

Multivariate Statistical Analysis of Public Transit Bus Driver Distraction

K. A. D'Souza, Ph.D., and S. K. Maheshwari, Ph.D., P.E.
Hampton University

Abstract

This paper examines the efficacy of a multivariate statistical modeling approach to analyze public transit bus driver distraction data collected through a self-administered driver survey. The distracting activities were classified into four risk zones according to distraction risk indices derived from distracting ratings, distracting durations, and driver perception of risks. A multinomial logistic regression model was formulated for highly-risky distracting activities using levels of distraction as the categorical dependent variable and correlating it with categorical and continuous independent variables responsible for the distraction. Results revealed that the common sources of distraction were due to passenger-related activities, which match two-thirds of simulated validation outputs. On-site route observations and discussions with transit staff revealed mixed results. The model could be used to identify drivers at highest distraction risk from their demographic backgrounds as well as driving schedules. The transit agency can use the results to implement relevant policies and training programs to mitigate distraction and improve transit performance.

Introduction

Over the past few years, distracted driving accidents have increased due to the proliferation of electronic devices use while driving and greater driver involvement in secondary tasks (U. S. DOT 2009). Research on transit bus driver distraction conducted in the U. S. is limited (D'Souza and Maheshwari 2012a and 2012b), although

sources of distraction are generally higher due to the driver performing required secondary tasks and attending to many passengers. The National Transit Database (NTD) was used to study the relationship between transit bus collisions and factors such as road design, weather, lighting, etc. (Yang 2007), but driver distraction was not included as a factor in this study. Due to lack of distraction-reporting by transit drivers, the associated risks and impact on performance is difficult to study and is not well-understood.

The increasingly complex nature of distraction data that consist of multiple predictors and categorical outcome variables requires an appropriate multivariate statistical model to relate the levels of distracting activities with factors that impact distraction. Multinomial logistic regression (MLR) or multinomial logit (MNL) models are widely used in transportation to study relationships between categorical dependent variables and sets of continuous and/or categorical independent variables (Washington et al. 2011; Yan et al. 2009; Morfoulaki et al. 2007; Gkritza et al. 2006).

This paper attempts to establish the likelihood that a transit bus driver's risk of getting distracted is related to his/her demographic background, driving hours per week, and location. It is based on an exploratory distracted-driving study conducted at a regional public transit agency serving six cities and surrounding suburbs within an area of 369 square miles and an annual ridership of 18 million. The distracting activities identified from a survey were classified into risk zones according to distraction risk indices derived from distracting ratings, distracting durations, and driver perception of risks. The MLR was used to model highly-risky distracting activities using levels of distraction as the categorical dependent variable and correlating it with driver demographics, location, and driving load as independent variables. The independent variables were identified from the literature review (Salmon et al. 2011), discussions with the transit agency, and sample route observations. Comparison of the MLR model's results with the simulated outputs show similarities for two-thirds of the model values. On-site field observations and discussions with transit staff were conducted to verify the discrepancies between MLR model and simulated results.

The remainder of this paper is organized as follows. The next section develops a Distraction Risk Index to classify distracting activities into their respective risk zones. This is followed by MLR modeling of highly-risky distracting activities and interpretation of results. The results are statistically assessed and validated using Monte Carlo simulation and on-site route observations. The last section concludes the paper and discusses some of the applications and limitations.

Classification of Distraction Activities

Data Collection and Analysis

The survey instrument used by Salmon et al. (2006) was redesigned to collect data on driver demographics, driving pattern, source, duration and perception of distraction. The survey was approved by the transit agency and the Hampton University Institutional Review Board (IRB). This self-administered survey was conducted during the summer of 2011 on a group of drivers located in the transit agency's Northside and Southside operation centers. The region was divided into two areas: the Northside and Southside due to the difference in population density, street layouts, and accident rates. The Southside is more commercialized and densely populated with a higher accident rate of 62 accidents/million miles compared to the Northside's rate of 54 accidents/million miles.

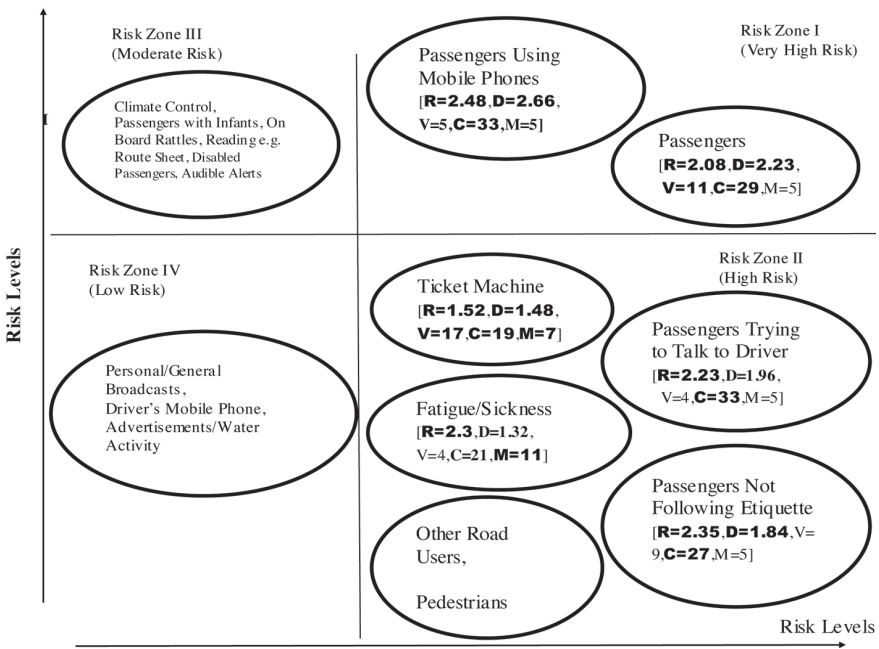
Completing the survey was voluntary, requiring the driver's written consent and assurance of confidentiality of their responses. An agency representative was assigned to distribute the surveys, deliver the introduction, answer questions, and assist in the survey process. A total of 265 surveys were distributed to 100 drivers on the Northside and 165 on the Southside. A total of 48 (19 from the Northside and 29 from the Southside) completed surveys were returned, resulting in an 18 percent response rate. The survey responses reflected the perception of the drivers who were the primary sources for distraction information. Their responses were fairly consistent and comparable with other transit bus surveys (Salmon et al. 2006).

