



## ANALYSIS OF CRASH SEVERITY AT INTERSECTIONS AND ROUNDABOUTS USING ORDERED AND GENERALIZED ORDERED PROBIT MODELS

Roya Eskandar Kamali<sup>1</sup>, Arash Mazaheri<sup>2</sup>, Amir Masoud Rahimi<sup>1</sup>

<sup>1</sup>University of Zanjan, Department of Civil Engineering, Zanjan, Iran

<sup>2</sup>Concordia University, Department of Building, Civil and Environmental

### Abstract

Accidents are considered as one of the leading causes of death around the world. Investigating the causes of various accidents, identifying the primary factors and accident blackspots can be used to prevent severe injuries and the associated costs and casualties. Many studies are conducted around the world to investigate the impacts of accidents and to identify the main factors. This study investigates the accidents and the severity of their injuries at intersections and squares. Accident data related to Zanjan is used to assess the severity of accidents at urban squares and intersections, which are modeled by SAS software. For this purpose, two models of OP and cumulative Probit are used. The results show that the cumulative probit model is better able to interpret the dependent variable based on the independent variables than the OP model. Based on the overall results of the models, it can be seen that the occurrence of an accident in daytime compared to nighttime, on holidays compared to non-holidays and in winter compared to other seasons, reduces the injury severity. Heavy vehicles reduce the injury severity compared to cars, while two-wheeled vehicles, including motorcycles and bicycles, increase the injury severity. Accidents at intersections also increase injury severity.

*Keywords: injury severity, intersections, squares, the OP model, crash*

### 1 Introduction

Road transport benefits both nations and individuals by facilitating goods and people's movement. The role of transport, its processes, and its functioning are very critical for national politics and the economy, and the economic assessment must be given special attention. Research shows that the impact of transportation is often underestimated in the planning of social and business costs, which must also take into account the influences of globalization, particularly those related to the elimination of market constraints and market liberalization [1]. Traffic accidents on the road have become an increasing problem that threatens the lives of many people around the world. Several factors cause traffic accidents. These factors fall into three main reasons for such accidents: the road (infrastructure), the vehicle (car condition), and the drivers/users (human factors) [2]. Road accidents are the second leading cause of death among people aged 5 to 29 years and the third leading cause among people aged 30 to 44 years. As the number of vehicles is increasing rapidly, this epidemic is growing fast in low- and middle-income countries like Iran. Low- and middle-income countries suffer a significant percentage of preventable deaths and injuries caused by crashes [3].

As shown by world health organization [4], road traffic collisions affect more than 1.2 million people and injury between 20 and 50 million people worldwide yearly, with more than 90% of fatalities in low- and middle-income nations.

Discrete traffic safety response models (often referred to as crash severity models), such as logit and Probit models, are typically used to analyse the correlation between the severity of the accident and its factors involved such as driver characteristics, vehicle characteristics, road conditions and road environmental factors [5]. The ordinal models can also be classified into ordered logit models, ordered Probit (OP) models, and ordered mixed logit models [6]. OP models are an excellent fit for evaluating the extent of crash injury due to the naturally ordered results [7].

In different parts of the road with different features, the number and severity of accidents will vary. Intersections are known as high-risk locations, which is because of the high number of approaches. Accidents are also more likely to occur at intersections due to angular collisions and left-turn collisions and the resulting injuries. Therefore, the safety of these places is considered essential [8].

In 2003, Mohamed Abdel Aty assessed the severity of driver injuries in different sections of roads in Central Florida using an OP model. The results of the fitted models at signalized intersections indicate that the main variables include age and gender of the driver, type of vehicle, non-use of a seat belt, collision location, speed rate, and location type. Also, a surprising result in these models is the negative impact of driver error on the injury severity, where the injury severity on the colliding driver is less [9]. In 2013, Yu Chiun Chiou et al. investigated the severity of two-vehicle accidents at intersections in Taipei using a bivariate generalized OP model. The results show that the intersection geometry is also an essential factor in the injury severity. Well-designed roads reduce the injury severity, while poor designs such increase injury severity [10]. In 2015, Richard Tay fitted an OP model with random parameters to evaluate accidents at rural and urban intersections. The results of this study show that accidents at rural intersections are more likely to cause injury and death than urban [11]. In 2007, Chih-Wei Pai & Wafaa Saleh assessed the severity of motorcyclist injuries under various traffic control conditions at three-legged junctions in England. The results of OP model show that driving on a dark street without lights and uncontrolled intersections simultaneously are fatal for motorcyclists [12]. Also, in 2007, Chih-Wei Pai & Wafaa Saleh assessed the injury severity of motorcyclists at T-junction intersections using an OP model. The results of this study show that the increased age of motorcyclists increases the injury severity. Men also face severe injuries more than women. In the case of vehicle features, heavier motorcycles increase the injury severity [13].

