CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

COMPARATIVE ASSESSMENT OF VARIOUS RANDOM PARAMETER ORDERED MODELS: A COMPREHENSIVE EVALUATION OF WORK ZONE COLLISION INJURY-SEVERITIES

Nahidi, S.A.[1,4]
[1] University of Waterloo, Canada
[4] snahidi@uwaterloo.ca

Tighe, S.L.[1,4]
[1] University of Waterloo, Canada
[4] sltighe@uwaterloo.ca

Abstract: Maintaining driver and worker safety on highways experiencing closures during construction, maintenance, and rehabilitation activities is crucial. Throughout the years, several studies have been conducted to identify the influences of various factors on the injury-severity level of the collisions occurring on highways. Developing statistical models can help identify the factors which significantly influence the injury-severity level. The clear and practical results of these models could be adopted by the contractors to improve the level of safety in the work zones. In recent years, several complicated and advanced statistical models were applied to injury-severity data; these models are often time inefficient, impractical, and hard to interpret by non-statistical experts. Also, these models are intended to have the issue of overfitting which limits the ability of the models to predict future events. This study collected historical data from four different US states (New York, Pennsylvania, Illinois, and Michigan), between 2013 and 2016, to develop an injury-severity level statistical model. Selected states have similar weather condition, pavement condition, and construction policies and regulations as Ontario, Canada. The authors believe due to the stated similarities between these US states and Ontario, the statistical models developed will be spatially transferable. Therefore, this paper aims to apply the random parameter concept to some of the well-known fixed parameter statistical models to overcome issues from both unnecessarily complicated models and statistically insufficient methodologies. Random parameter ordered Probit, random parameter ordered Logit, and random parameter ordered arctangent models will be developed to address stated issues. Then, a comparative assessment among all ordered models will be conducted to investigate which one of these models has statistical dominance. Finally, these models will be used to develop an Excel-based system which can be adapted by non-statistical experts to operate, understand, and plan. The Excel-based system will interpret the effect of each significant factor as well as predicting the possible severity level of future collisions in the work zones. This study intends to present a straightforward methodology, which addresses previous concerns in this field.

1 INTRODUCTION

In the last decade, many researchers have attempted to investigate the factors and causes of the crashes in highway work zones. Based on the 2016 Transport Canada Statistics study: 1898 fatalities, 10,322 major injury collisions, and 160,315 minor injury crashes occurred in work zones on Canadian highways. As such work zones can be considered high-risk locations. In a 2007 report by the same organization concerning the social costs of motor vehicle collisions, highway traffic delays, out-of-pocket expenses, hospital/health care, tow trucks, and Police, Fire and Ambulance Services were the main contributing factors. Using these factors, the social cost in Ontario for motor vehicle collisions was estimated to be approximately 18 billion dollars. Among all other injury-severity levels, fatalities had the highest share (11 billion dollars). With growth of the traffic volume on Ontario roadways, it is expected that the needs for setting up the work zones to maintain the serviceability of the highways will also increase. The nature of the work zones (i.e. narrower lanes and temporary changes in the geometry of the highways) may make these sections more vulnerable to collisions.

In a study by Khattak et al. (2002), it was concluded that the collision rate generally increases in the high-risk work zone sections of highways. Turochy et al. (2018) conducted a study to investigate most influential factors on work zone crashes. They concluded that evening and overnight closures, various manners of collision (head-on, rollover, and angle crashes), and excessive speed significantly affected the injury-severity level of the collisions occurring in the work zones of Alabama. Meng and Weng (2011) focused on the rear-end crashes in work zones and they stated that rear-end crashes are more likely to occur in expressways compared to arterial roadways. In addition, they determined that truck percentage and lane traffic flow are the most significant factors in work zone related crashes. Yang et al. (2015) found that the length of the work zones as well as the traffic volume increases the likelihood of crash occurrence in work zones. Yang et al. (2015) also claimed that although there were frequent attempts to determine a comprehensive relationship between work zone crashes and time, weather, and traffic control devices, a more meaningful relationship cannot be determined based on the current findings.