The transit bus drivers rated how distracting they found listed activities and the approximate duration they experienced these activities in a typical eight-hour shift (Salmon et al. 2006). The ratings and durations for each activity were averaged and ranked from highest to lowest. Each distracting activity's rating and duration were graded as a percentage (%) relative to the highest rating (2.48) and highest duration (2.66 hours).

The U.S. Department of Transportation (2010) has categorized distractions as Visual, Manual, and Cognitive and reported that the severity of distractions increases as it involves more than one category. The bus drivers were asked to categorize each distracting activity according to their perception. The total responses from the bus drivers were ranked from highest to lowest. The number of driver responses for distracting activities in each category was graded as a percentage (%) relative to the highest visual (19 driver responses), cognitive (33 driver responses), and manual (11 driver responses).

Development of the Distraction Risk Index

The graded scores for rating and duration of distraction, visual, cognitive and physical distractions are summarized in Table 1. The graded scores of each distracting activity were averaged to produce the Distraction Risk Index (DRI) that measures the potential risk associated with each risk zone activity. The DRIs ranged from 31–74 percent (excluding “Others”), with a mean of 56.77 percent and standard deviation of 11.5 percent. Following the approach of Peng and Nichols (2003), distracting activities scoring a DRI of at least one standard deviation above the mean, i.e., 70 percent or higher were identified as Risk Zone I (very high risk) activities. Those scoring between 60 and 69 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 between 50 and 59 percent, and the range for Risk Zone IV (low risk) was set at DRI scores below 50 percent. The graded scores of all distracting activities with the DRIs and assigned risk zones are shown in Table 1. Two distracting activities were classified into Risk Zone I, six into Risk Zone II, six into Risk Zone III, and remaining five into Risk Zone IV (Figure 1).



R = Distracting Rating; D = Distracting Duration; V = Visual Perception; C = Cognitive Perception; M = Manual/Physical Perception. The respective values are obtained from final study report. Bold indicates the values are critical for the assigned Risk Zone. The position of distracting activities within each quadrant is not related to its risk level.

Figure 1. Classification of Distracting Activities into Risk Zones

Table 1. Graded Scores and Distraction Risk Index for Each Distracting Activity

Distracting Activities	Distraction Rating (% of highest)	Distraction Duration (% of highest)	Eyes Off Road (% of highest)	Mind/Attention Off Road (% of highest)	Physical Interference (% of highest)	Avg. (Distraction Risk Index)	Risk Zone
Passengers Using Mobile Phones	100%	100%	26%	100%	45%	74%	I
Passengers (Moving Around, Standing Next to Driver's Cabin, Talking Next to Driver's Cabin)	84%	84%	58%	88%	45%	72%	I
Passengers Not Following Etiquette (Eating, Drinking, Smoking, Noisy)	95%	69%	47%	82%	45%	68%	II
Passengers Trying to Talk to Driver	90%	74%	21%	100%	45%	66%	II
Ticket Machine	61%	56%	89%	58%	64%	66%	II
Fatigue/Sickness	85%	50%	21%	64%	100%	64%	II
Other Road Users	79%	84%	32%	64%	55%	63%	II
Pedestrians	71%	69%	42%	64%	64%	62%	II
On-Board Rattles	75%	68%	37%	67%	36%	57%	III
Passengers with Infants	76%	59%	47%	82%	18%	56%	III
Climate Controls	56%	34%	58%	52%	64%	53%	III
Reading (e.g., Route Sheet)	57%	27%	100%	45%	27%	51%	III
Passengers with Disabilities	56%	38%	47%	52%	64%	51%	III
Audible Alerts	67%	46%	21%	79%	36%	50%	III
General Broadcasts	71%	57%	5%	85%	27%	49%	IV
Personal Broadcasts	67%	48%	21%	67%	36%	48%	IV
Driver Mobile Phone	64%	28%	26%	52%	27%	39%	IV
Advertisements/Water Activity	51%	24%	16%	48%	18%	31%	IV
Others (please specify)	20%	6%	5%	27%	9%	13%	IV

Modeling Distracting Activities

The research hypotheses were to determine the likelihood that the public transit bus driver getting slightly distracted, distracted, and very distracted with respect to not distracted is related to his/her demographic background, driving hours/week, and location. The MLR model was applied to eight distracting activities classified in Risk Zone I and II (Figure 1) to identify the factors having significant impact on the levels of distraction. The MLR model provides estimates of the sign of the independent variable coefficients and their magnitude relative to one another, the odds ratios (Yan et al. 2009; Morfoulaki et al. 2007; Peng and Nichols 2003), and the probability of occurrence of a distraction level relative to a reference distraction level (Field 2009; Yan et al. 2009).

The categorical dependent variable has four levels: Not Distracted, Slightly Distracted, Distracted, and Very Distracted. The independent variables included categorical variable—gender and location—and continuous variables—age, driving experience, and driving hours per week. This concept of categorizing distraction is similar to the one used by Morfoulaki et al. (2007) 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.

