ZANCO Journal of Pure and Applied Sciences
The official scientific journal of Salahaddin University-Erbil
https://zancojournals.su.edu.krd/index.php/JPAS

ISSN (print):2218-0230**, ISSN (online):** 2412-3986**, DOI: http://dx.doi.org/10.21271/zjpas**

RESEARCH PAPER

Analysis of the Contributory Factors to Accidents at Signalized Intersections using Generalized Estimating Equation with Negative Binomial Distribution

Nasreen A. Hussein[1], Rayya Hassan[2]

[1]Department of Civil Engineering, College of Engineering, University of Duhok-Duhok, Kurdistan Region, Iraq
[2]Department of Civil and Construction Engineering, Swinburne University of Technology, Victoria, 3122, Australia

A B S T R A C T:
Ensuring safety on roadways and controlling conflicting and merging streams of traffic at signalized intersections has become a major concern for transportation engineers and road users globally and locally. Many factors contribute to accident occurrence at road segments and intersections with driver behavior being the main contributor. Other factors include those related to the road and the vehicle. The primary aim of this study is to understand the impact of different factors causing traffic accidents at signalized intersections. It also include developing a descriptive study of traffic accident data to show the variation with different factors at both locations (intersections and roadway segments). To achieve this objectives, 22 signalized intersections and all roadways in Duhok city/Iraq were selected. Assessing the impacts of geometric characteristics and driver behavior on safety performance of the sites have been achieved through observations of trends from graphs and use of appropriate advanced statistical analysis such as Generalized Estimating Equation (GEE) with NB distribution and log link function. The results show that approach width, driver inattention, red light running and surface moisture condition are the significant contributor factors to accident occurrence. The study results highlight the importance of driver education, police enforcement and appropriate geometric design of intersections to the reduction of accident occurrence.

KEY WORDS: Signalized intersections, Accidents, Geometric characteristics, Driver behaviour, Generalized estimating equation.
DOI: http://dx.doi.org/10.21271/ZJPAS.35.2.4
ZJPAS (2023) , 35(2);29-40 .

1. INTRODUCTION:

Traffic accidents are one of the main causes of death, injury, disability, and property damage as well as financial costs in the world. Historical data from Duhok Traffic Directorate were used to evaluate traffic accidents during years 2010 to 2020. It was found that in Duhok Province on average 1000 accidents occur and more than 2,000 people were injured and about 140 people died in road accidents each year.

* **Corresponding Author:**
Nasreen Ahmed Hussein
E-mail: nasreen.hussein@uod.ac
Article History:
Received: 18/07/2022
Accepted: 02/10/2022
Published:20/04/2023

These include road segments, intersections (signalized and unsignalized) at both urban and rural areas. These statistics suggest that accidents at roadways should be a main concern for road agencies. According to Duhok Traffic Police accident reports, the main contributing factors are risky driving behavior (e.g. speeding, inattention, red light running,). Therefore, it would be necessary to focus also on human factors effect on road traffic accident occurrence. While investigating the effect of human factors, it would be of significant value to take into account the second factor (roadway geometric characteristics). Signalized intersections constitute only a small part of the highway system but they are the location of a disproportionate number of accidents and cause a high accident risk due to the different conflicting maneuvers. Therefore, these findings highlight a significant problem which requires a better understanding of the factors that contribute to accident occurrence at signalized intersections. Many factors contribute to accident occurrence at road segments and intersections in Detroit, Michigan, USA, with driver behavior being the main contributor (Schattler and Datta, 2004). Other factors include those related to the road and the vehicle. The three elements (human, vehicle and road/environment) along with their interactions differ in the strengths of their impact (Treat et al., 1977).

Ensuring safety on roadways and controlling conflicting and merging streams of traffic at signalized intersections has become a major concern for transportation engineers and road users. This can be done through providing sufficient geometric characteristics that control vehicle movements through intersections. Further, education and effective training of road users, information and awareness are also of primary importance and contribute to improve the safety.

Numerous studies have pointed out the relation between traffic volume, geometric characteristics, driver behavior and the number of accidents occurred at road network (Ogden et al., 1994; Chin and Quddus, 2003; Turner et al., 2012; Retallack and Ostendorf, 2020; Shams et al., 2021). Several studies showed that accident rate and severity reduce at divided roadway with the presence of raised medians (Sawalha and Sayed, 2003; Frawley and Eisele, 2004; Wang and Abdel-Aty, 2006). Lower accident frequency are related to intersection approaches with shared through/right turn lane (Fitzpatrick et al., 2006) and approaches at signalized intersections with a greater number of lanes are related to higher accident frequency (Abdel-Aty and Wang, 2006).