Crash records could possibly present useful information about the consequences of collisions and to identify crash severity. In previous literature, there has been very little agreement on severity of the crashes and their relationship with work zones. Some researchers determined that there is no significant difference between work zone crashes and regular crashes in terms of severity (Hall & Lorenz 1989; Chambless et al. 2002). However, there were some cases in which work zone crashes were found to be more severe than non-work zone crashes (Garber & Zhao 2002; Daniel et al. 2000). Schrock et al. (2004) determined that almost 6 percent of the 77 fatal crashes that occurred in the Texas were directly caused by work zones, and 39 percent were indirectly caused by work zones. Opposingly, based a survey completed by Benekohal and Shim (1999), the majority of 930 truck drivers surveyed perceived that driving through work zones is more hazardous than non-work zone sections of the highways.

Lin et al. (2004) explored the influence of the speed limit control on the safety of highway work zones. The interesting finding of this research was that not only lower speeds improve the safety of the work zones, but also reduction in the variance of the speed circuitously improves the safety.

Mohan and Gautam (2002) categorized the work zone collisions into two groups: worker-based and motorist-based collisions. Based on their analysis they claimed that the first group of collisions, where construction workers experienced an incident, accounted for 30% of all collisions, and the second group of collisions are those that involve motorists with remaining 70%. They also determined that driver error is the main reason of these collisions. An earlier study by Bryden and Andrew (1999) deduced that 15 percent of all serious injuries and over 40% of the fatal crashes involved pedestrian workers. Factors that significantly affected the injury-severity of the crashes occurred in the work zone were identified by Wong et al. (2011). They categorized these factors as location of the accident or work zone, duration of the work zone, time of day and type of activity in the work zone. Another factor that researchers believe that could possibly decrease the crash frequency is the use of traffic control devices in work zones (Bai and li 2011). Although these studies emphasized that using traffic control devices reduce the crash frequency, in some cases it was found that they can also be hazardous for drivers, as well as pedestrians and workers (Khattak et al. 2002). This demonstrates that there is not a clear view which constitutes the effective use of these devices, and further studies could be undertaken to draw more solid conclusions about their effects (Bligh et al. 1998; Bryden et al.1998). Hall and Lorenz (1989) analyzed the before and during-construction period of more than 1500 construction zones and they captured that rear-end collisions significantly increased due to construction zones. They recognized that almost 36 percent of the collisions in the work zones were rear-end. However, they also concluded that other crash types were less affected by these zones.

Mannering and Bhat (2014) conducted a detailed review on the various statistical modelling approaches in transportation safety analysis. Their review describes the chronological attempts of researchers in offering the most comprehensive solutions for the road safety concerns. Researchers such as Maycock and Hall (1984) were pioneers in applying count data models such as Poisson and negative binomial regressions to model accident frequency. Around the 1990s, researchers started adapting more advanced statistical models on crash data. They recognized that there were several segments in the roadway (or work zones) without any recorded collisions; this initiated the use of zero-inflated Poisson and negative binomial models. Miaou (1994) investigated the relationship between geometric design of the road segments and truck related crashes. He also conducted a comprehensive comparison between the results of zero-inflated Poisson and negative binomial models. In last two decades, more innovative methods with more complicated statistical methodologies have been examined. Bhat et al. (2014) developed a new method that accounts for endogenous covariates in count data models. More reliable analytical approximation, in addition to ability of evaluating asymptotic standard errors were the most highlighted results of his study. Chen et al. (2019) investigated the effect of the road geometry factors on various road classes by using bivariate negative binomial models. Shibata and Fukuda (1994) conducted an unconditional multiple logistic regression analysis to investigate the risk factors in fatal crashes. They found that motorcyclist helmet use, using seat belts, and low alcohol level can significantly avoid fatalities in collisions. This study offered a new standpoint toward solving crash related concerns, and more advanced crash injury severity methodologies have been introduced since then. Several subsets of Logit/ Probit models such as Multinomial Logit (Shankar and Mannering 1996; Ye and Lord 2014; Nahidi et al. 2017), Nested Logit (Chang and Mannering 1998; Yasmin and Eluru 2013), Sequential Logit/ Probit (Jung et al. 2010; Xu et al. 2013) have been introduced to crash-injury severity modelling. In late 2000s, random parameter modelling was instigated to eliminate the unobserved heterogeneity issue. Eluru and Bhat (2007), as well as Anastasopoulos and Mannering (2009) were the innovators who applied random parameter method on crash-injury severity and accident frequency models, respectively.