The MLR model for highly-risky distracting activities in Risk Zone I and II was converted into binary logistic regression models, with each response variable level compared to a reference level (Moutinho and Hutcheson 2011). Hence, four distracting levels ($k = 4$) produced three ($k - 1$) binary logistic regression models for each distracting activity. The general MLR model proposed by Moutinho and Hutcheson (2011) is expressed as:

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

Where,

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

X_1, X_2, \dots, X_k are the categorical or continuous independent variables.

β_0 is the Y axis intercept (constant term) and $\beta_1, \beta_2, \dots, \beta_k$ are the common slope coefficients.

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

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

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

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

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

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

The logit models 2, 3, and 4 provide three estimates of 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.

For modeling the transit bus driver distraction, the multinomial dependent variable Y_i (*logit*), which measures the total contribution of the five factors (independent variables), is expressed as:

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

Where,

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

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

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

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

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

The parameters $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ are estimated by the maximum-likelihood method, which maximizes the likelihood of reproducing the data with a given parameter estimate (Field 2009; Peng and Nichols 2003).

The coefficients estimated by the MLR models are used to predict the probability of a driver getting distracted relative to the reference level (Not Distracted). In logistic regression, the dependent variable Y_i in Equation (5) is a logit, which is the natural log of the odds ratio.

Taking the antilog of the logit (Y_i) and natural log of the odds ratio results in the following binary logistic function (Field 2009):

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

Where,

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

$e = 2.71828$ is the base of natural logarithms.

A distracting activity is very unlikely to occur if $f(Y_i)$ is close to 0 and very likely to occur if it is close to 1.

Equations (2), (3), and (4) were fitted to the survey data using SPSS 17.0 (2008) to test the hypothesis stated in the Modeling Distracting Activities section. SPSS 17.0 (2008) included all the independent variables as direct entry. A stepwise (MLR) procedure then eliminated non-significant variables until a good fit was achieved with the significant variables. The MLR output is split into three tables since the categorical dependent variables are compared in pairs. The analysis was conducted for each of the eight distracting activities in Risk Zones I and II (Figure 1) to estimate the function Y_{ij} (logit) of the MLR model that best fits the data for each distracting activity. Out of eight MLR models developed for Risk Zones I and II activities, six were found to be highly significant ($p < 0.05$) and exhibited a good fit (Table 2).

Table 2. MLR Model Functions for Risk Zone I and II Activities

Activity	Slightly Distracted (2)	Distracted (3)	Very Distracted (4)
Passengers Using Mobile Phone (1)	$(Y12)^* = -105.49 - 9.48 \text{LOCAT} + 82.41\text{SEX} + 1.65\text{AGE} + 2.57\text{EXP} + 1.89\text{DRIVING/WK}$	$(Y13) = 156.58 - 5.82\text{LOCAT} + 20.06\text{SEX} - 2.72\text{AGE} + 3.67\text{EXP} - 3.79\text{DRIVING/WK}$	N/S**
Passengers (2)	$(Y22) = -2.20\text{LOCAT} + 16.05\text{SEX} + 0.13\text{DRIVING/WK}$	$(Y23) = -224.35 + 235.99\text{SEX} + 0.20\text{EXP} + 4.53\text{DRIVING/WK}$	$(Y24) = 0.47\text{DRIVING/WK}$
Fatigue/Sick (3)	N/S	N/S	$(Y34) = 137.74\text{SEX}$
Passengers Not Following Etiquette (4)	$(Y42) = -4.47\text{LOCAT} + 53.49\text{SEX}$	$(Y43) = 323.22 - 6.52\text{LOCAT} - 6.26\text{AGE} - 7.99\text{DRIVING/WK}$	$(Y44) = 152.61\text{SEX}$
Ticket Machine (5)	$(Y52) = 1050.21 - 11.51\text{LOCAT} - 68.67\text{SEX} - 20.30\text{AGE} + 23.86\text{EXP} - 26.51\text{DRIVING/WK}$	N/S	N/S
Other Road Users (6)	$(Y62) = 55.88 - 1.28\text{DRIVING/WK} - 1.04\text{AGE}$	N/S	$(Y64) = 67.26 - 2.66\text{DRIVING/WK} - 1.93\text{AGE}$
Passenger Trying to Talk To Driver (7)	N/S	N/S	N/S
Pedestrians (8)	N/S	N/S	N/S

Note: SPSS 17.0 sets the reference level Not Distracted = 0 with Slightly Distracting (2), Distracting (3), and Very Distracting (4) set = 1; Northside = 1 and Southside = 0; Male = 1 and Female = 0.

*(Y_{ij}) is the estimated function that measures the total contribution of each significant factor where, $i = 1$ to 6, $j = 2$ to 4.

** MLR Final Model or individual independent variables were not significant (N/S).

Model Assessment and Validation of Results

Model Assessment

Due to space limitation, only model assessment for “Passengers” is presented in Table 3. The likelihood ratio test using model fitting information shows that the difference in the -2Log Likelihood between the intercept only (without any independent variables) and the final model (with all the independent variables) provides the chi-square (χ^2) = 36.61 (18) signifying a good improvement in the model fit. It follows that the independent variables contribute significantly to the outcome variable. The values of the AIC initial/final values (114.22/104.16); the BIC initial/final values (145.06/140.14) get smaller during the stepwise process indicating a good fit for the final model.

The model’s Goodness of Fit as indicated by the p-values for Pearson and Deviance chi-square (χ^2) = 1.00, (p = 1), indicating 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 R2 (0.59, 0.65, and 0.32) are reasonably high values and when used as supplementary tests (Peng and Nichols 2003) also indicates a good fit. Table 3 presents the outputs from three binary logistic regression models along with the coefficients (β), Wald Statistic, and odds ratios (Exp [B]), along with the 95% Confidence Interval (CI) values.