In a study conducted by McDonald (1953) at 150 signalized intersections in California/USA it was found that accident rates at intersections are more related to minor road volume than to major road volume. Baguley (1988) noted that there is a negative correlation between the numbers of red light running violations and the volume of traffic (i.e. there is an increase in red light running violations with high volume of traffic). However, Persaud et al. (2002) found that at signalized intersections, traffic volume on both minor and major roads was the significant factor in accident prediction models.

Using Negative Binomial model to study the effect of roadway lighting at intersections safety, Donnell et al. (2010) found that the difference in total estimated accidents using aggregated results of day and night models are a 3% decrease in accidents at intersections with lighting as compared to intersections without lighting. Al-Reesi et al. (2013) stated that unsafe driving in general and aggressive destruction of traffic rules in particular are main risk factors. Al-Atawi (2013) found that number of approaches, road width and speed significantly contribute to red light violation. Furthermore, safety performance deterioration in signalized intersections is based primarily on the increase in traffic speed (Alghafli et al., 2021). Dong et al. (2014) found that traffic volume, percentage of truck, lighting condition, and intersection angle are significant factors affecting signalized intersection safety. In A recent research conducted by Retallack and Ostendorf (2020) at 120 intersections in Adelaide, Australia, it was found that at lower traffic volumes there is a nearly linear relationship between traffic volume and accident frequency and accident frequency increases at a higher rate.

The primary goal of this study is to understand the impact of different factors causing traffic accidents at signalized intersections in the city of Duhok/Iraq. The specific objective of this study is to assess the effects of different driver behavior, volume and geometric characteristics of intersections and road segments on safety performance in terms of accident frequency through observation of trends from graphs and statistical analysis. This can be done by developing a descriptive study of the traffic accident data for the study area between the period 2017-2021 to show the variation with different factors at both locations (intersections and roadway segments).

It also covers developing models linking significant contributing parameters to accident occurrence at signalized intersections. Such method requires the use of appropriate advanced statistical analysis such as Negative Binomial (NB) regression model. With repeated observations of accident frequency exist in the data, the application of Generalized Estimating Equation (GEE) with NB distribution and log link function can address this problem.

2. SITE SELECTION

The main principle for site selection is the availability of relevant data to enable performance assessment of the selected sites. Considering accessibility to the required data, Duhok city roadways (including segments and intersections) was selected as the study area, refer to Figure (1). Relevant data that were available for this research cover the study period of 2017-2021. For achieving each objective, the sites were selected as follows:

- For developing a descriptive study of the traffic accident data to show the variation with different factors, the analysis involves all roadways and intersections (signalized and un-signalized) in Duhok city road network.
- For developing models linking significant contributing parameters to accident occurrence, the analysis involves using the details for a sample of 22 signalized intersections in Duhok city with different geometric characteristics.