2 METHODOLOGY

2.1 Ordered Probability Models

There are many cases in the transportation safety field where researchers have used discrete ordered data as a dependent variable. Results of questionnaires with qualitative scales, hierarchical opinions (such as agree, neutral, disagree), or categorical frequency data (such as KABCO classifications) are some examples for these types of data. In the mid-70s, ordered probability models were introduced in transportation field; the main idea of this methodology depends on a variable referred as z, which estimates the ordinal ranking of the data. This unobserved variable (z) could be formulized as Equation 1:

[1] $z = \beta X + \varepsilon$

In this equation X represents the vector of significant independent variables, β is the set of estimated coefficients for each independent variable, and ε is the random disturbance (Washington et al. 2010).

In Equation 2, the dependent variable (y) is categorized as 1,2, 3, ..., k; predictions can be made by comparing the boundary limits (thresholds) with z,

[2] $y = 1 \;\; if\; z \le \mu_0$

$y = 2 \;\; if\; \mu_0 < z \le \mu_1$

$\dots$

$y = k \;\; if\; z \le \mu_{k-1}$

The probability of each category could be calculated as Equation 3,

$$[3]\ P(y = 1) = \mathrm{F}(-\beta X)$$
$$P(y = 2) = \mathrm{F}(\mu_1 - \beta X) - \mathrm{F}(-\beta X)$$
$$P(y = 3) = F(\mu_2 - \beta X) - \mathrm{F}(\mu_1 - \beta X)$$
$$\ldots$$
$$P(y = k) = 1 - \mathrm{F}(\mu_{k-1} - \beta X)$$

$\mathrm{F(z)}$ is the cumulative distribution function. Any changes to type of the cumulative distribution function., e.g. Probit, Logit, and arctangent, could alter the estimation procedure and affect the final outcomes. Note that the first threshold in these types of estimations are generally considered as zero. Consequently, Equation 4 is the general equation for calculating the probabilities:

$$[4]\ P(y = i) = \mathrm{F}(\mu_i - \beta X) - \mathrm{F}(\mu_{i+1} - \beta X)$$

Also, based on the underlying mathematical structure, the likelihood and log-likelihood functions could be formed as Equation 5 and 6, respectively,

$$[5]\ L(y|\beta_1, \ldots, \beta_s, \mu_1, \ldots, \mu_{k-1}) = \prod_{n=1}^{N} \prod_{i=1}^{k} [\mathrm{F}(\mu_i - \beta X_n) - \mathrm{F}(\mu_{i+1} - \beta X_n)]^{\delta_{in}}$$

$$[6]\ LL(y|\beta_1, \ldots, \beta_s, \mu_1, \ldots, \mu_{k-1}) = \sum_{n=1}^{N} \sum_{i=1}^{k} \delta_{in}\, LN[\mathrm{F}(\mu_i - \beta X_n) - \mathrm{F}(\mu_{i+1} - \beta X_n)]$$

N is the number of observations and δ_{in} is equal to 1 for observation n if it the observed outcome is i, and it is equal to 0 otherwise. The most crucial assumption through estimation procedure is $0 \leq \mu_1 \leq \cdots \leq \mu_{k-1}$. One of the concerns associated with ordered probability models is that βs cannot give a full picture of their effect on each category. The positive sign of the βs shows that it increases the probability of the very high category and reduces the probability of first category, but it doesn't represent details about the changes in intermediate categories. To have better insight about effect of each β on the interior categories, it is recommended to calculate the marginal effect, which could be estimated as Equation 7 (Washington et al. 2010),

$$[7]\ \frac{\partial P(y = i)}{\partial X} = [f(\mu_i - \beta X) - f(\mu_{i+1} - \beta X)]\acute{\beta}$$

where $f(z)$ represents the standard density function.