Interpretation of MLR Results for High Risk Distracting Activities

Out of the eight distracting activities in Risk Zone I and II shown in Table 2, six significant activities were analyzed for impact of the sign and magnitude of the significant independent variable’s coefficients on the dependent variable. In the interest of space, interpretation of results for very high risk distracting activity “Passengers” is illustrated here and results for the remaining activities are summarized in Table 4.

The MLR model functions Y_{ij} (*logit*) for “Passengers” distracting levels are reproduced from Table 2 as follows:

$$\text{Slightly Distracted: } (Y_{22}) = - 2.20LOCAT + 16.05SEX+0.13DRIVING/WK \quad (7)$$

$$\text{Distracted: } (Y_{23}) = -224.35 + 235.99SEX + 0.20EXP+4.53DRIVING/WK \quad (8)$$

$$\text{Very Distracted: } (Y_{24}) = 0.47DRIVING/WK \quad (9)$$

Table 3. MLR Model Parameter Estimates for “Passengers”

Model Chi-Square (χ^2) = 36.61 (18)*** Pearson Stat (NS) Deviance Stat(NS)		R2 = 0.590 (Cox & Snell); 0.649 (Nagelkerke); 0.317 (McFadden)		AIC initial/final values: 114.22/104.16 BIC initial/final values: 145.06/140.14	
Independent Variables and Interactions	Coeff β (SE)	Wald Statistic	Odds Ratio Exp (B)	95% CI	
Slightly Distracted vs. Not Distracted					
Intercept	N/S	-			
LOCAT = 1	-2.20 (1.04)**	4.44	0.11	[0.14 – 0.86]	
LOCAT = 2	0.00				
SEX =1	16.05 (6.04)**	7.07	9340926	[67.82 – 1.29E12]	
SEX = 2	0.00				
AGE	N/S	-	N/A		
EXP	N/S	-	N/A		
DRIVING/WK	0.13 (0.07)*	3.64	1.14	[1.00 – 1.30]	
AGE*DRIVING/WK	N/S	-	N/A		
SEX=1*DRIVING/WK	-0.34 (0.13)****	6.87	0.71	[0.55 – 0.92]	
AGE*EXP	N/S	-	N/A		
Distracted vs. Not Distracted					
Intercept	-224.35 (6.95)****	1042.79			
LOCAT = 1	N/S	-	N/A		
LOCAT = 2	0.00				
SEX =1	235.99 (1.53)****	23736	3.08E102	[1.53E103 – 6.20E103]	
SEX = 2	0.00				
AGE	N/S	-	N/A		
EXP	0.20 (0.10)**	3.79	1.22	[1.0 – 1.48]	
DRIVING/WK	4.53 (0.10)****	1947	93.15	[76.16 – 113.94]	
AGE*DRIVING/WK	N/S	-	N/A		
SEX=1*DRIVING/WK	N/S	-	N/A		
AGE*EXP	N/S	-	N/A		
Very Distracted vs. Not Distracted					
DRIVING/WK	0.47 (0.21)**	5.00	1.6	[1.06 – 2.41]	

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

Table 4. Impacts of Independent Variable Coefficients on Risk Zone I and II Distracting Activities Relative to Not Distracting Activity

High Risk Activity	Location	Sex	Age	Driving Exp	Driving Hrs/Wk
Passengers Using Mobile Phone	Southside drivers are more likely to get Slightly Distracted, followed by Distracted.	Male drivers are more likely to get Slightly Distracted, followed by Distracted.	Older drivers are more likely to get Slightly Distracted but less likely to get Distracted.	Drivers with more driving experience are more likely to get Slightly Distracted, followed by Distracted.	Drivers doing more driving hours are more likely to get Slightly Distracted but less likely to get Distracted.
Passengers	Southside drivers are more likely to get Slightly Distracted.	Male drivers are more likely to get Distracted, followed by Slightly Distracted.	N/S	Drivers having more driving experience are more likely to get Distracted.	Drivers doing more driving hours are more likely to get Distracted, followed by Very Distracted, or Slightly Distracted.
Fatigue/Sickness	N/S	Male drivers are more likely to get Very Distracted.	N/S	N/S	N/S
Passengers Not Following Etiquette	Southside drivers are more likely to get Slightly Distracted, followed by Distracted	Male drivers are more likely to get Very Distracted, followed by Slightly Distracted	Older drivers are less likely to get Distracted.	N/S	Drivers doing more driving hours are less likely to get Distracted.
Ticket Machine	Southside drivers are more likely to get Slightly Distracted.	Female drivers are more likely to get Slightly Distracted	Older drivers are less likely to get Slightly Distracted.	Drivers having more driving experience are more likely to get Slightly Distracted.	Drivers doing more driving hours are less likely to get Slightly Distracted
Other Road Users	N/S	N/S	Older drivers are less likely to get Slightly Distracted, followed by Very Distracted.	N/S	Drivers doing more driving hours are less likely to get Slightly Distracted, followed by Very Distracted.
Passengers Trying To Talk To Driver	N/S	N/S	N/S	N/S	N/S
Pedestrians	N/S	N/S	N/S	N/S	N/S

N/S – MLR models were not significant.

Location

SPSS 17.0 (2008) coded Northside = 1 and Southside = 0 and used Not Distracted as the reference level. The negative coefficient (-2.20) associated with LOCAT in Equation (7) implies that holding all other independent variables constant, the Southside drivers were more likely than their Northside counterpart to get Slightly Distracted by passengers. The odds ratio (0.11) is < 1 , and the 95% CI does not include 1 (Table 3), which indicates less likelihood of a Northside driver getting Slightly Distracted than a Southside driver. The odds of a Northside driver getting Slightly Distracted compared to the odds of Not Distracted by passengers are 0.11. The reciprocal of 0.11 indicates that drivers in the Northside were 9 times more likely to get “Not Distracted” than “Slightly Distracted.” The Northside had a lower population density than the Southside and, hence, fewer passengers, which could possibly lead to less passenger distraction.

