UNRULY KIDS (SD)	PASS MOB PHONE (SD)
P01	0.032015	0.423359	0.062973	0.139434	0.995954	0.998517	0.996495	1.000000
P02	0.080542	0.999601	0.500000	0.006306	0.082413	0.093638	0.352059	1.000000
P03	0.051077	0.141001	0.062973	0.139434	0.997498	0.999057	0.500000	1.000000
P05	0.012334	0.981035	0.500000	0.006306	0.989460	0.996330	0.001908	1.000000
P07	0.380601	0.995318	0.500000	0.500000	0.381780	0.388410	0.500000	1.000000
P09	0.000672	0.999783	0.500000	0.006306	0.838891	0.946849	0.352059	1.000000
P11	0.025271	0.996073	0.500000	0.006306	0.994857	0.998139	0.001908	1.000000
P13	0.000059	0.552496	0.062973	0.139434	0.318646	0.647941	0.500000	1.000000
P15	0.019918	0.877504	0.062973	0.003834	0.993465	0.997666	0.996495	1.000000
P16	0.032015	0.994024	0.500000	0.006306	0.995954	0.998517	0.352059	1.000000
P17	0.188314	0.383433	0.062973	0.001027	0.999410	0.999758	0.352059	1.000000
P18	0.032015	0.999873	0.500000	0.210818	0.033118	0.040002	0.352059	1.000000
P19	0.380601	0.999955	0.500000	0.210818	0.999775	0.999903	0.001908	1.000000
P20	0.000000	0.994862	0.500000	0.210818	0.003758	0.019265	0.001908	1.000000
P21	0.002699	0.991481	0.500000	0.210818	0.953846	0.984909	0.352059	1.000000

Table 24. Guide for Scheduling Driver Training

DRIVER	PEDESTR (SD)	PEDESTR (D)	PASSENG (SD)	PASSEN (D)G	OTHER RD USERS (SD)	OTHER RD USERS (D)	UNRULY KIDS (SD)	PASS MOB PHONE (SD)
P01	NO TRG	NO TRG	NO TRG	NO TRG	TRG	TRG	TRG	TRG
P02	NO TRG	TRG	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	TRG
P03	NO TRG	NO TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P05	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P07	NO TRG	TRG	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	TRG
P09	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P11	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P13	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	TRG
P15	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	TRG	TRG
P16	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P17	NO TRG	NO TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P18	NO TRG	TRG	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	TRG
P19	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG
P20	NO TRG	TRG	NO TRG	NO TRG	NO TRG	NO TRG	NO TRG	TRG
P21	NO TRG	TRG	NO TRG	NO TRG	TRG	TRG	NO TRG	TRG

5. Conclusions

This project attempted to combine independent procedures for studying transit bus driver distraction into a modular research framework. It is one of only a few studies to consolidate 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.

As additional information is available from studies conducted at other agencies, the framework can be updated accordingly. The expanded data set can be used for validation as well as further refinement of the proposed framework. The 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.

5.1. Limitations

This study was conducted on a sample of regional and urban transit bus drivers in the Commonwealth of Virginia (Figures 1 and 2). The framework outlined in Figure 3 was developed from these studies and could be used by any transit agency in the Commonwealth of Virginia. In order to be applicable in other U. S. states, it may be necessary to conduct a few studies in the northern, southern, and western states.

The models suffer from important limitations and cannot provide a conclusive answer as to which distraction activity poses the greatest risks of crash involvement. One of the reasons for this is that the actual study sample size (number of cases) was grossly below the number required in Table 25 to avoid *numeric instability* [11]. The logistic regression used to solve the MLR model tends to systematically overestimate ORs or beta coefficients when the ratio of cases to variables is small [42]. Hence, extremely high coefficients (69.36), standard error (34,032 and OR (35,482) were computed by SPSS 17.0 [40] for some of the coefficients of the IVs (Appendix 10), thus lessening the power of analysis.

Table 25. Required Sample Size

TRANSIT AGENCY	TOTAL DRIVER POPULATION	PRIOR SAMPLE SIZE REQUIREMENT [27]	*RULE OF 10* SAMPLE SIZE [11]	ACTUAL SAMPLE SIZE
Regional	480	220	170	77
Urban	150	110	170	53

The GLM methodology used to develop the MLR model has the following limitations according to Agresti [3]:

- $\pi(x)$ produces probability values of 0 and 1 while actual DV values can take continuous values between 0 and 1.
- $\pi(x)$ can be < 0 and > 1 depending on the values of x. Hence, the model is valid over a finite range of x values.
- The least square estimators of the parameters are not optimal.
- The variance of the binary outcome Var (Y) = $\pi(x)[1 - \pi(x)]$ is not constant for all x.
- The maximum likelihood (ML) estimate for this model is computed by the statistical software since no formal formula is applicable.

5.2. Recommendations for Improving Model Results

Limitation caused by the small sample cannot be rectified at this stage. As indicated in Table 1, some transit agencies did not participate in the study. Among those that participated, the response rate was not very encouraging. It is strongly recommended that the Federal Transit Agency (FTA) and the American Public Transportation Association (APTA) get involved in public transit driver distraction studies and encourage the chief operating officer of transit agencies to conduct studies in their organizations. In order to be successful, it is imperative that the head of the organization is supportive of such studies. If larger number of drivers from more transit agencies participate, a larger sample can be drawn which will overcome the problems caused by small sample sizes.

Missing values have already been corrected in Section 4.2.2.1. Further improvement in the quality of data and model structure could be done by eliminating outliers (in excess of ± 2 standard deviations) from the continuous variables (Section 4.2.2.1). Interactions have an impact on the estimated coefficients of the fitted model, hence it is important to determine if there is interaction between any of the IVs. A simple way to check for statistical interaction recommended by Hosmer et al. [11] is to consider models with a pair of IVs: one is dichotomous (SEX) and one is continuous (EXP). Three models are fitted: a model containing only SEX; a model containing SEX and EXP; and a model containing SEX, EXP, and their statistical interaction SEX*EXP. Based on the p-values ($p<0.05$), one can decide on the best model to use out of the three models.