As previously mentioned, several $\mathrm{F(z)}$ cumulative functions could be adapted for probability estimation. Three of the cumulative density functions that are used in this study are Probit, Logit, and arctangent; their respective $f(z)$'s and $F(z)$'s are presented in Table 1. In the ordered Logit model, it is assumed that ε has a standard logistic distribution instead of normal distribution which is the assumption of Probit model. The ordered arctangent model is generally used to capture any asymmetric distributions (Greene 2002).

2.2 Random Parameters

Random parameter modelling is a robust estimation method to eliminate the unobserved heterogeneity issue. The underlying formulation of the random parameter model is based on conditional probability. Contrarily to the fixed parameter models which estimates one coefficient for each factor, in random parameter modelling there is the luxury to estimate a set of coefficients for each factor which are changing across the observations.

General formulation of random parameter modelling is shown in Equation 14,

$$[14]\ g(y_{it}|x_{it}, z_i) = f(y_{it}, x_{it}, z_i, a_i)$$

where f is the density function for the ith individual observed dependent variable at time t, $y_{it}$is the ith individual dependent variable at time t, x_{it}, z_i are measured covariates, and a_i is a factor specific parameter vector which alters randomly across each observations, with mean a and covariance matrix Ω.

Table 1: List of Probability Density Functions and Cumulative Density Functions[1]

Probit	Logit	Arctangent
[8] $f(z) = \frac{1}{\sqrt{2\pi}} e^{-[\frac{x^2}{2}]} = \phi(z)$	[10] $f(z) = \Lambda(z)[1 - \Lambda(z)]$	[12] $f(z) = \frac{2}{\pi} * \frac{1}{1+z^2}$
[9] $F(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-[\frac{x^2}{2}]} dx = \Phi(z)$	[11] $F(z) = \frac{e^z}{1+e^z} = \Lambda(z)$	[13] $F(z) = \frac{2}{\pi} arctan(z) = G(z)$

[1] $f(z)$ and $F(z)$ are representing Probability density function and Cumulative Density Function, respectively.

Random parameter structure could be revised based on the model structure. For ordered probability models, random parameter formulation is,

[15] $Probability[y_{it} = j \,|x_{it}, \beta_i] = F\left(y_{it} = j, \mu, \beta'_i x_{it} + \ a_i\right), i = 1, 2, \ldots, N, t = 1, 2, \ldots, T_i.$

where F could be any of mentioned distributions (Logit, Probit, arctangent). The assumption is that parameters are randomly distributed with heterogenous parameters generated by Equation 16,

[16] $E[\beta_i | z_i] = \beta + \Delta z$

Finally, the model could generate the coefficients with use of Equation 17,

[17] $\beta_i = \beta + \Delta z_i + \Gamma v_i$

where β is the fixed constant terms in the means of the distributions for the random parameters, Δ is the coefficient matrix, z_i is a set of observed variables which are not altered by time and the means of the random parameters, Γ is a lower triangular matrix which produces the covariance matrix of the random parameters, and finally v_i is the unobservable latent random term in the ith observation in β_i.

3 DATA

National Highway Traffic Safety Administration (NHTSA) aimed to present a detailed dataset in the Fatality Analysis Reporting System (FARS) encyclopedia. Data used in this study is a subset of the database which consists of work zone collision data from rural and urban interstate highways in New York, Indiana, Michigan, and Pennsylvania which were gathered within a 4-year period (2013-2016). The data involves 256 work zone collisions, with various injury-severity levels (property damage only or *PDO*, minor injuries, major injuries and fatalities). Each work zone accident had various sets of information such as crash, vehicle, driver, and pre-crash information. Crash related information consist of atmospheric condition, hour of the collision, day of the collision, month of the collision, heavy truck involvement, road signage, light condition and several other factors. Driver information summarizes age of the driver, drug and alcohol involvement, gender, speeding, and other person related factors. Vehicle data represents the type of vehicle, make of vehicle, age of vehicle, and safety system of the vehicle. Finally, pre-crash datasets recapped evidence related to the manner of collision, the driver's view (i.e. obstructed or unobstructed), driver distraction and so on. This dataset also contained the injury level of the people involved in the collision. Based on the dataset there were 77 property damage only collisions, 58 minor injuries, 38 major injuries, and 83 fatalities.