Therefore, it is essential to use accident analysis models to identify the factors affecting it; there are different models that, given the situation, each one can be used. The models can be classified as ordered, non-ordered, or a combination of the two. Since the injury severity of the accident has an ordered nature, ordered models also require smaller sample sizes and thus reduce error rates in the event of data scarcity, it can be said that these models are suitable [5]. Non-ordered models allow variables to have different effects on injury severity groups [14]. It should be noted that Regrettably, after cardiovascular disease, road accidents are considered the second leading cause of death in Iran [15]. In 2014, more than 16000 fatalities due to traffic crashes were observed in Iran [16]. There have been few research studies to directly compare different models of crash severity, although each model type has its unique advantages and limitations [17]. The multinomial logit models and OP models were found to be the most common types of methods used for the study of road crash severity [5].

Therefore, in this study, the OP model and the generalized OP model will be used to assess the severity of accidents occurred in different road components, including intersections and roundabouts and the results of the model will be compared with each other.

## 2 Methodology and data

### 2.1 Mathematical formulas and equations

In this study, the OP and cumulative probit models are used to assess accidents on different road components, including intersections and roundabouts.

The OP model contains discrete ordered data. In order to develop an OP model, it is assumed that all hidden parameters have a normal distribution. It is essential to consider the Ordered nature of injury data (e.g., no injury, possible injury, slight injury, incapacitating injury, and death) in modelling accident injury. The most common approach for generating these models is to specify a hidden parameter  $Z_i$  that is used as the basis for data ordered ranking modelling. This hidden variable is usually specified as a linear function for each accident observation, such as  $Z = \beta X$ . However, the ordered data related to the  $Y$  injury observed is as Eq. (1):

$$Y = k \quad \text{if} \quad \mu_{k-1} < z \leq \mu_k \quad (1)$$

Where  $\mu$  is an acceptable limit that describes  $Y$  as an ordered integer, and  $K$  is the most dependent ordered integer variable.  $\mu$  is a parameter evaluated by  $\beta$ . If error  $\varepsilon$  during the observation has a normal distribution with a mean of zero and a variance of one, an OP model is obtained. The probability that the injury severity belongs to group  $k$  is as Eq. (2):

$$P(y_i=k) = \phi(u_k - \beta X) - \phi(u_{k-1} - \beta X) \quad (2)$$

$U_{k \text{ and } k-1}$  are the upper and lower bounds for the injury severity  $i$ , and  $\phi$  is the standard cumulative distribution function. If the corresponding error is assumed to have a logistic distribution, and Ordered logit model is obtained [9].

In this study, SAS statistical software was used to create the model. This software has a faster processing speed than other statistical software, so it is best to be used in specialized statistical work involving complex information. It also uses various commands to fit ordered response models [17].

Accident data for Zanjan, which is located in the central part of Iran used for modeling and estimating accidents and injury severity. Data on various accidents have been obtained from the traffic police of Zanjan province, including information on accidents that have occurred in different parts of the city. This data includes information about the place of accident, the cause of the accident, the number of deaths, injuries and damages, and information on the driver, the occupants, and the injured. This information includes accidents related to the years 2015-2017 in different parts of Zanjan. The dependent variable of the modelling in this study is the injury severity, which is divided into two groups according to the KABCO method. These groups include 1- No injury / pure damage only, and 2- injury crash.

### 2.2 Data analysis

#### 2.2.1 OP Model (Binary)

In this study, the data of urban roads were fitted into two types of models, including the OP model and the cumulative probit model. Since the level of injury severity only includes two groups (1- No injury / pure damage only, and 2 - injury crash), the models are binary. The results of OP modelling data for Zanjan roundabouts are shown in Table 1.

**Table 1** Estimated coefficients and the significance level of the OP model of the roundabouts

Parameter	Estimate	95% Confidence Limits		Chi-Square	Pr > ChiSq
spring	0.5109	0.1169	0.9049	6.46	0.011
summer	0.4596	0.0822	0.837	5.7	0.017
autumn	0.4041	0.0216	0.7867	4.29	0.0384
holiday	-0.439	-0.8075	-0.0704	5.45	0.0196
heavy vehicle	-1.1205	-1.6729	-0.5681	15.81	0.0001
two-wheeled vehicles	1.5898	1.0113	2.1684	29.01	0.0001
speed violation	-0.8623	-1.5392	-0.1855	6.24	0.0125
sudden redirection	-0.7347	-1.2478	-0.2215	7.87	0.005
failure to observe primacy	-0.9078	-1.2295	-0.5861	30.59	0.0001
other causes	-0.8739	-1.4052	-0.3425	10.39	0.0013