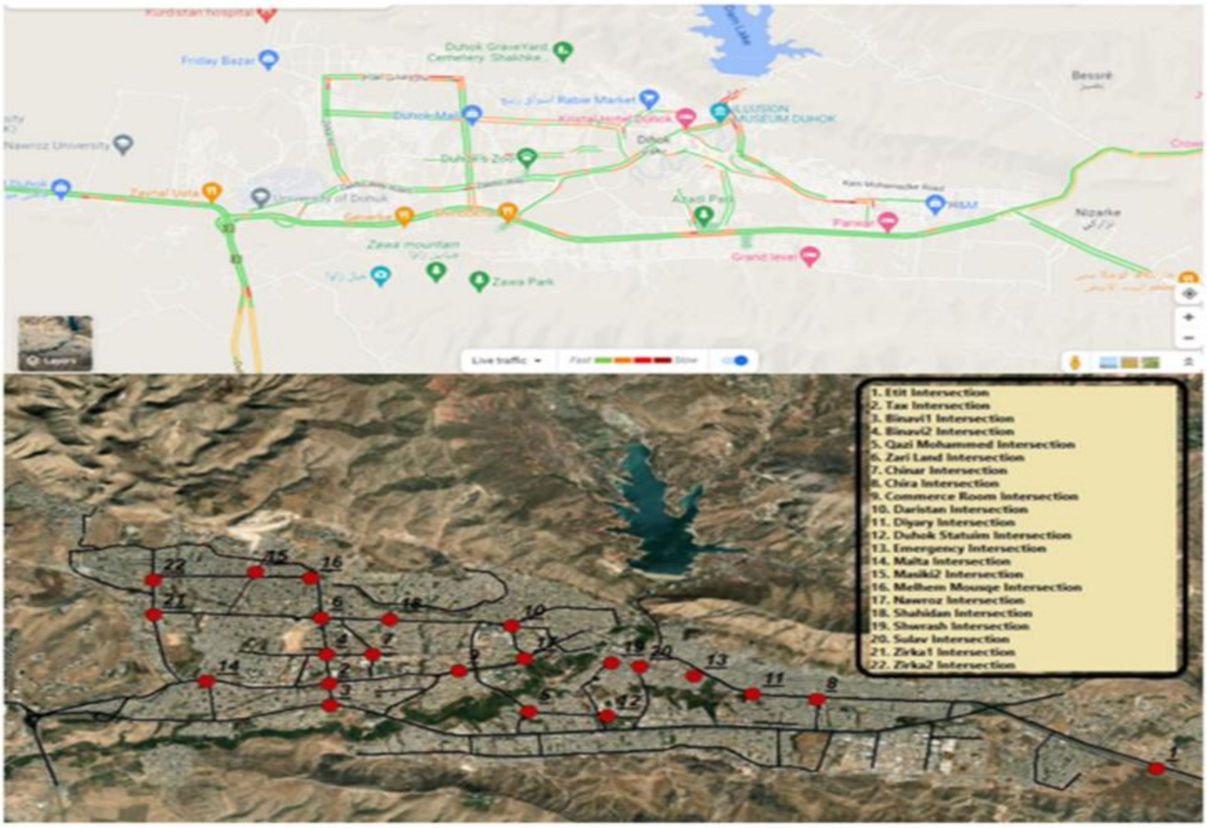


Figure 1: Duhok city study area

3. METHODOLOGY AND DATA COLLECTION

The present study is achieved by collecting and preparing relevant data for each site (intersections and road segments). For trend analysis and to develop accurate accident prediction models the data collected include detailed accident data, driver behavior, geometric characteristics, traffic volume data and other situational factors such as lamination condition and surface moisture condition. It also involved the employment of GEE with Negative Binomial regression model to identify the contributing factors to the occurrence of accidents at signalized intersections.

3.1 Data Preparation

This section provides details of data collection, preparation and statistical techniques that were used in addressing the objective of this study. Accident data were obtained from Duhok Traffic Directorate using 1387 accident reports which contain information on accidents that occurred during the 5-year period (2017-2021) recorded and provided by the Duhok traffic police. In addition to selection of a large enough time interval, it was necessary to identify accidents that occurred within the intersection or its approaches to use them in the modeling analysis. To achieve more precise results in accident analysis, accidents should be allocated to their exact sites. This would help in relating intersection characteristics to total accident occurrence at intersections and approaches, so their causes and effects can be studied in the analysis. The intersection models estimate the expected frequency of accidents that occur within the limits of an intersection i.e. those that occur on the intersection legs within 100m (HSM, 2009).

Accident data collected include all accidents covering all severity levels (fatal, injuries and property damage only), type of accident (collision with pedestrian, angle accident, head on, side swipe, rear-end, collision with fixed object and roll over), location/site type (intersection and non-intersection), lamination condition when they occurred (day or night), road surface condition (wet or dry) and cause of accident (driver inattention, red light running and speeding). Accident data are reported on forms by traffic police staffs. It is important to note that not all accidents are reported, and not all reported accidents are recorded accurately. According to HSM (2009), errors may occur at any step of the collection and recording of accident data and may be due to data entry/typo errors, inaccurate entry/ the use of general terms to define a location, level of accident severity, road surface, vehicle types, etc. and subjectivity/where data collection relies on the personal judgment of an individual, contradiction is likely, for example excessive speed. It is also important to note that, the count of reported fatality accidents may contain incomplete information, as traffic police are responsible to record an accident as fatality, when a person involved in the accident died immediately at the time of accident. Otherwise, the data are recorded in the General Directorate of Health/Duhok.

The information of driver behavior was collected from the accidents report forms, in which they are reported as causes of accidents. These were classified into three groups, namely: inattention, red light running and speeding. The traffic volume data required for achieving the objective of this study are annual average daily traffic (AADT). Traffic volume data in terms of AADT of the year 2007 for the identified intersections and road segments was extracted from Master plan of Traffic of Duhok city (2009). For each site, traffic volume data were collected for the 5 years under study, covering the same number of years of available accident data. Traffic growth factor was used in estimating traffic volume for the years of missing data over the study period.

The intersection geometric data provide information about the physical features of the accident site. For signalized intersections, data collection was required for each individual approach of the intersection itself. The geometric characteristics that have been collected are approach width, intersection clearance distance (the crossing distance from one approach to the opposite approach), presence of median and presence of shared lane. Data of speed limit was collected by reviewing the observed speed limit at each intersection approach. In this study, site characteristics were checked and found to be the alike over the study period (Duhok Traffic Directorate, 2022). 10 four-legged intersections and 12 three-legged intersection were chosen for geometrics characteristics assessment.