4 RESULTS AND DISCUSSION

Based on the developed models, all three ordered probability models had exceptional performance on predicting the injury-severity level of the collisions in the work zones. As shown in Table 2, several factors were found to be significant, however, the magnitude of each factor is not the best presentation of its actual influence on its respective crash-injury severity class. A positive value shows enhancement in probability of fatal crash occurrence, as well as, reduction in probability of PDO. A negative value represents the opposite. The marginal effect, as the representor of the effect of each factor on intermediate classes, has been estimated and the results will be used throughout the discussion.

The results illustrate that if a crash occurs between 8 PM and 6 AM, the probability of experiencing fatal crash could increase. This finding could emphasize on several issues of night time work zones such as poor visibility/ lack of vision and the increased likelihood of impaired drivers. Drivers under the age of 25 years and older than 50 are also more probable to have fatal collisions. This could be attributed to the lack of the experience for the former group and ageing or health related issues for latter. Better young driver education, as well as, stricter driver's license renewal regulations for elderly persons could possibly reduce the risks of more serious collisions and reduce this factor's effect in future models. Impaired driving also had a significant influence on injury-severity of the collisions in the work zones. As expected, it is found that not using alcohol and/or drugs reduces the probability of fatal crashes.

Because of work zones' nature, it is always recommended to have speed control in these vulnerable areas. This could significantly improve the safety for both drivers and workers. It is found in this study that speed limits less than 50 mph (or 80 km/h) could reduce the probability of fatal crash incidences. Highways with four or more than four lanes are also more probable to experience fatal crashes. Usually highways with these characteristics are separated and they accommodate higher number of vehicles with higher speeds. Due to these facts it was expected to capture higher probability of fatal crashes in these highways.

Careless driving was another crucial factor that was found to be statistically significant. Improper lane changes, failing to remain in proper lane and running off the road increase the probability of fatal crashes. The results of the models also demonstrated the importance of using traffic control devices; using these devices could desirably lead to less severe collisions. Warning signs were also found to be very influential in reducing the severity of the collisions. Rear-end crashes are also less probable to cause fatal collisions. To investigate the real impact of this factor, the marginal effect should be used; results illustrate that rear-end collision increase the probability of having PDO or minor injuries rather than fatalities and major injuries.

Three parameters were found to be significant as random parameters including: heavy truck involvement, the effect of queuing and the occurrences of vehicle roll-over. The heavy truck involvement factor increased the probability of fatal crashes in 83% of the occasions. Therefore, limiting the access of trucks could vastly improve the safety of the work zones. In remaining 17% of fatal crashes, the probability was reduced due to queuing. Lower speeds and congestions sometimes avoid fatal collisions; however, it might increase the probability of PDO and minor injuries. In addition, in 84% of the collisions with roll-over, the probability of fatal crashes increases. Surprisingly, the remaining 16% of collisions involving roll-overs only increased the probability of PDO occurrence, which might be the result of roll-overs of vehicles within the construction zones with very low speed. The study also captured that females are more probable to have fatal accidents.

Table 3 presents the prediction performance of all three developed models. Results show the high prediction performance of these models; however, Logit models shows slightly more satisfying performance over the other two models. This model predicts all PDO and minor injury collisions correctly. Only 1 in each major injury and fatality categories was predicted incorrectly. The other two models have less accurate prediction performance in major injury class, which could be due to lower number of observations of this category in comparison with other ones. Also, the Logit model has the lowest value for log-likelihood function. Therefore, it could be concluded that this model would be the most appropriate choice for this study.

One of concerns that authors were eager to address was to present the most practical way to use the developed models. An excel-based form could be designed with surveys related to the significant factors found in this study. The underlying model would start predicting the level of the crash-injury severity by inserting the corresponding values for each survey question; thus, contractors could arrange and modify their work zones by having insight about the possible injury-severity risks. Figure 1 represents a sample of the recommended excel-based form.