The original model had a four-category outcome variable (DISTRACTION ACTIVITY), with the following levels/categories: 0 = *Not Distracted*, 1 = *Slightly Distracted*, 2 = *Distracted*, and 3 = *Very Distracted*. Although the MLR can be used for any number of levels, Hosmer et al. [11] recommends three levels for simplification of the model. The SPSS 17.0 [40] Case Processing Summary is examined for dichotomous variables having very uneven splits of 90-10% or worse [15]. For Pedestrians, the most uneven split for the dichotomous IVs was 76.5 – 23.5% but in case of the DV, it was only 7.8% for the *Very Distracting* category. Hence, it was decided to reduce the number of DV levels from four to three by collapsing *Distracted* and *Very Distracted* levels raising the combined level to 25.4%. Hence, by combining the last two levels, the three levels were as follows: 0 = *Not Distracted*, 1 = *Slightly Distracted*, 2 = *Distracted*/*Very Distracted*. The first level, 0 = *Not Distracted* was used as the reference value.

The results from the original and transformed MLR model outputs are shown in Tables 26 and 27. The Classification Table 26 compares the observed versus predicted responses for all combinations of the dependent variable after and before the transformation. The case of the distracting activity *Pedestrian,* the overall performance for predicting the responses has improved from 82.4% 92.2% and the split is 23.5%, 52.9%, and 23.5% as against 25.5%, 47.1%, 19.6%, and 7.8% with four levels for the DV.

Table 26. Comparison of Classifications for Pedestrians

Classification for Pedestrians After Transformation

Observed	Predicted: NOT DISTRACTING	SLIGHTLY DISTRACTING	DISTRACTING/VERY DISTRACTING	Percent Correct
NOT DISTRACTING	11	2	0	84.6%
SLIGHTLY DISTRACTING	1	24	0	96.0%
DISTRACTING/VERY DISTRACTING	0	1	12	92.3%
Overall Percentage	23.5%	52.9%	23.5%	92.2%

Classification for Pedestrians Before Transformation

Observed	Predicted: NOT DISTRACTING	SLIGHTLY DISTRACTING	DISTRACTING	VERY DISTRACTING	Percent Correct
NOT DISTRACTING	11	2	0	0	84.6%
SLIGHTLY DISTRACTING	2	20	3	0	80.0%
DISTRACTING	0	2	7	0	77.8%
VERY DISTRACTING	0	0	0	4	100.0%
Overall Percentage	25.5%	47.1%	19.6%	7.8%	82.4%

The MLR Model output in Table 27 compares the Goodness of Fit and estimated coefficients after and before the transformation. Almost all tests indicated a good fit for the new transformed model with three levels of distractions. The likelihood ratio test using model fitting information 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) computes the chi-square (χ^2) = 73.090 (22), $p<0.001$, signifying a good improvement in the model fit. The model's Goodness of Fit as indicated by multiple statistics such as: the p-values for Pearson ($p = 0$) and Deviance ($p = 1$) proving significance and no significance. Hence, it can be partly 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 (0.76, 0.87, and 0.69) are reasonably similar and high values of R^2 indicating a good fit. The Table 27 presents outputs for the transformed and original logistic regression models for "Pedestrians "along with the coefficients, Wald Statistic, and OR and 95% CIs. A total of 10 variables are significant in the transformed model as against 4 variables in the original model.

Table 27. Transformed and Original MLR model outputs for Pedestrians

Transformed MLR model outputs for Pedestrians.

Model Chi-Square (χ^2) = 73.090 (22)**** Pearson Stat (SIG) Deviance Stat(NS)	R^2 = 0.76 (Cox & Snell); 0.87 (Nagelkerke); 0.69 (McFadden)	AIC initial/final values: N/A BIC initial/final values: N/A		
Independent Variables and Interactions	Coeff β (SE)	Wald Statistic	Odds Ratio Exp (B)	95% CI
Slightly Distracted vs. Not Distracted				
Intercept	14.92 (5.694)	6.869	N/A	N/A
DRIVE HRS/WK	0.088 (0.049)	3.274	1.092	0.99 – 1.20
PEAK = 0	-7.815 (2.706)	8.343	0.00 0	0.00 – 0.81
EDU = 0	-2.074 (1.067)	3.778	0.126	0.02 – 1.02
DAY = 0	-4.922 (2.290	4.618	0.007	0.00 – 0.65
EXP	- 0.868 (0.206)	17.813	0.420	0.28 – 0.63
Distracted/Very Distracted vs. Not distracted				
Intercept	-72.595 (26.289)	7.626	N/A	N/A
DRIVE HRS/WK	1.350 (0.454)	8.860	3.859	1.59 – 9.39
PEAK = 0	-11.143 (6.091)	3.346	0.000	0.00 – 2.22
AGE	0.565 (0.249)	5.151	1.760	1.08 – 2.87
MARITAL = 0	-8.294 (3.414)	5.901	0.000	0.00 – 0.20
DAY = 0	-14.143 (6.978)	4.108	0.000	0.00 – 0.627

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

Original MLR model outputs for Pedestrians.

Model Chi-Square (χ^2) = 80.420 (33)**** Pearson Stat (NS) Deviance Stat(NS)	R^2 = 0.79 (Cox & Snell); 0.87 (Nagelkerke); 0.66 (McFadden)	AIC initial/final values: N/A BIC initial/final values: N/A		
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	- 5.930 (3.458)	2.942	0.003	0 – 2.33
DRIVE HRS/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; ****p < 0.001. N/S = Not Significant.