Based on the results presented in Table 1, The general equation of this model for the significant variables is as Eqn (3):

$$Y = 0.5109X_1 + 0.4596X_2 + 0.4041X_3 - 0.439X_4 - 1.1205X_5 + 1.5895X_6 - 0.8623X_7 - 0.7347X_8 - 0.9078X_9 - 0.8739X_{10} \quad (3)$$

The accident season variable indicates that winter accident has the least injury severity, which may be due to the drivers being more cautious this season. Accident occurrence on holidays reduce the injury severity. This can be attributed to the increase in the number of trips and, therefore, the number of vehicles on the roads, which results in a slower speed and a higher degree of caution for drivers. Concerning the type of vehicle involved in the accident, the results show that heavy vehicles reduce the injury severity, while two-wheeled vehicles increase injury severity. Concerning the cause of an accident, the results related to causes of exceeding the speed limit, sudden redirection is disregard of primacy and distraction from the front, and other causes of the accident are significant. The highest severity is attributed to distraction from the front. The least severity is attributed to non-observing primacy.

The results of OP modelling data for Zanjan intersections show the least injury severity of the accident is related to winter accidents. Results for vehicle type show that the fitted model is only significant for cars and vehicles with two wheels, which can be due to the low number of accidents in other groups. Two-wheeled vehicles increase the injury severity compared to cars. The results of the accident cause variable show that the fitted model is significant only in the group of distracted from the front and the group of other causes. Most injury is due to the distraction from the front. It can be interpreted that due to the driver's distraction from the front of himself, the reaction time of the driver will increase if there is a critical situation, resulting in a more severe collision. Also, daytime accidents reduce the injury severity compared to the night, which can be due to driver glare, lack of sufficient vision, and lighting at night.

### 2.2.2 Cumulative probit model

The results of the cumulative probit model for intersections are shown in Table 2.

**Table 2** Estimated coefficients and the significance level of cumulative Probit model of intersections

Parameter	Estimate	Wald Chi-Roundabout	Pr > ChiSq
spring	0.3681	5.3434	0.0208
summer	0.4589	9.4359	0.0021
autumn	0.3494	4.9352	0.0263
two-wheeled vehicles	1.5801	31.0832	0.0001
other causes	-0.3847	5.0706	0.0243
Light condition	-0.4197	12.2561	0.0005

The general equation of this model for the significant variables is as Eq. (4):

$$Y = 0.3681X_1 + 0.4589X_2 + 0.3494X_3 + 1.5801X_4 - 0.3847X_5 - 0.4197X_6 \quad (4)$$

Based on the results, accident occurrence in seasons, apart from winter, increases the injury severity. Results for vehicle type are significant only for the group of two-wheeled vehicles and indicate an increase in the injury severity of this group compared to car drivers. Also, the results for accident causes are significant only for the other causes group and indicate a decrease in the injury severity of this group compared to the group of distractions from the front. The most severe injury in this model is attributed to the distraction from the front. The results of the lighting condition of the route also show a decrease in the severity of the accident occurring during the day compared to night.

The results of the cumulative probit model for roundabouts show that occurrence of an accident on holidays has less injury severity than non-holidays. The results of the variable vehicle type show that the injury severity of the accident in heavy vehicles is reduced compared to cars, while the injury severity of two-wheeled vehicles increases compared to cars. Also, the results show that speeding violations, sudden redirection, disregard for the priority of crossing, and other causes of accidents reduce the injury severity compared to roll over and distraction from the road.

### 2.2.3 Comparison of models

Table 3 presents Log-likelihood and Percent Concordance indices for model comparison.

**Table 3** Comparison of models

	Model	Log-likelihood	Percent concordance
Intersection	OP	-367.78	-
	Cumulative Probit	-355.9	72.5
Roundabout	OP	-234.12	-
	Cumulative Probit	-219.7	70.6

Larger values for Log-likelihood statistics indicate a better fit of the model. The presented results show that the cumulative model is better than the OP model to describe the dependent variable based on independent variables. The results of the Percent Concordance statistic for the cumulative Probit model show that the fitted model for intersections is better than roundabouts.

In SAS software in the OP model, it is not possible to automatically select variables by software, and manual selection of variables is time-consuming and can cause errors. Therefore, it can be concluded that the cumulative Probit model is better able to describe the dependent variable based on the independent variables than the OP model. Table 4 presents values for comparing the cumulative Probit models for different road segments. Based on this table, the lower value for the Brier Score statistic that is a value between zero and one, indicates a good fit to the model. By comparing this statistic, the intersection model has better fitness than the roundabout model. Also, the cumulative model of intersections has the least error. The AIC and BIC indices are used to evaluate the fitness of the model, which lower values indicating better fitness of the model. By comparing the AIC and the BIC, it can be seen that the roundabouts model has fewer values than the intersections model. Considering the values in the table 4, it can be concluded that the fitted roundabouts model is better than the intersections model.