3.2 Statistical Approaches

Detailed data collected and prepared in this study were utilized statistically to assess the safety performance of signalized intersections. One of the characteristics of accident frequency data is that the variance of accident counts is greater than the mean (over dispersion). The equi-dispersion assumption of Poisson regression requires that the mean and variance of accident count data is equal. Poisson regression is a restricted form of the NB model with $\alpha = 0$. The parameter (α) is referred to as the dispersion parameter. Lord and Mannering (2010) stated that the Negative Binomial is an appropriate model to overcome the over-dispersion problem in the data. Through letting the mean number of accidents be $\mu i=\lambda \exp(\varepsilon)$, the NB model can be derived as:

$$\mu_i = \exp(X_i^T \beta + \varepsilon_i) \qquad (1)$$

Where

$\exp(\varepsilon i)$ = gamma distributed error term with variance α and mean 1.

X^T_i= vector of explanatory variables, and

β= vector of regression coefficients

The NB model permits the variance to differ from the mean (Carson and Mannering, 2001), when $\alpha \neq 0$ since,

$$Var[Y]=\mu_i+\alpha\mu_i^2 \qquad (2)$$

Where,

Var [Y]= variance of accident count data

In this study, count regression was applied for accident data analysis using the available geometric characteristics, traffic volume, driver behavior and other situational factors such as lamination condition and surface moisture condition. Traffic volume was transformed to its natural log and included in the model as a main effect. The accident data are not independent due to the repeated observation of each intersection on four conditions (day-wet, day-dry, night-wet and night-dry) and for all three categories of driver behavior (inattention, red light running and speeding). This clustering needs to be considered in the statistical analysis to determine appropriate standard errors. The GEE method used herein, is an extension of the Generalized Linear Model (GLM) that takes into consideration the correlation among clustered data (Liang and Zeger, 1986).

4. RESULTS AND DISCUSSION

The effects of driver behavior and geometric characteristics on intersections safety performance are assessed here through observation of trends from graphs and regression analysis. This section starts with preliminary analysis of the collected data including descriptive analysis for distribution of accidents by several variables. Also reported in this section are the statistical approaches used to assess the effects of different explanatory variables on accident occurrence at signalized intersections.

4.1 Accident Frequency Distribution

During the 5-years period (2017-2021) investigated herein, the road network in Duhok city was associated with nearly 1387 accidents. These accidents occurred along roadway segments and intersections. Descriptive analysis is performed to study the distribution of accident data according to several variables, including:

- Accident type (rear end, head on, right angle, side swipe, collision with pedestrian, collision with fixed object and roll over)
- Accident severity (fatality, serious injury, other injury, property damage only (PDO))
- Road surface moisture condition (dry, wet)
- Lamination condition (day, night)
- Cause of accident (driver inattention, red light running and speeding)

It is important to note that there is a slight difference in the number of accidents when comparing distribution of accidents by different variables. This is due to lack in information for some accidents relevant to specific variables. Distribution of accidents by type is given in Figure (2). It indicates that rear-end, right angle and side collision comprise a large proportion of accidents. It can be noticed that large portion of right angle (37%) and rear-end accidents (31.7%) of total accidents occurred at all types of intersections and particularly at signalized intersections with 36.8% right angle accidents and 36.4% rear-end accidents. This is due to the fact that driver decision has a great impact on the probability of rear-end and right angle accidents at signalized intersections.

The right angle accidents occur when drivers are unable to clear the intersection safely before the conflicting movements get green light. Inadequate signal timing, excessive speed, slippery pavement could also be possible contributory factors to right angle accidents at signalized intersections (HSM, 2009). The rear-end accident occurs when drivers stop suddenly (Kennedy and Sexton, 2009). Other possible contributory factors for rear-end accidents include inappropriate approach speeds, unexpected lane changes on approach, poor visibility of signals, narrow lanes, excessive speed and slippery pavement (HSM, 2009).