6. References

[1] AAA Foundation for Traffic Safety. "Distraction in Everyday Driving." *University of North Carolina at Chapel Hill, Highway Safety Research Center*, 2003. www.aaafoundation.org.

[2] Afifi, A., May, S., and V. Clark. Practical Multivariate Analysis. 5th Edition (2012). CRC Press. Taylor and Francis Group.

[3] Agresti, Alan. (2007). *An Introduction to Categorical Data Analysis*, 2nd edition. *John Wiley and Sons, Inc., New York, NY 10158*.

[4] Blower, D., Green, P. E., & Matteson, A. (2008). "Bus Driver Types and Driver Factors in Fatal Bus Crashes: Results from the Buses Involved in Fatal Accidents Survey." *Federal Motor Carrier Safety Administration, DTMC75-06-H-00003 Task C*, Center for National Truck and Bus Statistics. University of Michigan Transportation Research Institute, 2901 Baxter Road, Ann Arbor Michigan 48109-2150, June 2008.

[5] D'Souza, K. A., and Maheshwari, S. K.. "Research framework for studying transit bus driver distraction." *Proceedings of the Urban Transport 2013 Conference*, Kos, Greece, May 29-31, 2013. Editors: J.W.S. Longhurst and C.A. Brebbia. Wessex Institute of Technology Press, Southampton, U. K, pp. 137 - 148

[6] 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,* Vol. 15, No. 3, 2012, pp. 1-23.

[7] D'Souza, K. A., and Maheshwari, S. K.. "Improving performance of public transit buses by minimizing driver distraction." *Proceedings of the Urban Transport 2012 Conference,* A Coruňa, Spain, May 13-16, 2012. Editors: J.W.S. Longhurst and C.A. Brebbia. Wessex Institute of Technology Press, Southampton, U. K: 281–293.

[8] Field, A. 2013. *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications Ltd., 4th Edition.

[9] Gkritza, K., D. Niemeier, and F. Mannering. 2006. Airport Security Screening and Changing Passenger Satisfaction: An Exploratory Assessment. *Journal of Air Transport Management* 12: 213 – 219.

[10] Hickman, J. S., Hanowski, R. J., & Bocanegra, J. (2010). Distraction in Commercial Trucks and Buses: Assessing Prevalence And Risk in Conjunction with Crashes and Near-Crashes. FMCSA-RRR-10-049. U.S. Department of Transportation, Federal Motor Carrier Safety Administration, Office of Analysis, Research and Technology, 1200 New Jersey Ave., SE, Washington, DC 20590.

[11] Hosmer, D. W., Lameshow, S., and R. X. Sturdivant. *Applied Logistic Regression*. 3rd Edition (2013). Wiley Series in Probability and Statistics. Wiley and Sons, Inc., New York, NY 10158.

[12] Hurtado, J. "The Effect of Marital Status on Car Insurance Rates". EasyFinance.com, July 24, 2012.

[13] HRT Part II of the Study on Bus Driver Distraction at Hampton Roads Transit (HRT), Hampton, VA 23661, June – August 2012.

[14] Hutcheson, G. D., and Moutinho, L. *Statistical Modeling for Management*. Sage Publications, Ltd., London, 2008.

[15] Hutcheson, G., and N. Sofroniou. *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models*. Sage Publications, Ltd., London, 1999.

[16] Johnson, Allie and John Egan. "Who are the Most Distracted Drivers." InsuranceQuotes.com.

[17] Lagarde, Emmanuel*†; Chastang, Jean-François*†; Gueguen, Alice*†; Coeuret-Pellicer, Mireille*†; Chiron, Mireille‡; Lafont, Sylviane. "Emotional Stress and Traffic Accidents: The Impact of Separation and Divorce."
Epidemiology: November 2004 - Volume 15 - Issue 6 - pp 762-766

[18] Lind, D. A., Marchal, W. G., & Wathen, S. A. (2006). Basic statistics for business and economics. *McGraw-Hill Irwin*, 5th. Edition

[19] Massie, D. L., and Campbell, K. L. "Analysis of Accident Rates by Age, Gender, and Time of the Day Based on the 1990 Nationwide Personal Transportation Survey." Final Report. The University of Michigan Transportation Research Institute, Ann Arbor, Michigan 48109-2150. 1993.

[20] McHugh, M. L. "The odds ratio: calculation, usage, and interpretation". *Biochemia Medica*. The Journal of Croatian Society of Medical Biochemistry and Laboratory Medicine, 2010. http://www.biochemia-medica.com/content.

[21] Metro Magazine. "Study: Transit Drivers Involved in Collisions Twice as Likely to Use Phone While Driving." *Metro Magazine. Industry News*. January 18, 2011.

[22] Mertler, C. A., and Vannatta, R. A. *Advanced and Multivariate Statistical Methods: Practical Application and Implementation*. 4th Edition. Pyrczak Publishing, Glendale, CA 91225.

[23] Morfoulaki, M., Y. Tyrinopoulos, and G. Aifadopoulou. 2007. Estimation of Satisfied Customers in Public Transport Systems: A New Methodological Approach. *Journal of the Transportation Research Forum* 46 (1): 63-72.

[24] Moutinho, L., and G. Hutcheson. 2011. Store Choice and Patronage: a Predictive Modelling Approach. *International Journal of Business Innovation and Research*. Vol. 1, No. 3, 2007, pp. 233 -252.

[25] 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.

[26] Norman, L. G. "Road Traffic Accidents: Epidemiology, Control, and Prevention." *World Health Organization*, Geneva, 1962.

[27] Orcher, Lawrence, T. (2007). Conducting a Survey: Techniques for a Term Project. *Pyrczak Publishing, P. O. Box 250430, Glendale, CA 91225*, pp. 46 – 47.

[28] Owens, J. M., and R. Lehman. "The Effects of Age and Distraction on Reaction Time in a Driving Simulator." *1st International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, Aspen, CO, August 14 – 17, (2001): 147-152.