Table 2: Ordered Probability [Logit, Probit and Arctangent] Models Results

List of statistically significant parameters	Probit Model		Logit Model		Arctangent Model	
	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
Constant	5.561	7.140	12.311	7.280	10.458	7.070
Collision Hour (1 if collision occurred between 8PM and 6 AM, 0 otherwise)	1.348	4.320	3.551	5.660	3.089	5.650
Driver's Age (1 if it is less than 25 or more than 50, 0 otherwise)	1.172	4.710	2.518	5.170	2.148	5.110
Alcohol or Drug involvement (1 if any or both is/are not involved, 0 otherwise)	-0.997	-3.830	-2.112	-4.220	-1.831	-4.250
Speed at the time of collision (1 if it is less than 50 mph, 0 otherwise)	-1.510	-4.850	-2.870	-4.810	-2.451	-4.770
Number of lanes (1 if the highway has 4 or more lanes, 0 otherwise)	5.865	5.200	15.559	5.570	13.274	5.530
Driver related factor (1 if failing to keep in proper lane or run off road or following improperly, 0 otherwise)	1.522	3.920	4.506	5.420	3.802	5.320
Month of collision (1 if the collision occurred between April and July, 0 otherwise)	1.473	5.110	3.407	5.760	2.924	5.700
Traffic control devices (1 if used, 0 otherwise)	-2.331	-4.360	-4.631	-4.460	-4.003	-4.490
Manner of collision (1 if rear-end, 0 otherwise)	-1.449	-3.600	-1.842	-2.490	-1.534	-2.430
Work zone signage (1 if warning sign is available, 0 otherwise)	-1.660	-3.160	-4.296	-4.150	-3.762	-4.230
Heavy truck involvement (1 if it is involved, 0 otherwise)	1.903	5.140	3.838	5.370	3.337	5.370
(Std. dev. of parameter dist.– normally distributed)	*2.053*	*9.070*	*4.672*	*8.520*	*4.001*	*8.210*
Roll-over (1 if roll-over didn't occur, 0 otherwise)	-2.930	-6.760	-6.735	-7.070	-5.721	-6.870
(Std. dev. of parameter dist.– normally distributed)	*2.927*	*10.340*	*5.164*	*8.890*	*4.408*	*8.530*
Gender of the driver (1 if male, 0 if female)	-1.197	-4.410	-2.087	-4.180	-1.762	-4.110
(Std. dev. of parameter dist.– normally distributed)	*1.976*	*9.010*	*6.493*	*8.940*	*5.515*	*8.570*
Threshold (01)	2.793	8.110	6.053	7.650	5.164	7.440
Threshold (02)	4.431	10.080	9.643	8.990	8.222	8.630
Log-likelihood function	-303.201		-302.937		-302.980	
Likelihood at zero	-372.337		-372.337		-372.337	

In the example shown in the Figure 1, the underlying model estimated the value of 11.209 from the survey questions. Based on the recommended use of the Logit model, this value would be compared with thresholds presented in Table 1 for Logit model. Since 11.209 is greater than 9.643, it could be concluded that the probability of experiencing fatal collisions in that specific work zone is higher than other injury-severity categories. It is also shown in Figure 1 that in the prepared electronic form there is an option for choosing the answers from set of possible responses. These cells are directly connected to the “Prediction” cell, which is the outcome of the developed Logit model; and this design choice illustrates the final decision of this study.

Table 3: Prediction Performance of Developed Models

		Probit Model Prediction Performance			
		PDO	Minor Injury	Major Injury	Fatality
Actual	PDO	77	0	0	0
	Minor Injury	0	58	0	0
	Major Injury	0	6	31	1
	Fatality	0	0	3	80
		Logit Model Prediction Performance			
		PDO	Minor Injury	Major Injury	Fatality
Actual	PDO	77	0	0	0
	Minor Injury	0	58	0	0
	Major Injury	0	0	37	1
	Fatality	0	0	1	82
		Arctangent Model Prediction Performance			
		PDO	Minor Injury	Major Injury	Fatality
Actual	PDO	77	0	0	0
	Minor Injury	0	58	0	0
	Major Injury	0	3	33	2
	Fatality	0	0	1	82