Gender

SPSS 17.0 (2008) coded Male = 1 and Female = 0 and used Not Distracted as the reference level. The positive coefficients (16.00 and 235.99) associated with SEX in Equations (7) and (8) implies that holding all other independent variables constant, male drivers were more likely than female drivers to get Distracted, followed by Slightly Distracted by passengers. The estimation of separate coefficients for both functions indicates that gender is not considered equally in the functions (Washington et al. 2011), with Distracted having the highest likelihood, followed by Slightly Distracted. The odds ratios are > 1 , and the 95% CI does not include 1 (Table 3), which indicates greater likelihood of a male driver getting Distracted, followed by Slightly Distracted than female drivers. The odds for the male drivers getting Distracted to Slightly Distracted compared to the odds of Not Distracted by passengers are very high values (Table 3).

Driving Hours/Week

SPSS 17.0 (2008) used Not Distracted as the reference level. The positive coefficients (0.13, 4.53, and 0.47) associated with DRIVING/WK in Equations (7), (8), and (9) implies that, holding all other independent variables constant, the higher the driving hours/week, the more likely the driver would get Distracted, followed by Very Distracted and Slightly Distracted. The estimation of separate coefficients for the three functions indicates that driving hours/week is not considered equally in the functions with Distracted having the highest likelihood, followed by Very Distracted and Slightly Distracted. The odds ratios are > 1 for all three functions, and the 95% CI does not include 1 for Distracted and Very Distracted, but for Slightly

Distracted the lower limit of the CI is 1; hence, it was considered as “not including” 1 (Table 3). If a driver increases his/her driving hours/week by one hour, the relative risk of getting Slightly Distracted, Distracted, and Very Distracted relative to Not Distracted would increase by 1.14, 93.15, and 1.60 times, respectively, given the other independent variables are held constant.

Driving Experience

SPSS 17.0 (2008) used Not Distracted as the reference level. The positive coefficient (0.20) associated with EXP in Equations (8) implies that, holding all other independent variables constant, the higher the driving experience, the more likely the a driver would get Distracted. The odds ratios are > 1 , but the lower limit of the CI is 1, i.e., the odds ratio will be significant at any confidence level of $\alpha > 0.05$ (Table 3). Hence, for this analysis it is considered significant. If a driver increases his/her driving experience by one year, the risk of getting Distracted relative to Not Distracted would increase by 1.22 times, given the other independent variables are held constant. This appears contrary to popular belief, where experience made a driver better at handling distraction. There was a significance difference in driving experience in both locations, with drivers on the Southside having more driving experience (15 years) as compared to the Northside drivers (8 years). It must be noted that the more experienced Southside drivers also have a high accident rate.

The above interpretation for “Passengers” covered significant independent variables LOCAT, SEX, DRIVING/WK, and EXP. The variable AGE was not significant for “Passengers” but is an important factor related to accidents, with younger drivers more prone to distracted driving and accidents (U. S. DOT 2009). This study reveals the positive and negative impact of age on other distraction activities, which are discussed in D’Souza and Maheshwari (2012a and 2012b).

The impact of coefficients of all MLR functions listed in Table 2 is summarized in Table 4. Following the approach of Washington et al. (2011), for functions having separate coefficients, interpretation is provided only for the largest positive or smallest negative coefficient.

Validation of MLR Results

The MLR model functions for the Risk Zones I and II activities provide estimates of the current levels of distraction at the transit agency. How effective are these functions, and how can one validate these results? Are the results generated from the MLR models for each risk zone activity valid for a large random population of transit bus drivers? Simulating the models using probabilistic distributions generates

driver distraction events that would occur in practice over a range of random factors. Monte Carlo simulation using discrete distribution that incorporates random variability into the model was applied to validate the results. Each MLR function in Table 2 was simulated for a large sample of drivers with randomly selected location, age, sex, driving experience, and driving hours/week.

Following the approach of Smith et al. (2005), these five independent variables were simulated for 1,000 replications one at a time, keeping the remaining variables constant. The impact of independent variables' coefficients (Table 4) were validated by comparison with the simulated probability values. For each Risk Zone I and II activity, the simulation model generated average probability values from Equation (6) for 1,000 drivers getting Slightly Distracted, Distracted, and Very Distracted by the distraction factors. The results for location, gender, and driving hours/week related to distracting activity "Passengers" are presented as follows.

Location

The MLR model results presented in Table 4 indicate that the Southside drivers have a higher likelihood of getting Slightly Distracted and Distracted. The simulation output (Figure 2) validates these results for all the passenger-related activities. The probability values for Southside drivers are higher compared to Northside drivers.

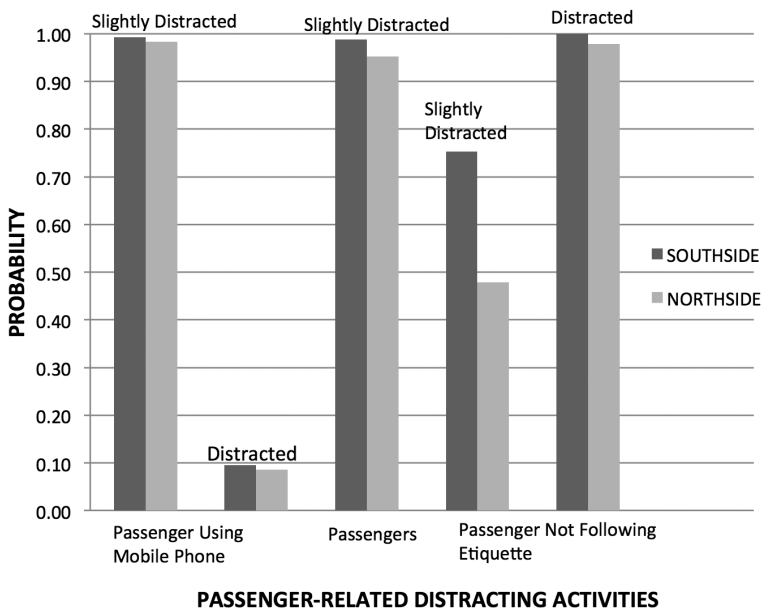


Figure 2. Simulation Results for Location

Gender

The MLR model results indicate that male drivers have a higher likelihood of getting Distracted, followed by Slightly Distracted. The simulation output (Figure 3) validates these results for all the passenger-related activities. The probability values for male drivers are higher compared to female drivers.