[29] Peng, C. Y and R. N. Nichols. "Using Multinomial Logistic Models to Predict Adolescent Behavioral Risk." *Journal of Modern Applied Statistical Methods* 2 (1), 2003, pp. 1 – 13.

[30] Petrucci, Carrie J. "A Primer for Social Worker Researchers on How to Conduct a Multinomial Logistic Regression. *Journal of Social Service Research* 35(2), (2009): 193 – 205.

[31] Potomac and Rappahannock Transportation Commission Transit Development (PRTC) "A Study of Bus Operator Distraction." September – December 2012.

[32] Powell, N. B., Schechtman, K. B., Riley, R. W., Guilleminault, C, Ping-Ying Chiang, R., and E. M. Weaver. "Sleepy Driver Near-Misses May Predict Accident Risks." *SLEEP*, 30 (3), 2007, pp. 331 – 342.

[33] Ranney, T. A. "Driver Distraction: A Review of the Current State-of-Knowledge." *Report Number DOT HS 810 787*, 2008. National Highway Traffic Safety Administration Final Report. 1200 New Jersey Avenue SE.14. Washington, DC 20590.

[34] Salmon, P. M., K. L. Young, and M. A. Regan. **"**Distraction 'On the Buses': A Novel Framework of Ergonomics Methods for Identifying Sources and Effects of Bus Driver Distraction." *Applied Ergonomics* 42, (2011): 602-610.

[35] Salmon, P.M., Young, K.L., & M. A. Regan (2006). "Analysis of risk for State Transit Authority New South Wales Bus Drivers. Bus Driver Distraction Stage 1: Final Report: September 2005 – June 2006." *Monash University Accident Research Centre*, Wellington Road, Clayton, Victoria, 3800, Australia.

[36] Salvucci, D. D., Chavez, A. K., and F. J. Lee (2004). Modeling Effects of Age in Complex Tasks: A Case Study in Driving. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*.

[37] Salvucci, D. D (2002). Modeling Driver Distraction from Cognitive Tasks.. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Pp. 457 – 492.

[38] Santen, I., Esser, J., Neubert, I., Wahle, J., Schadschneider, and M. Schadschneider (1998). Online Simulation of Urban Traffic Using Cellular Automata. *Proceedings of the 3rd IMA International Conference on Mathematics in Transport Planning and Control* Edited by J. D. Griffiths. Pergamon Elsevier Science, pp. 301–309.

[39] Smith, D. I., J. Chang, D. Cohen, J. Foley, and R. Glassco. "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.

[40] SPSS 17.0 (2008). SPSS 12.0, SPSS Inc., Chicago, IL 60606-6412.

[41] Statista. "Age of U. S. Urban Transit Bus Vehicle." WWW.Statista.com/statistics/ 2012.

[42] Tabachnick, B. G., and Fidell, L. S. *Using Multivariate statistics. (sixth edition). Pearson Education, Inc., New Jersey 07458*. 2013.

[43] Tefft, Brian C. "Motor Vehicle Crashes, Injuries, and Deaths in Relation to Driver Age: United States, 1995–2010." *AAA Foundation for Traffic Safety*. November 2012. www.aaafoundation.org.

[44] Transportation. "Code Title 7. Vehicles and Traffic Subtitle C." *Rules Of The Road Chapter 541. Definitions*.

[45] Trick, L. M., Enns, J. T., Mills, J. and J. Vavrik. "Paying Attention Behind the wheel: a framework for studying the role of attention in Driving." *Theoretical Issues in Ergonomics Science*. 5 (5), pp 385 to 424

[46] U. S. Department of Transportation. "An Examination of Driver Distraction as Recorded in NHSTA Databases." *DOT HS 811 216*, 2009.

[47] U.S. Department of Commerce. "American Community Survey" *U. S . Census Bureau*, Washington, DC, October 2010.

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

[49] U. S. Department of Transportation. "Statistics and Facts about Distracted Driving." 2010. www.distraction.gov/stats-and-facts.

[50] U. S. Department of Transportation. “Blue Print for Ending Distracted Driving.” NHTSA, *DOT HS 811 629*, June 2012.

[51] Vittinghof, E., and C. E. McCulloch. Relaxing the Rule of Ten Events Per Variable in Logistic and Cox Regression. *American Journal of Epidemiology*. Vol. 165, pp. 710 – 718.

[52] Virginia Department of Motor Vehicle. “Virginia Traffic Crash Facts.” *Virginia Crashes Involving Buses*. Virginia Highway Safety Office, (2011), pp. 15-17..

[53] Washington, S. P., M. G. Karlaftis, and F. L. Mannering. 2011. Logistic Regression, Discrete Outcome Models and Ordered Probability Models. In *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC: 303 - 359.

[54] Wong, Jinn-Tsai and Shah-Hsuan Huang. “Modeling Driver Mental Workload for Accident Causation and Prevention”. *Journal of the Eastern Asia Society for Transportation*, Vol. 8, 2009.

[55] 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**, pp. 983-995, 2005.

[56] Yan, X., E. Radwan, and K. K. Mannila. “Analysis of Truck-Involved Rear-End Crashes Using Multinomial Logistic Regression.” *Advances in Transportation Studies an International Journal* Section A 17, (2009): 39 – 52.

APPENDICES

Appendix 1. Hampton University Transportation Center Bus Driver Distraction Survey.

Instructions for filling the survey- Please answer all questions completely to the best of your ability. All of the information provided will be used as source information for analysis purposes only and no names are requested or required.