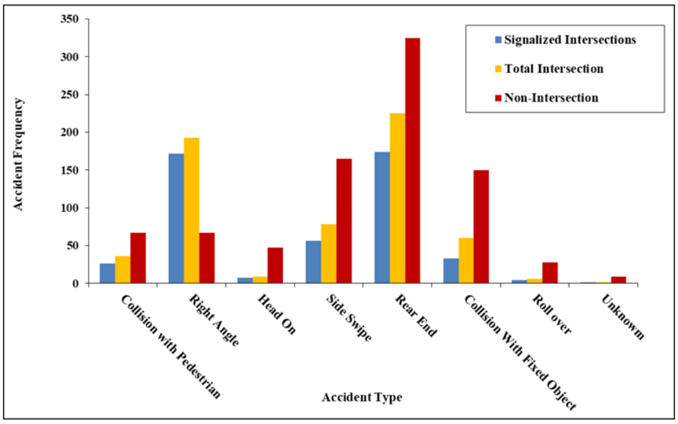
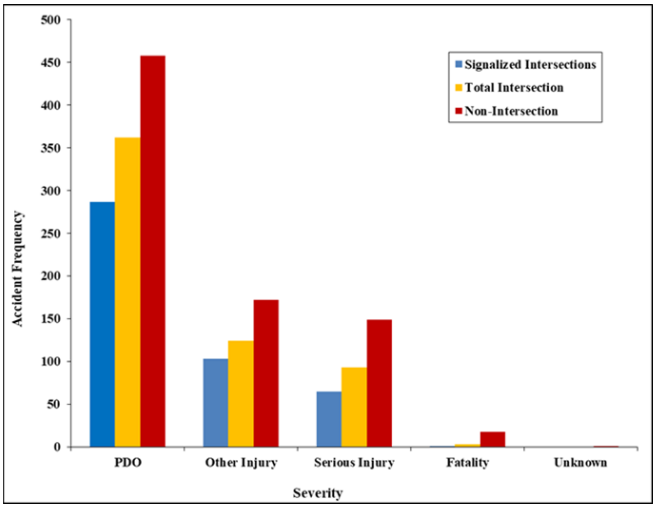


Figure 2: Distribution of accidents by type

Distribution of accidents by severity level for the three site types in the data set i.e. all intersections (total intersections), signalized intersections only and non-intersection is presented in Figure (3). This figure shows that, the



majority of accidents are of property damage only (PDO) followed by injuries (serious and other) with fatal ones being the least. The count of reported fatality accidents may contain incomplete information, as traffic police are responsible to record an accident as fatality, when a person involved in the accident died immediately at the time of accident.

Figure 3: Distribution of accidents by severity

Distribution of accidents by road surface moisture condition for the three site types (or locations) considered herein, is presented in Figure (4). It indicates that, for all data sets, most accidents occurred in dry surface conditions than wet conditions. Distribution of accidents by lamination condition for all three locations is presented in Figure (5). It indicates that, for the three data sets, most accidents occurred during day time than night time. The results from Figures (4) and (5) are consistent with what is reported in Yan et al. (2005) and Turner et al. (2012). The high number of accidents at day and dry conditions is believed to be related to higher volumes/travel during these condition; dry days than wet nights.

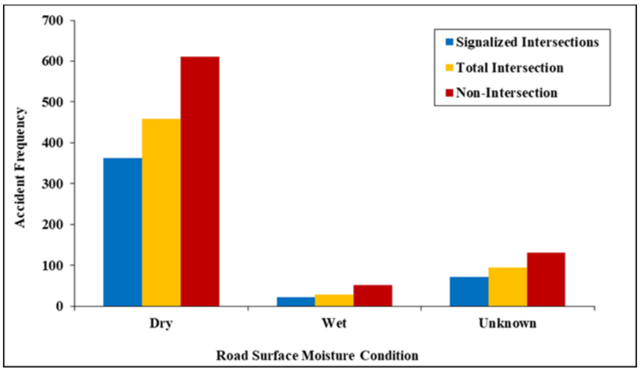


Figure 4: Distribution of accidents by road surface moisture condition

Distribution of accidents by probable cause from police accident reports is presented in Figure (6). This reveals that driver inattention is a major contributory factor to accident occurrence followed by speeding and the red light running being the last one.

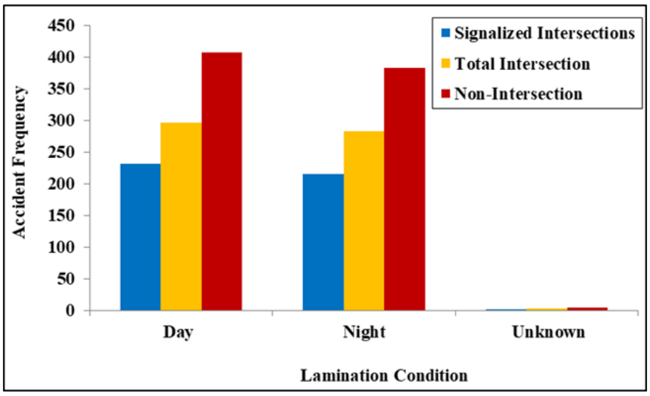


Figure 5: Distribution of accidents by lamination condition

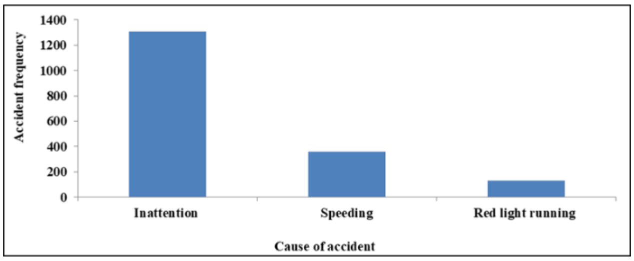


Figure 6: Distribution of Accidents by Probable Cause

4.2 Geometric Characteristics and Accidents-General Trends

Descriptive analyses are performed to demonstrate the distribution of accident frequency with different geometric characteristics. A summary of observations from the analysis is given in Table (1).