No.	Questions	
1	What is the planned closure time?	between 8 PM and 6 AM
2	What the average age of the driver in the region?	Less th
3	Is you construction zone close to areas that might be affected by impaired drivers?	Y
4	What is the precieved speed in or around the construction zone during the work period? (speed limit at the work zone)	More than 50 mph
5	Does this section of highway has 4 or more lanes?	Yes
6	Are there any reports regarding the slippery road condition, inappropriate curves, low visibility or any other issues in or around the work zone which causes issues such as immediate lane changes, run-off the road, etc?	Yes
7	Are you planning to use traffic control devices during your closure?	Yes
8	Based on the experience of contactor, what is most probable crash type in this specific location?	Rear-end
9	Which month are you planning to set up the closure?	Between April and July
10	Are you going to use "Warning Sign" before and after your closure as described in the guidelines [Book 7]?	Yes
11	Does this section accomodates a high percentage of heavy trucks? [more than 25 %]	Yes
12	Is the area has a geomtric design deficiencies which leads to roll-over? [e.g. narrow lanes, inappropriate closures, slippery roads, etc]	Yes
13	What is the dominant driver gender in the area? [e.g. more female drivers close to shopping malls, more male drivers in rural highways]	Male
	Prediction	Fatality

Please select a value

Figure 1: Sample of the Excel-Based Form

5 CONCLUSION

The safety of the highways, especially work zones, is crucial and any negligent act that increases the probability of fatal crashes should be avoided. Previous literature emphasized that several factors such as the drivers' behaviours, geometric design of highways, work zone set up guidelines can significantly enhance the probability of harmful incidences in these high-risk areas. This study aimed to investigate the possible solutions and factors that can improve the safety of the highways. Crash data from 4 states (New York, Pennsylvania, Illinois, and Michigan) had been used to develop statistical models. It was assumed that selected states typically have similar environmental condition, pavement condition, and construction policies and regulations as Ontario, Canada. The authors believe that the random parameter ordered probability models could identify some of the crucial factors to improve road safety. The ordered Logit model presented better results in comparison with Probit and arctangent models. The analysis of several random parameter techniques was used to eliminate the unobserved heterogeneity issue with ordered probability models. As ordered probability models are moderately easy to interpret, determining a way to improve the usability of random parameter models was one of the main objectives of this study. This objective was met through the creation of an efficient and straightforward solution for non-statistical experts in the field that will help them to have a better understanding about the possible scenarios of work zone set ups and conditions.

6 REFERENCES

Anastasopoulos, P., & Mannering, F. 2009. A note on modelling vehicle accident frequencies with random-parameters count models. *Accident Analysis and Prevention, 41*(1), 153-159.

Bai, Y., & Li, Y. 2011. Determining the drivers' acceptance of EFTCD in highway work zones. *Accident Analysis and Prevention*.

Benekohal, R., & Shim, E. 1999. Multivariate Analysis of Truck Drivers' Assessment Of Work Zone Safety. *Journal of Transportation Engineering, 125*(5), 398.

Bhat, C., Born, K., Sidharthan, R., & Bhat, P. 2014. A count data model with endogenous covariates: Formulation and application to roadway crash frequency at intersections. *Analytic Methods in Accident Research*.

Bligh, R., Mak, K., & Rhodes Jr., L. 1998. Crash testing and evaluation of work zone barricades. *Transportation Research Record*.

Bryden, J., & Andrew, L. 1999. Serious and Fatal Injuries to Workers on Highway Construction Projects. *Transportation Research Record, 1657*, 42-47.

Bryden, J., Andrew, L., & Fortuniewicz, J. 1998. Work Zone Traffic Accidents Involving Traffic Control Devices, Safety Features, and Construction Operations. *Transportation Research Record: Journal of the Transportation Research Board, 1650*(1), 71-81.

Chambless, J., Ghadiali, A., Lindly, J., & McFadden, J. 2002. Multistate work-zone crash characteristics. *ITE Journal (Institute of Transportation Engineers), 72(5)*, 46-50.

Chang, L.-Y., & Mannering, F. 1998. Predicting Vehicle Occupancies from Accident Data: An Accident Severity Approach. *Transportation Research Record, 1635*, 93-104.

Daniel, J., Dixon, K., & Jared, D. 2000. Analysis of Fatal Crashes in Georgia Work Zones. *Transportation Research Record: Journal of the Transportation Research Board, 1715*, 18-23.