Part I – Demographic Details

1. **What is your age in years**? ☐

2. **Are you male or female**? ☐ Male ☐ Female

3. **Education Level** ☑ **(check one)**

☐ Less than High School ☐ High School or GED ☐ Some College ☐ 2-year College ☐ 4-year college or more

4. **Marital Status** ☑ **(check one)**

☐ Married ☐ Separated ☐ Divorced ☐ Never Married ☐ Don't want to reveal

Part II – Driving experience and travel patterns

1. **How many years have you been driving a bus**? ☐

2. **On an average, how many hours per week do you spend driving a bus?** ☐

3. **When do you normally drive a bus?** ☑ **(Check one which is most applicable)**

☐ During the day ☐ At night ☐ During peak hours ☐ Non-peak times ☐Other

4. **Where do you spend most of your time driving a bus?** ☑ **(Check one which is most applicable)**

☐ Commuter ☐ Metro Feeder ☐ Local Bus ☐ Others, _______________

Part III – Equipment

1. **Type of equipment you normally drive** ☑ **(Check one which is most applicable)**

☐ Orion V 40' ☐ Gillig 40'High floor ☐ MCI ☐ Gillig 40' low floor ☐ Gilling 30" Low floor

Parts IV –Source and Extent of Distraction - Please indicate by ☑ HOW distracting you find the following sources while driving the bus. **Please answer for all activities.**

Activity	Not Distracting	Slightly Distracting	Distracting	Very Distracting
Dispatch Broadcasts				
Communication with Dispatch				
Passengers (moving around, standing and talking next to driver's cabin)				
Passengers using a mobile phone				
Passengers trying to talk to driver				
Disabled passengers				
Passengers not following etiquette (eating, drinking, smoking, noisy)				
Passengers with Infants				
General/other Broadcasts				
Mobile Data Terminals (MDTs)				
Ticket Machine/Farebox				
Announcing Bus Stops				
Reading e.g. Route Sheet				
Driver's Mobile Cell phone				
Looking at Advertisement				
Climate Controls				
Audible alerts				
On-board rattles				
Fatigue/Sickness				
Other Road Users				
Pedestrians				
Food and Other Smells				
Unruly Kids				

Part V –Duration of Distraction- Please indicate by ☑ approximately how long per shift (assuming you were working a typical 8-hour shift), do you experience the following distracting activities while driving the bus. **Please answer for all activities.**

Activities	N/A	0-1hr	1-2hrs	2-3hrs	3-5hrs	+5hrs
Dispatch Broadcasts						
Communication with Dispatch						
Passengers (moving around, standing and talking next to driver's cabin)						
Passengers using a mobile phone						
Passengers trying to talk to driver						
Disabled passengers						
Passengers not following etiquette (eating, drinking, smoking, noisy)						
Passengers with Infants						
General/other Broadcasts						
Mobile Data Terminals (MDTs)						
Ticket Machine/Farebox						
Announcing Bus Stops						
Reading e.g. Route Sheet						
Driver's Mobile Cell phone						
Looking at Advertisement						
Climate Controls						
Audible alerts						
On-board rattles						
Fatigue/Sickness						
Other Road Users						
Pedestrians						
Food and Other Smells						
Unruly Kids						

Parts VI –Perceived Effect of Distraction- Please indicate which effect each activity has upon your driving? (**Check ☑ all that apply for each activity**).

Activities	Eyes off the road	Mind/Attention off the road	Physical Interference
Dispatch Broadcasts			
Communication with Dispatch			
Passengers (moving around, standing and talking next to driver's cabin)			
Passengers using a mobile phone			
Passengers trying to talk to driver			
Disabled passengers			
Passengers not following etiquette (eating, drinking, smoking, noisy)			
Passengers with Infants			
General/other Broadcasts			
Mobile Data Terminals (MDTs)			
Ticket Machine/Farebox			
Announcing Bus Stops			
Reading e.g. Route Sheet			
Driver's Mobile Cell phone			
Looking at Advertisement			
Climate Controls			
Audible alerts			
On-board rattles			
Fatigue/Sickness			
Other Road Users			
Pedestrians			
Food and Other Smells			
Unruly Kids			

Appendix 2. Route Observation Form (Illustrated)

DISTRACTING ACTIVITY	MLR MODEL AND SIMULATION RESULTS	ROUTE OBSERVATIONS
Passenger Using Mobile Phone	This distraction was impacted by Location, Sex, Age, Driving Experience, and Driving Hours/Week.	***Very few passengers used mobile phone and could hardly be heard by driver.***
Passengers	This distraction was impacted by Location, Sex, Driving Experience, and Driving Hours/Weeks.	***Passenger having very loud conversation in front of bus. Distracting to all on the bus.***
Passenger Talk to Driver	No significant factors.	***No passengers talked to the driver while the bus was in motion and those who talked while stopped were asking for the best route***
Ticket Machine	This distraction was impacted by Location, Sex, Driving Experience, and Driving Hours/Weeks.	***No distraction observed. Ticket machine was operated during stops.***
Fatigue/Sick	This distraction was impacted by Sex.	***Driver was just starting shift so she was alert and focused***
Non-Etiquette Passengers	This distraction was impacted by Location, Sex, Age, and Driving Hours/Weeks.	***No distraction observed.***
Other Road Users	This distraction was impacted by Age, Driving Hours/Weeks.	***No distraction observed.***
Pedestrians	No significant factors.	***No distraction observed.***

Appendix 3. Sources Urban Driver Distraction Ranked On The Average Distraction Rating.

Rank	Activity	Rating
1	Unruly Kids	2.85
2	Passenger Using Mobile Phone	2.70
3	Driver's Mobile Phone	2.60
4	Passengers Not Following Etiquette	2.37
5	Passengers	2.36
6	Fatigue/ Sickness	2.34
7	Passengers Trying to Talk to Driver	2.30
8	Mobile Data Terminals	2.29
9	Looking at Advertisements	2.25
10	Pedestrians	2.10
11	Other Road Users	2.08
12	Communication with Dispatch	2.00
13	General Broadcasts/ Other	1.95
14	Passengers with Infants	1.93
15	On-board Rattles	1.93
16	Food and Other Smells	1.90
17	Reading	1.83
18	Dispatch Broadcast	1.70
19	Climate Control	1.68
20	Audible Alerts	1.54
21	Disabled Passengers	1.52
22	Announcing Bus Stops	1.48
23	Ticket Machine/ Farebox	1.35

Appendix 4. Duration of Urban Driver Distraction Ranked on the Average Distraction Durations.