4.3 Modeling Safety Performance of Signalized Intersections-Statistical Analysis

The main purpose of this part of study is to evaluate the effects of geometric characteristics and driver behavior on safety performance at signalized intersections. Modeling of accident data was performed based on different contributory factors to establish those that have significant impact on accident occurrence. Accident frequency models at signalized intersections (approach/departure level analysis) were fitted by applying GEE with log link function for accident analysis using Statistical Package for the Social Sciences (SPSS) GENLIN procedure (SPSS, 2022). In this study, the model takes into account the effect of driver behavior (inattention, speeding and red light running), lamination condition and surface moisture on accident occurrence and for the signalized intersections under study (22 intersections). Accident data was observed to vary by time of day and moisture condition and resulted in repeated observations, i.e. each intersection or site is repeated twelve times (inattention with night-wet, night-dry, day-wet and day-dry), (speeding with night-wet, night-dry, day-wet and day-dry) and (red light running with night-wet, night-dry, day-wet and day-dry). Therefore, each intersection was considered as a cluster with twelve repeated observations.

Traffic volume in terms of logAADT, lamination condition (day coded as 1 and night coded as 0) and surface moisture condition (dry coded as 1 and wet coded as 0), driver behavior (inattention coded as 2, speeding coded as 1 and red light running coded as 0), were explored for their effect on accident occurrence. The geometric explanatory variables including approach width, presence of shared lane, presence of median and clearance distance were also factors of interest in analyzing accident frequency data.
As shown in the trend analysis (refer to Table 1), both approach width and clearance distance tend to have more than two levels. Therefore, a collection of dummy variables was required to represent these variables in statistical analysis. Accordingly, approach width was categorized into three levels and includes approach width < 10m coded as 2, approach width from 10-15m coded as 1 and approach width ≥ 15m coded as 0, with the majority of intersection approaches being in the range of 5-10m wide and account for 60% of the sample. The clearance distance was categorized into a distance < 20m and coded as 2, clearance distance from 20-30m coded as 1 and clearance distance ≥ 30m coded as 0 with the majority of intersections having depths ranging between 20-30m.

Table 1: Summary of observations from descriptive analysis

Geometric Characteristics	Summary of Observations
Width of approaches	The widthes of the selected approaches range between 5 m and 25 m with the majority of intersection approaches and departures is 5-10 m wide and exists in 60% of the sample Accident frequency decreases as the approach width increases beyond 15m.
Clearance Distance	The distance of the selected intersections range between 10 m and 49 m with the majority of intersections having depths ranging between 20-30m that account for 52% of all intersections. Accident frequency increases with increase in clearance distance.
Presence of median	Medians are present at 95% of approaches in the sample set. Approaches with medians are associated with higher accident frequency than approaches with no medians.
Daily traffic volume	The sample considered in this study consists of a range of low, medium and high traffic volume approaches and departures with a high number of sample having 10000 to 20000 vehicles per day Sites with high traffic volumes (> 10000) are associated with higher accident frequency than other traffic categories except at very high volume approaches (> 25000).
Presence of shared lane	Shared lanes are present at 50% of approaches in the sample set. Approaches with shared lanes are associated with 46% of accident frequency

4.4 Evaluation of Over Dispersion in Accident Data

Numerous criteria were used to select the proper probability model for accident data analysis. These criteria are: testing whether the dispersion parameter > 0 using the Lagrange multiplier test and examining whether the dispersion statistic is greater than one. The histogram of accident data includes high number of zero values and a long positive skewness of the dependent variable (DV i.e. accident frequency). Furthermore, the variance of accident counts is greater than the mean proposing over dispersion in accident data.

The value of the dispersion statistic (Pearson chi square/degree of freedom) for the Poisson model with intercept only is 2.9 and is greater than 1 which proposes that accident frequency is over dispersed. In addition, the dispersion parameter (alpha, α) denotes the parameter used in NB model to consider over-dispersion. Poisson model accepts it to be zero; however, the Negative Binomial model permits it to be greater than zero. The dispersion parameter for NB model with intercept only is 4.02 and is greater than zero which indicates that the dependent variable (accident frequency) is over dispersed.