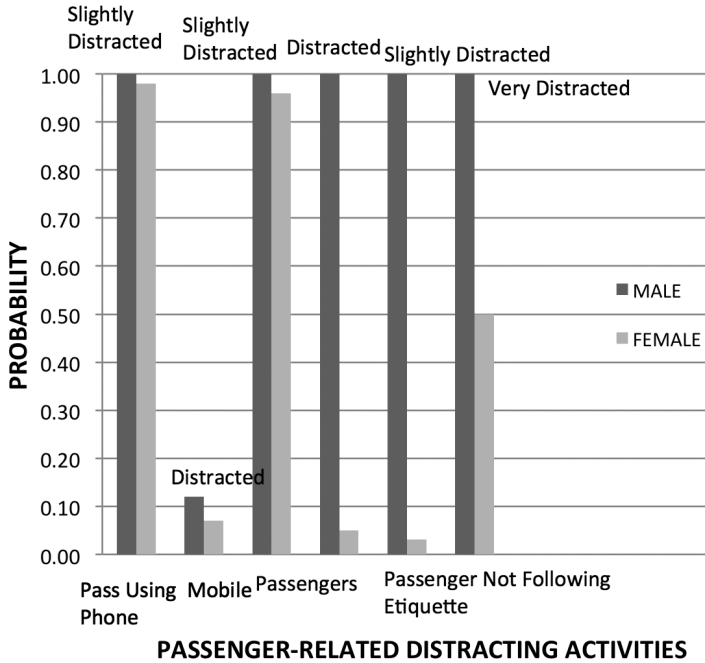


Figure 3. Simulation Results for Gender

Driving Hours/Week

The MLR model results indicate that drivers with more driving hours/week have a higher likelihood of getting Distracted, followed by Very Distracted and Slightly Distracted. The simulation output (Figure 4) for passenger-related activities validates these results with the exception of Passengers Using Mobile Phones (Distracted) and Passenger (Very Distracted).

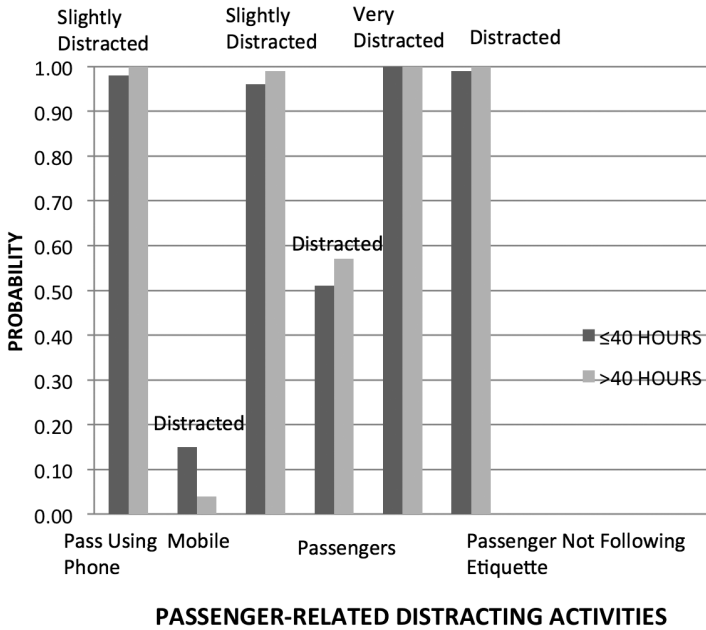


Figure 4. Simulation Results for Driving Hours/Week

Comparison of predicted results generated by the MLR model with simulated outputs for Risk Zones I and II activities show similarities for two-thirds of the model results (D’Souza and Maheshwari 2012a and 2012b). Simulation and predicted results match for certain distraction activities like Passengers Using Mobile Phone and Climate Control for all independent variables namely sex, age, location, experience and driving per week. Two other distraction activities— Passengers Not Following Etiquette and Passengers—match simulation and predicted results for all but one independent variable. However, distraction activity Ticket Machine does not show any convergence between predicted and simulated results for any independent variable. It is possible that survey respondents have confounded between two distraction activities: Ticket Machine and Passengers. In practice, most of the ticket machine distraction can be attributed to passengers. Therefore, those two factors can easily be confounded. Thus, some degree of disagreement between simulation and statistical prediction is possible. A larger sample and better explanation of survey to the respondents should improve result convergence between statistical model and simulation.

Route Observations

Eight route observations covering duration of 15 hours were conducted in the Northside and Southside in winter 2011 and spring 2012 during the late mornings or early afternoon hours are summarized in Table 5. Passengers Using Mobile Phone and Passengers were classified as Risk Zone 1 distracting activities. During the route observations, very few passengers were talking on mobile phones but were not loud enough to be heard by others, including the driver. However, Passengers distraction was observed several times. For example, passengers were observed standing in the “no-go” designated area and the bus driver did not tell them to move back behind the no-standing zone. In certain situations, passengers were standing next to the driver’s cab and talking continuously to the driver, causing distractions.