Rank	Activity	Rating
1	Pedestrians	2.45
2	Other Road Users	2.25
3	Announcing Bus Stops	2.00
4	Ticket Machine/ Farebox	1.98
5	On-board Rattles	1.89
6	Passengers Using Mobile Phone	1.84
7	Passengers Not Following Etiquette	1.76
8	Mobile Data Terminals	1.75
9	Audible Alerts	1.61
10	Passengers	1.56
11	Climate Control	1.55
12	Passengers with Infants	1.44
13	Passenger Trying to Talk to Driver	1.40
14	Food and Other Smells	1.23
15	Dispatch Broadcasts	1.22
16	Disabled Passengers	1.20
17	Driver's Mobile Phone	1.19
18	General Broadcasts/ Other	1.19
19	Looking at Advertisements	1.18
20	Communication with Dispatch	1.12
21	Unruly Kids	1.10
22	Reading	1.09
23	Fatigue/ Sickness	1.02

Appendix 5. Average Distractions Ratings for Male and Female Drivers.

Male

Female

Rank	Activity	Rating	Rank	Activity	Rating
1	Unruly Kids	2.81	1	Passengers Using Mobile Phone	3.09
2	Passengers Using Mobile Phone	2.58	2	Unruly Kids	2.92
3	Driver's Mobile Phone	2.45	3	Driver's Mobile Phone	3.00
4	Passengers Trying to Talk to Driver	2.25	4	Mobile Data Terminals	2.90
5	Passengers Not Following Etiquette	2.25	5	Passengers Not Following Etiquette	2.73
6	Passengers	2.24	6	Looking at Advertisements	2.64
7	Fatigue/ Sickness	2.15	7	Passengers	2.67
8	Looking at Advertisements	2.10	8	Looking at Advertisements	2.64
9	Mobile Data Terminals	2.04	9	Other Road Users	2.55
10	Pedestrians	1.93	10	Pedestrians	2.50
11	Other Road Users	1.89	11	Communication with Dispatch	2.50
12	Communication with Dispatch	1.82	12	Passenger Trying to Talk to Driver	2.45
13	General Broadcasts/ Other	1.81	13	On- board Rattles	2.42
14	Food and Other Smells	1.79	14	Passengers with Infants	2.42
15	Passengers with Infants	1.73	15	Reading	2.30
16	On- board Rattles	1.71	16	Climate Control	2.25
17	Reading	1.62	17	General Broadcasts/ Other	2.18
18	Dispatch Broadcasts	1.58	18	Dispatch Broadcasts	2.08
19	Announcing Bus Stops	1.52	19	Food and Other Smells	2.04
20	Disabled Passengers	1.47	20	Audible Alerts	1.83
21	Climate Control	1.45	21	Disabled Passengers	1.67
22	Audible Alerts	1.41	22	Announcing Bus Stops	1.50
23	Ticket Machine/ Farebox	1.33	23	Ticket Machine/ Fare	1.38

Appendix 6. Average Distraction Durations for Male and Female Drivers.

Male			Female		
Rank	Activity	Average Duration	Rank	Activity	Average Duration
1	Pedestrians	2.45	1	Mobile Data Terminals	2.88
2	Other Road Users	2.26	2	Pedestrians	2.45
3	On-board Rattles	2.07	3	Announcing Bus Stops	2.41
4	Ticket Machine/ Farebox	1.95	4	Passengers Using Mobile Phone	2.35
5	Audible Alerts	1.77	5	Ticket Machine/ Farebox	2.32
6	Announcing Bus Stops	1.72	6	Other Road Users	2.30
7	Passengers Not Following Etiquette	1.64	7	Passengers Using Mobile Phone	2.19
8	Driver's Mobile Phone	1.60	8	Passengers	2.06
9	Passengers Using Mobile Phone	1.52	9	Passengers Not Following Etiquette	2.00
10	Looking at Advertisements	1.27	10	Climate Control	1.83
11	Mobile Data Terminals	1.23	11	Passengers Trying to Talk to Driver	1.70
12	Passengers	1.14	12	Food and Other Smells	1.63
13	Unruly Kids	1.12	13	Reading	1.58
14	Passengers with Infants	1.09	14	Communication with Dispatch	1.50
15	Passenger Trying to Talk to Driver	1.08	15	General Broadcasts/ Other	1.50
16	Fatigue/ Sickness	1.06	16	On- board Rattles	1.45
17	Disabled Passengers	1.04	17	Disabled Passengers	1.40
18	Food and Other Smells	1.03	18	Audible Alerts	1.29
19	Dispatch Broadcasts	1.02	19	Dispatch Broadcast	1.25
20	Communication with Dispatch	0.86	20	Looking at Advertisements	1.00
21	General Broadcasts/ Other	0.85	21	Fatigue/ Sickness	0.90
22	Reading	0.70	22	Unruly Kids	0.83
23	Looking at Advertisements	0.37	23	Driver's Mobile Phone	0.50

Appendix 7. Perceived Effect of Distraction- Number of Bus Drivers (of a possible 53)