Results of the Lagrange multiplier test for accident data indicate that the Z-score test is 4.83 with a t-probability of $P < 0.001$. A significant Lagrange indicates that the model dispersion parameter is different from 0. These results verify that the hypothesis of no over dispersion is rejected and that actual over dispersion exists in the data set. Consequently, accident data ought to be modeled as negative binomial.

4.5 Modeling Accident Data

Based on the goodness of fit evaluation of the model described in the previous section, this section uses the traditional NB model as a first step to find the dispersion parameter for the probable model including all variables. GEE with NB distribution and log link function was applied to analyze accident data as the dependent variable. Exchangeable correlation matrix was used as appropriate correlation structure to account for correlation among repeated observations. The value of dispersion parameter (α) was inserted to GEE negative binomial model with all predictors. The GEE was fitted using the value of $\alpha = 0.89$. The detail output results are given in Table (2). It is important to note that observations for all explanatory variables for the sample are not available (156 out of 1320 observations are missing volume data which are related to 13 approaches and departures).

4.6 GEE Model Interpretation

This section presents the interpretation of the results for analysing the effect of different contributory factors on accident data. It is important to note that the models include the significant and non-significant independent variables. The results indicated that not all variables had a statistically significant effect to describe the variation in accident frequency. Results in Table 2 illustrate that, presence of median, presence of shared lane; clearance distance, lamination condition and traffic volume are not significant in accident occurrence (p value > 0.05). However, approach width, surface moisture condition (dry vs. wet) and driver behavior are significant in accident occurrence with p value of < 0.05.

The intercept is the expected number of accidents when all variables in the model are evaluated at zero. That is, at an approach width less than 10 m, a shared lane and a median, during a dry day and a clearance distance less than 20 m, the expected total accident count is exp (-1.387) = 0.250 accidents per year if all other variables take their mean values.

The main effect of approach width in this table indicates that generally wider approaches with approach width ≥15m tend to have higher accident frequency than narrower approaches as indicated by the positive coefficient of 0.95. Similar findings have been reported in Othman et al. (2009) where the authors related this finding to lane changing maneuvers, overtaking behavior and higher speeds on wide carriageways. The significant effect of surface moisture condition with negative coefficient of -2.795 implies that wet surface has lower accident frequency than dry surface. These findings are related to higher exposure during dry conditions due to higher number of vehicles entering the intersections. Whereas traffic volume decreases at night during wet conditions as drivers are less likely to travel during these conditions hence, reducing the likelihood of accident occurrence (Keay and Simmonds, 2005).

The effect of driver behavior with positive coefficients of 1.5 for inattention characteristic and 1.159 for red light running shows that the probability of accident occurrence is more likely when drivers are inattentive and run the red light signal as compared to speeding behavior. Klauer et al. (2006) indicated that inattentive drivers while drowsy results in a four- to six-times higher accident risk compared to alert drivers. Furthermore, Kamyab et al. (2000), stated that an increase in the traffic volume and signal cycle time results in an increase in the frequency of red light running violations. Therefore, to plot the effect of significant categorical predictors on accident frequency, it was necessary to calculate the predicted value of accidents as a function of approach width, surface moisture condition and

Table 2: GEE regression with NB and log link for accident dat

Parameter	Coefficients β	Std. Error	P- Value	Exp (β)	95% Wald Confidence Interval for Exp(β)	
					Lower	Upper
(Intercept)	-1.387	1.0547	0.188	0.250	0.032	1.974
Approach width ≥15m =0	0.950	0.3916	0.015 0.062	2.586	1.200	5.572
Approach width 10m – 15m =1	0.368	0.1971	.	1.445	0.982	2.126
Approach width < 10m =2*	0	.		1	.	.
Median, not present=0	-0.414	0.3492	0.236 .	0.661	0.333	1.311
Median, present=1*	0	.		1	.	.
Shared lane, not present=0	-0.198	0.1623	0.223 .	0.821	0.597	1.128
Shared lane, present=1*	0	.		1	.	.
Clearance distance ≥ 30m =0	0.509	0.3222	0.114	1.664	0.885	3.129
Clearance distance 20m – 30m =1	0.257	0.2245	0.253	1.292	0.832	2.007
Clearance distance < 20m =2*	0	.	. 0.840	1	.	.
Lamination condition, Night=0	0.021	0.1024	.	1.021	0.835	1.248
Lamination condition, Day=1*	0	.		1	.	.
Surface MC, Wet=0	-2.795	0.1625	0.000 .	.061	0.044	0.084
Surface MC, Dry=1*	0	.		1	.	.
Driver Behavior, inattention=0	1.500	0.1656	0.000 <.001	4.483	3.241	6.201
Driver Behavior, Red Light Running=1	1.159	0.1829		3.186	2.226	4.560
Driver Behavior, Speeding=2*	0	.	.	1	.	.
LogAADT	-0.026	0.1194	0.829	0.975	0.771	1.232
(Negative binomial)	0.89					
Number of intersection approaches (number of clusters)	97					
Size of cluster	12					
Number of observations	1164					