Eluru, N., & Bhat, C. 2007. A joint econometric analysis of seat belt use and crash-related injury severity. *Accident Analysis and Prevention, 39*(5), 1037-1049.

Garber, N., & Zhao, M. 2002. Distribution and Characteristics of Crashes at Different Work Zone Locations in Virginia. *Transportation Research Record: Journal of the Transportation Research Board, 1794*, 19-25.

Greene, W. H. 2000. Econometric analysis. Upper Saddle River, N.J: Prentice Hall.

Hall, J., & Lorenz, M. 1989. Characteristics of Construction-Zone Accidents. *Transportation Research Record: Journal of the Transportation Research Board, 1163*, 4-14.

Hall, R., & Maycock, G. 1984. Accidents at 4-arm roundabouts. *Transport and Road Research Laboratory*.

Jung, S., Qin, X., & Noyce, D. 2010. Rainfall effect on single-vehicle crash severities using polychotomous response models. *Accident Analysis and Prevention, 42*(1), 213-224.

Khattak, A., Khattak, A., & Council, F. 2002. Effects of work zone presence on injury and non-injury crashes. *Accident Analysis and Prevention, 34*(1), 19-29.

Lin, P., Kang, K., & Chang, G. 2004. Exploring the effectiveness of variable speed limit controls on highway work-zone operations. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 8*(3), 155-168.

Mannering, F., & Bhat, C. 2014. Analytic methods in accident research: Methodological frontier and future directions. *Analytic Methods in Accident Research, 1*, 1-22.

Maycock, G., & Hall, R. 1984. Accidents at 4 roundabouts. *Transportation Research Record: Journal of the Transportation Research Board*.

Meng, Q., & Weng, J. 2011. Evaluation of rear-end crash risk at work zone using work zone traffic data. *Accident Analysis and Prevention, 43*(4), 1291-1300.

Miaou, S. 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis and Prevention, 26*(4), 471-482.

Mohan, S., & Gautam, P. 2002. Cost of Highway Work Zone Injuries. *Practice Periodical on Structural Design and Construction, 7*(2), 68-73.

Nahidi, S., Fountas, G., Sarvani, S. P., Sarwar, M. T., & Anastasopoulos, P. C. 2017. Project Discrepancies in roadway construction and Preservation: a statistical analysis of Public–Private Partnership contract Types in the Us. Frontiers in Built Environment, 3, 15.

Schrock, S., Ullman, G., Cothron, A., Kraus, E., & Voigt, A. 2004. *An Analysis of Fatal Work Zone Crashes in Texas.* Texas Transportation Institute.

Shankar, V., & Mannering, F. 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research*.

Shibata, A., & Fukuda, K. 1994. Risk factors of fatality in motor vehicle traffic accidents. *Accident Analysis and Prevention, 26*(3), 391-397.

Turochy, R., Jehn, N., & LAMondia, J. 2018. Analysis of Work-Zone Crash Reports to Determine Factors Associated with Crash Severity. *Transportation Research Record.* Washington D.C., United States.

Washington, S., Karlaftis, M. G., & Mannering, F. L. 2010. Statistical and econometric methods for transportation data analysis. Boca Raton: Chapman & Hall/CRC.

Wong, J., Arico, M., & Ravani, B. 2011. Factors influencing injury severity to highway workers in work zone intrusion accidents. *Traffic Injury Prevention, 12*(1), 31-38.

Xu, C., Tarko, A., Wang, W., & Liu, P. 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis and Prevention, 57*, 30-39.

Yang, H., Ozbay, K., Ozturk, O., & Xie, K. 2015. Work Zone Safety Analysis and Modelling: A State-of-the-Art Review. *Traffic Injury Prevention*.

Yasmin, S., & Eluru, N. 2013. Evaluating alternate discrete outcome frameworks for modelling crash injury severity. *Accident Analysis and Prevention, 59*, 506-521.

Ye, F., & Lord, D. 2014. Comparing three commonly used crash severity models on sample size requirements: Multinomial logit, ordered probit and mixed logit models. *Analytic Methods in Accident Research*.