Activities	Eyes off the road (1)	Mind/Attention off the road (2)	Physical Interference (3)	No Response (N/A)
Dispatch Broadcast	3	28	6	16
Communication with Dispatch	8	23	7	15
Passengers using a mobile phone	7	25	3	18
Passengers	13	21	4	15
Passengers trying to talk to driver	7	24	4	18
Passengers not following etiquette (eating, drinking, smoking, noisy)	8	21	4	20
Passengers with Infants	5	20	6	22
General/ Other Broadcasts	3	27	4	19
Disabled Passengers	7	9	17	20
Mobile/ Data Terminals	10	21	5	17
Ticket Machine/ Farebox	11	21	3	18
Announcing Bus Stops	4	22	6	21
Reading	13	14	4	22
Driver's Mobile Phone	8	18	4	23
Looking at Advertisements	13	16	3	21
Climate Controls	11	14	6	22
Audible Alerts	3	26	2	22
On- board Rattles	8	20	4	21
Fatigue/ Sickness	5	15	14	19
Other Road Users	13	16	3	21
Pedestrians	12	19	4	18
Food and Other Smells	7	21	4	21
Unruly Kids	10	20	6	17

Appendix 8. Handling Missing Dada

VARIABLE	TYPE OF MISSING DATA	REPLACEMENT APPROACH
Age	Nonrandom missing data	Regression approach (Imputation)
Experience	Nonrandom missing data	Regression approach (Imputation)
Driving hrs/week	Missing at random (MAR)	Regression approach (Imputation)
Marital Status	Nonrandom missing data	Regression approach (Imputation)
Education Level	Nonrandom missing data	Regression approach (Imputation)
Equipment	Missing at random (MAR)	Regression approach (Imputation)
Day	Missing at random (MAR)	Prior Knowledge/well-educated guess
Peak	Missing at random (MAR)	Prior Knowledge/well-educated guess
Location	Missing at random (MAR)	Prior Knowledge/well-educated guess
Distracting Activity Ratings	Missing Completely at Random (MCAR)	Mean Substitution

Appendix 9. Summary of Model Fitting for Very High and High Risk Distracting Activities.

	STEP SUMMARY				OVER DISPERSION		GOODNESS OF FIT			PSEUDO R SQUARE	
DISTRACTING ACTIVITY	-2 LOG LIK (INITIAL, FINAL)	SIG	-2 LOG LIKELIHOOD χ^2 (INITIAL,FINAL)	SIG	$\frac{\chi^2_{PEAR}}{df}$	$\frac{\chi^2_{DEV}}{Df}$	PEAR SIG	DEV SIG	COX & SNELL	NAG	McFAD
PEDESTRIAN	(70, 42)	0.000	(123, 42)	0.000	0.443	0.362	1.000	1.000	0.793	0.872	0.655
PASSENGERS	N/A	N/A	(128, 93)	0.242	1.500	0.773	0.000	0.969	0.497	0.541	0.274
OTHER ROAD USERS	N/A	N/A	(113, 65)	0.018	0.737	0.539	0.986	1.000	0.614	0.659	0.429
UNRULY KIDS	N/A	N/A	(130, 95)	0.230	0.960	0.79	0.610	0.957	0.500	0.542	0.272
PASS MOBILE PHONE	(94, 71)	0.036	(131, 71)	0.008	0.651	0.620	0.999	1.000	0.691	0.749	0.459
MOB DATA TERMINAL	(80, 69)	0.010	(126, 69)	0.006	1.517	0.591	0.000	1.000	0.673	0.735	0.452
PASS NOT ETIQUET	N/A	N/A	(131, 79)	0.006	1.115	0.656	0.183	0.999	0.644	0.697	0.401
FARE BOX	(52, 46)	0.020	(60, 46)	0.255	1.159	1.188	0.229	0.196	0.234	0.339	0.227

Appendix 10. DISTRACTING ACTIVITIES WITH SIGNIFICANT VARIABLES.

ZONE I											
ACTIVITY		B	STD ERROR	WALD	Df	SIG	ODDS RATIO	LOWER BOUND	UPPER BOUND	OR > OR < 1	CL INCL 1
PEDESTRIAN (SD)	Intercept	0.472	5.59	0.007	1						
	Experience	-0.487	0.183	7.10	1	0.008	0.615	0.430	0.879	< 1	No
PEDESTRIAN (D)	Intercept	-13.47	8.33	2.61	1						
	Drive Hrs/Wk	0.178	.0.082	4.65	1	0.031	1.194	1.016	1.403	> 1	No
	Age	0.264	0.147	3.21	1	0.073	1.302	0.976	1.737	> 1	Yes
	Peak=0	-5.93	3.46	2.94	1	0.086	0.003	3E-6	2.332	< 1	Yes
PASSENGER (SD)	Intercept	5.03	3.74	1.81	1						
	Peak=0	-2.70	1.65	2.69	1	0.102	0.67	0.003	1.71	< 1	Yes
PASSENGER(D)	Intercept	-2.17	4.88	0.197	1						
	Peak=0	-5.56	2.43	5.25	1	0.022	0.004	3E-5	0.448	< 1	No
	Day=0	-5.06	2.30	4.92	1	0.027	0.006	7E-5	0.56	< 1	No
	Equip=0	3.74	2.03	3.39	1	0.066	42.16	0.786	2262	>1	Yes
OTHER ROAD USER (SD)	Intercept	26.08	4958	0.00	1						
	Marital=0	8.88	4.91	3.27	1	0.07	7194	0.477	1E8	>1	Yes
	Exp	-0.482	0.225	4.59	1	0.032	0.618	0.397	0.960	< 1	No
OTHER ROAD USER (D)	Intercept	22.26	4958	0.00	1						
	Marital=0	9.69	4.97	3.80	1	0.051	16,179	0.953	3E8	>1	Yes
	Exp	-0.454	0.009	3.94	1	0.047	0.635	0.405	0.994	< 1	No
UNRULY KIDS (SD)	Intercept	9.03	7.05	1.64	1						
	Loc=0	-5.65	3.31	2.92	1	0.088	0.004	5E-6	2.31	< 1	Yes
	Day=0	-6.26	3.02	4.28	1	0.0.038	0.002	5E-6	0.716	< 1	No
FARE BOX (SD)	Intercept	-0.535	2.963	0.033	1						
	SEX = 0	3.00	1.419	4.48	1	0.034	20.14	1.248	325.0	> 1	No
	EDU = 0	2.180	1.235	3.11	1	0.078	8.814	0.786	99.56	> 1	Yes
	DAY = 0	-2.635	1.612	2.67	1	0.100	0.072	0.003	1.688	< 1	Yes

Appendix 10 (Continued).