Passengers Trying to Talk to the Driver, Ticket Machine, Fatigue/Sickness, Passengers Not Following Etiquette, Other Road Users, and Pedestrians were classified under Risk Zone II. Talking to the driver was a common distraction but often unavoidable, since less route information was available at the bus station or on the bus compelling passengers to ask the drivers questions. On some routes, passengers talked (personal) continuously with the drivers, causing distraction. Passengers talking briefly with the driver when necessary, for example, during an emergency is permitted by the transit agency, but personal chatting is prohibited. Passengers trying to talk to the driver did not cause visual and physical distractions, but it certainly caused cognitive distraction since it took away the driver’s mind from driving. The ticket machine did not appear to cause distractions since it was used by passengers when entering the bus at a stop, except in one instance when a driver was observed operating the ticket machine while pulling away from a bus stop.

The atmosphere in the bus was quiet during non-peak hours but noisy when filled to capacity. On long routes, driver fatigue and restlessness were observed, causing one of the largest apparent distractions for male and female drivers. This is contrary to the results of the MLR model and simulation, which indicated that male drivers get very distracted with fatigue/sickness.

Personal Broadcast (PB) was classified under Risk Zone IV but appeared the most distracting to the drivers. On one route, the driver received three calls/hour on the PB, which took away the attention of the driver from the road. Other distracting activities not included in the study such as unusual sounds emitted from the dashboard, driver seat readjustment, etc., will require further analysis.

Table 5. Summary of Route Observations

Distracting Activity for Zone I And II	MLR Model and Simulation Results	Route Observations
Passengers Using Mobile Phones (I)	This distraction was impacted by Location, Gender, Age, Driving Experience, and Driving Hours/Week.	Many passengers were using mobile phones but only a few could be heard at the front by the driver.
Passengers (I)	This distraction was impacted by Location, Gender, Driving Experience, and Driving Hours/Weeks.	Passengers talking to other passengers while standing in the “no-standing zone” next to the driver’s cab.
Passengers Talking to Driver (II)	No significant variables.	Passengers were continuously talking to driver mostly asking for information and making personal talk.
Ticket Machine (II)	This distraction was impacted by Location, Gender, Driving Experience, and Driving Hours/Weeks.	Ticket machine was operated during stops and was a distraction when passengers did not have the correct change ready when boarding.
Fatigue/Sickness (II)	This distraction was impacted by Gender of the driver.	Fatigue and restlessness were observed on long routes for male and female drivers.
Passengers Not Following Etiquette (II)	This distraction was impacted by Location, Gender, Age, and Driving Hours/Weeks.	Passengers were noisy when the bus was filled to capacity. Use of profane language was observed during one route.
Other Road Users (II)	This distraction was impacted by Age, Driving Hours/Weeks.	No distraction observed.
Pedestrians (II)	No significant factors.	No distraction observed.
Other Distracting Activities	None	<ol style="list-style-type: none"> 1. High pitch buzzing sound from bus dashboard. 2. Driver was required to write while driving. 3. Driver’s back rest required constant adjustments.

Conclusions and Limitations

The results of the MLR models, simulation, and route observations indicate that passenger-related activities classified under Zones I and II are most distracting to the driver. The results of the MLR model and simulation do show that passengers using mobile phones caused the highest distraction, which contradicts the route observations. The insufficient number of route observations conducted mostly during the non-peak hours did not identify this distraction. Also, the cognitive

type of distraction caused by passengers using mobile phones is more challenging to detect by the route observer. A larger sample of route observations during peak hours (early morning and late afternoon) could possibly detect this 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 mobile phones in the transit system. Although PB appeared to be distracting to the route observer, it was classified in Risk Zone IV (Low Risk). The drivers did not perceive PB to be distracting since it is considered a part of the driving tasks. Also, the route observers' understanding of distraction may differ from the driver's understanding, especially for cognitive distractions.

It is a challenge for a 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 behavior etc. Providing route maps in the bus and at the stops would reduce talk between passenger and driver. Personal use of electronic devices could be allowed 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 passengers to stand next to the driver's cab to avoid unnecessary communications with passengers, and appropriate sign boards need to be posted in the bus. If conversation cannot be avoided, it must be done cautiously while driving or when the bus is stopped.

The design of ticket machines, control panels, and other devices must be user-friendly and not require long glances away from the roadway. An educational training program on the proper use of technological devices mounted in the cab or issued to the driver and hazards associated with using these devices while driving should focus on drivers who are likely to be distracted by these devices.

How could the transit agency use the MLR models developed in this study? They could be applied to predict distraction for varying driver demographic backgrounds and driving patterns. It is observed that drivers are affected differently by distracting activities, which possibly could be 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. Furthermore,

other transit agencies could use this study as a framework for conducting similar distraction analysis of their drivers (D'Souza et al. 2012).

The sample size of 48 drivers amounts to a 9.6:1 ratio for the number of cases to independent variables, which is near the minimum case-to-variable ratio of 10:1 but below the preferred case-to-variable ratio of 20:1 favored by researchers (Petrucci 2009). The literature has not recommended any specific approach for computing the sample size. In addition to the five independent variables, other variables such as environmental, vehicle, roadway designs, etc., which could also have an impact on driver distraction (Washington et al. 2011; Morfoulaki et al. 2007; Yan et al. 2009; Gkritza et al. 2006), need to be included in the MLR model.

The presence of multicollinearity that was detected from the Pearson correlation analysis makes it difficult to determine the relative importance of each independent variable on the MLR model and the effects on the dependent variable (Washington et al. 2011). The increase in sample size may reduce the standard error, thereby mitigating the threat of multicollinearity. Furthermore, due to the exploratory nature of the study and the small sample size, the MLR data were not tested to show that they meet the Independence of Irrelevant Alternative (IIA) specifications, which require the ratio of probabilities of selecting any two alternatives to be independent of the third choice (Small and Hsiao 1985). This test is planned for future studies covering larger samples of bus drivers.