**Table 4** Comparison of cumulative Probit model in different road components

Road type	Intersection	Roundabout
Brier Score	0.150828	0.184615
BIC	818.3386	511.458
AIC	743.8721	463.4114
Log-Likelihood	-355.9	-219.7

### 3 Conclusion and discussion

Since the models built to depend on database features, including sample size, data quality, number of independent variables, and their quality, therefore, modelling with different databases can present different results. Overall, it can be said that the fitness of a model depends on the quality of the database, which is in line with the relevant literature.

Based on all the results presented, it can be found that cumulative Probit models are capable of interpreting the dependent variable better than the OP model. Also, in cumulative models, unlike OP models, the software can automatically select independent variables, thus reducing the error rate. Based on the overall results of the models, it can be seen that the occurrence of an accident in daytime compared to nighttime, on holidays compared to non-holidays and in winter compared to other seasons, reduces the injury severity. Heavy vehicles reduce the injury severity compared to cars, while two-wheeled vehicles, including motorcycles and bicycles, increase the injury severity. Accidents at intersections also increase injury severity. Concerning the cause of the accident, the results show that distraction from the front increases the injury severity compared to other causes of the accident. Of 802 accidents of intersections, 78.18% are at the highest level of injury, while of the 453 accidents of roundabouts, 65.12% have a higher severity level. Therefore, it can be seen that the occurrence of an accident at intersections increases the severity of injury compared to roundabouts, which can be due to the high interference flows at intersections.

## References

- [1] Antić, B., Vujančić, M., Lipovac, K., Pešić, D.: Estimation of the traffic accidents costs in Serbia by using dominant costs model, *Transport*, 26 (2011) 4, pp. 433-440
- [2] Norman, L.G.: Road traffic accidents: epidemiology, control, and prevention, World Health Organization, 1962.
- [3] Anjuman, T., Hasanat-E-Rabbi, S., Siddiqui, C.K.A., Hoque, M.M. editors. Road traffic accident: A leading cause of the global burden of public health injuries and fatalities. International Conference on Mechanical Engineering 2007, pp. 29-31, Dhaka, Bangladesh, 29-31 December 2007.
- [4] Global status report on road safety: time for action, World Health Organization, 2009.
- [5] Ye, F., Lord, D.: Comparing three commonly used crash severity models on sample size requirements: multinomial logit, ordered probit and mixed logit models, *Analytic methods in accident research*, 1 (2014), pp. 72-85, DOI: 10.1016/j.amar.2013.03.001
- [6] Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A.: The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives, *Accident Analysis & Prevention*, 43 (2011) 5, pp. 1666-1676
- [7] Yu, R., Abdel-Aty, M.: Using hierarchical Bayesian binary probit models to analyse crash injury severity on high speed facilities with real-time traffic data, *Accident Analysis & Prevention*, 62 (2014), pp. 161-167
- [8] Ma, Z., Lu, X., Chien, S.J., Hu, D.: Investigating factors influencing pedestrian injury severity at intersections, *Traffic injury prevention*, 19 (2018) 2, pp. 159-164
- [9] Abdel-Aty, M.: Analysis of driver injury severity levels at multiple locations using ordered probit models, *Journal of safety research*, 34 (2003) 5, pp. 597-603
- [10] Liu, Y.: Weather impact on road accident severity in Maryland, 2013.
- [11] Tay, R.: A random parameters probit model of urban and rural intersection crashes, *Accident Analysis & Prevention*, 84 (2015), pp. 38-40
- [12] Pai, C.W., Saleh, W.: An analysis of motorcyclist injury severity under various traffic control measures at three-legged junctions in the UK, *Safety science*, 45 (2007) 8, pp. 832-847
- [13] Pai, C.W., Saleh, W.: Exploring motorcyclist injury severity resulting from various crash configurations at T-junctions in the United Kingdom—An application of the ordered probit models, *Traffic Injury Prevention*, 8 (2007) 1, pp. 62-68
- [14] Patil, S., Geedipally, S.R., Lord, D.: Analysis of crash severities using nested logit model—accounting for the underreporting of crashes, *Accident Analysis & Prevention*, 45 (2012), pp. 646-653
- [15] Rahmati, F., Doosti, M., Bahreini, M.: The cost analysis of patients with traffic traumatic injuries presenting to emergency department; a cross-sectional study, *Advanced journal of emergency medicine*, 3 (2019) 1
- [16] Mehmandar, M., Soori, H., Mehrabi, Y.: Predicting and analysing the trend of traffic accidents deaths in Iran in 2014 and 2015, *International journal of critical illness and injury science*, 6 (2016) 2, pp. 74-78
- [17] Park, H.M.: Regression models for ordinal and nominal dependent variables using SAS, Stata, LIM-DEP, and SPSS, 2015.