ZONE II											
ACTIVITY		B	STD ERROR	WALD	Df	SIG	ODDS RATIO	LOWER BOUND	UPPER BOUND	OR > OR < 1	CL INCL 1
PASS MOBILE PHONE (D)	Intercept	69.36	34.90	3.95	1						
	Peak=0	-31.82	2.16	217.13	1	0.000	0.000	0.000	0.000	< 1	No
PASS MOBILE PHONE (SD)	Intercept	58.69	34032	2.92	1						
	Peak=0	-30.33	1.96	239.07	1	0.000	0.000	0.000	0.000	< 1	No
MOB DATA TERMINAL (SD)	Intercept	8.35	4.64	3.24	1						
	Age	-0.156	0.084	3.40	1	0.065	0.856	0.725	1.010	< 1	Yes
	Experience	-1.431	0.604	5.61	1	0.018	0.239	0.073	0.781	< 1	No
MOB DATA TERMINAL (VD)	Intercept	56.90	7.73	54.12	1						
	Peak=0	-56.06	2.21	641.12	1	0.000	0.000	0.000	0.000	< 1	No
	Sex=0	3.57	1.91	3.51	1	0.061	35.64	0.849	1497	> 1	Yes
	Equip=0	3.750	2.232	2.823	1	0.093	42.50	0.536	3372	>1	Yes
PASS NOT ETIQUET (SD)	Intercept	8.00	4.25	3.54	1						
	Equip=0	3.28	1.85	3.15	1	0.076	26.68	0.708	1005	>1	Yes
	Experience	-0.218	0.100	4.75	1	0.029	0.804	0.061	0.978	< 1	No
PASS NOT ETIQUET (D)	Intercept	6.14	5.05	1.48	1						
	Peak=0	-5.48	2.77	3.91	1	0.048	0.004	0.000	0.952	< 1	No
	Day=0	-5.73	2.72	4.46	1	0.035	56.37	0.741	4288	< 1	No
	Equip = 0	4.032	2.21	3.33	1	0.035					
	Experience	-0.38	0.179	4.54	1	0.033	0.684	0.482	0.970	< 1	No
PASS NOT ETIQUET (VD)	Intercept	9.42	6.23	2.28	1						
	Age	-0.391	0.184	4.50	1	0.034	0.676	0.471	0.971	< 1	No
	Marital=0	6.12	3.19	3.69	1	0.055	454.46	0.880	234,732	>1	Yes
	Edu=0	10.49	5.16	4.13	1	0.042	35482	1.452	88500	>1	No

Appendix 11. Response Variables Correctly Predicted.

DISTRACTING ACTIVITIES	NOT DISTRACTING	SLIGHTLY DISTRACTING	DISTRACTING	VERY DISTRACTING	OVERALL PERCENTAGE
PEDESTRIAN	84.6%	80.0%	77.8%	100%	82.4%
PASSENGERS	33.3%	88.0%	55.6%	50.0%	66.7%
OTHER ROAD USERS	84.6%	880.0%	77.8%	100%	82.4%
UNRULY KIDS	50.0%	69.2%	63.2%	46.7%	58,8%
PASS MOBILE PHONE	75.0%	83.3%	50.0%	61.5%	66.7%
MOB DATA TERMINAL	61.5%	79.2%	80.0%	77.8%	74.5%
PASS NOT ETIQUET	50.0%	87.0%	60.0%	75.0%	72.5%
FARE BOX	94.6%	35.7%	N/A	N/A	78.4%

Appendix 12. MLR Linear Predictors (Functions) for Risk Zones I and II Distracting Activities.

Activity	Slightly Distracted (1)	Distracted (2)	Very Distracted (3)
PEDESTRIAN	Y_{11} = – 0.487*Exp	Y_{12} = -13.47 +0.178*Drive Hrs/Wk + 0.264*Age – 5.937*Peak=0	NS
PASSENGERS	Y_{21} = – 2.70*Peak=0	Y_{22} = – 5.56*Peak=0 – 5.06*Day=0 + 3.74*Equip=0	NS
OTHER ROAD USERS	Y_{31} = 8.88*Marital=0 – 0.482*Exp	Y_{32} = 9.69*Marital=0 – 0.454*Exp	NS
UNRULY KIDS	Y_{41} = 5.65*Loc=0 – 6.26*Day=0	NS	NS
PASS MOBILE PHONE	Y_{51} = 69.36 – 31.82*Peak=0	Y_{52} = 58.69 – 30.33*Peak=0	NS
MOB DATA TERMINAL	NS	NS	Y_{63} = 56.90 – 56.06*Peak=0 + 3.57*Sex=0 + 3.75*Equip=0
PASS NOT ETIQUET	Y_{71} = 8.00 + 3.28*Equip=0 – 0.218*Exp	Y_{72} = – 5.48*Peak=0 – 0.38*Exp -5.73*Day=0 + 4.032*Equip=0	Y_{73} = – 0.391*Age + 6.12*Marital=0 + 10.19*Edu=0
FARE BOX	Y_{81} = 3.00*Sex=0 + 2.18*Edu=0 – 2.64*Day=0	NS	NS

Note: SPSS 17.0 sets the reference level Not Distracted = 0 with Slightly Distracted (1), Distracted (2), and Very Distracted (3). (Y_{ij}) is the estimated utility function that measures the total contribution of each significant factor where, i = 1 to 8, and j = 1 to 3. N/S = MLR Model individual independent variables (IVs) were not significant.