This paper is one of only a few to include the full range of distractions and associated risks in public transit buses. The results support the hypothesis that the likelihood of driver distraction is related to his/her demographics, driving hours/week, and location in two-thirds of cases. But it has limited applications due to the localized sample size and limitations discussed above; hence, discrepancies need to be followed up with fresh inputs from an expanded study covering a larger sample of drivers from other agencies. The expanded data set, thus, can be used for validation as well as further refinement of the proposed models. Knowing the activities that cause a high risk of distraction as well as the responsible factors may provide additional input to law and policy makers while crafting legislation and regulations statewide or nationwide.

References

D'Souza, K. A, S. K. Maheshwari, and Z. Banaszak. 2012. Research framework for studying driver distraction on Polish city highways. To appear in *Journal of*

Foundations of Management, Faculty of Management, Warsaw University of Technology, Warsaw, Poland.

- D'Souza, K. A., and S. K. Maheshwari. 2012a. Improving performance of public transit buses by minimizing driver distraction. *Proceedings of the Urban Transport 2012 Conference*, A Coruña, Spain, May 13-16, 2012. Editors: J.W.S. Longhurst and C.A. Brebbia. Wessex Institute of Technology Press, Southhampton, U. K: 281–293.
- D'Souza, K. A., and S. K. Maheshwari. 2012b. An analysis of transit bus driver distraction using multinomial logistic regression models. *On-line Proceedings of the Journal of the Transportation Research Forum*, Tampa, FL, March 15–17. www.trforum.org/forum/2012.
- Field, A. 2009. Logistic regression. In *Discovering Statistics Using SPSS*, 3rd Edition. SAGE Publications Ltd.: 264–315.
- 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.
- 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.
- Moutinho, L., and G. Hutcheson. 2011. Multinomial logistic regression. In *The SAGE Dictionary of Quantitative Management Research*. SAGE Publications Ltd.: 208–212.
- Peng, C. Y., and R. N. Nichols. 2003. Using multinomial logistic models to predict adolescent behavioral risk. *Journal of Modern Applied Statistical Methods* 2(1): 1–13.
- Petrucci, Carrie J. 2009. A primer for social worker researchers on how to conduct a multinomial logistic regression. *Journal of Social Service Research* 35(2): 193–205.
- Salmon, P. M., K. L. Young, and M. A. Regan. 2011. Distraction “on the buses”: A novel framework of ergonomics methods for identifying sources and effects of bus driver distraction. *Applied Ergonomics* 42: 602–610.

- Salmon, P. M., K. L. Young, and M. A. Regan. 2006. Bus driver distraction stage 1: Analysis of risk for State Transit Authority New South Wales bus drivers. Final Report. Monash University Accident Research Center, Victoria, Australia.
- Small, K. A. and C. Hsiao. 1985. Multinomial logit specification tests. *International Economic Review* 26(3): 619–627.
- Smith, D. I., J. Chang, D. Cohen, J. Foley, and R. Glassco. 2005. A simulation approach for evaluating the relative safety impact of driver distraction during secondary tasks. 12th World Congress on ITS, San Francisco, November 6–10, Paper 4340.
- SPSS 17.0. 2008. SPSS Inc., Chicago, IL.
- U. S. Department of Transportation. 2009. An examination of driver distraction as recorded in NHSTA Databases. DOT HS 811 216.
- U. S. Department of Transportation. 2010. Statistics and facts about distracted driving. www.distraction.gov/stats-and-facts.
- 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.
- Yan, X., E. Radwan, and K. K. Mannila. 2009. Analysis of Truck-Involved Rear-End Crashes Using Multinomial Logistic Regression. *Advances in Transportation Studies an International Journal* Section A 17: 39 – 52.
- Yang, C. Y. D. 2007. Trends in Transit Bus Accidents and Promising Collision Countermeasures. *Journal of Public Transportation* 10 (3): 119 – 136.

Acknowledgments/Disclaimer

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This research was funded by the U. S. Department of Transportation University Transportation Centers Program: DTRT06-G-0029. The U.S. Government or Hampton Roads Transit (HRT) assumes no liability for the contents or use thereof.

About the Authors

KELWYN A. D'SOUZA (*kelwyn.dsouza@hamptonu.edu*) is a Professor of Management Science in the School of Business at Hampton University. He received his Ph.D. in Industrial Engineering from the University of South Florida in December 1991 and has 10 years of prior industrial experience. His research interests include modeling and analysis of transportation systems, transportation safety, and logistics. He has been associated with transportation research and training activities for over 15 years and has been awarded ~ \$2.5 million from USDOT in the past 5 years for conducting transportation-related research and educational programs. He is the founding Director of the Tier II University Transportation Center (UTC) at Hampton University. He has published in refereed journals and conference proceedings and is the University's contact person for transportation.

SHARAD K. MAHESHWARI (*sharad.maheshwari@hamptonu.edu*) is an Associate Professor of Management at Hampton University and Associate Director for Technology Transfer for the Hampton University UTC. He received his Ph.D. in Industrial Engineering from the University of South Florida in May 1992. His current research and teaching interests are in business research, quality management, transportation safety management, distracted driving, and business education. He has published 25 articles in the refereed journals, including the *International Journal of Production Research*, *International Journal of Computer Integrated Manufacturing*, and has presented more than 35 articles in professional conferences. He is the editor of *Academy of Information and Management Science Journal* and *Journal of Business Research* and a recipient of seven best paper awards. He is a registered Professional Engineer in the Commonwealth of Virginia and a Certified Quality Engineer (CQE) from the American Society of Quality Control.