and driver behavior individually with representative values of traffic volume as shown in Figure (7). It can be seen that the trends are consistent with the results shown in the Table (2). Depending on the results of Negative Binomial model shown in Table 2, the regression equation with significant predictors can be written as:

$$\text{The expected accident count} = \exp(-1.387 + 0.95 * (\text{Approach Width} > 15\text{m}) - 2.795 * (\text{Surface Condition, wet}) + 1.5 * (\text{Driver behavior, inattention}) + 1.16 * (\text{Driver behavior, red light running})$$

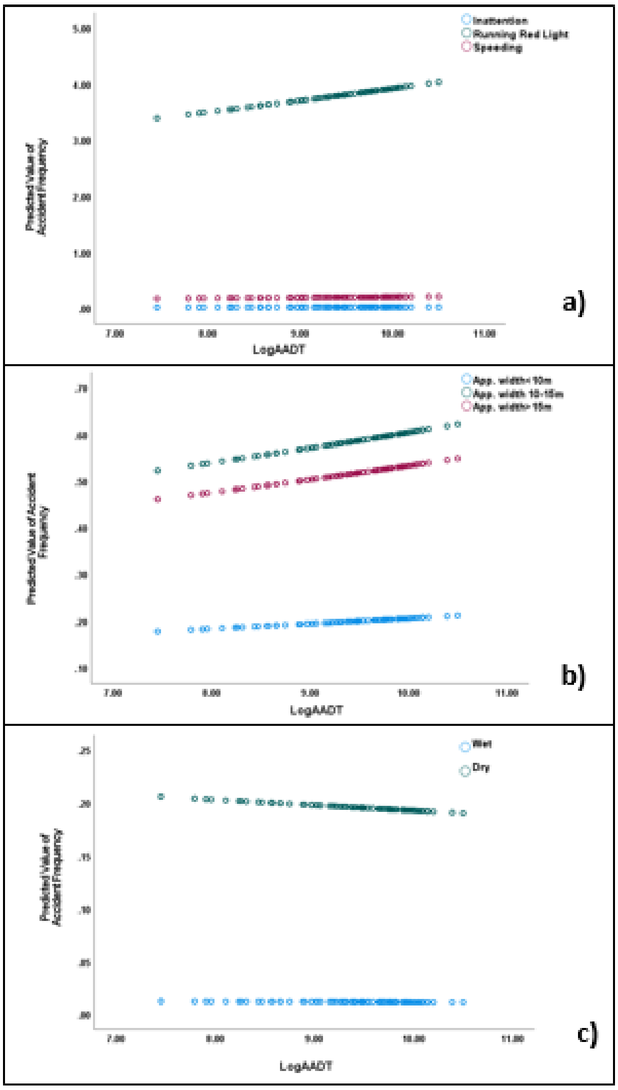
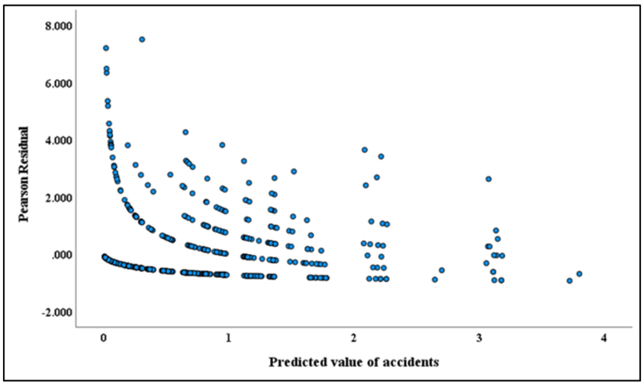


Figure 7: Effect of a) driver behavior, b) approach width and c) road surface condition on predicted value of accident frequency

4.7 Model Evaluation

To assess whether the Negative Binomial model adequately fits the data, predicted value of accidents were plotted versus Pearson residuals (Hilbe, 2011). One of the criteria for selecting the final model includes falling 95% of residuals between -2 and +2 (Jackson et al., 2006). In this analysis, predicted value of accident frequency was plotted versus Pearson residuals as given in Figure (8). The figure shows that 1105 out of 1164=95% of the residuals fell within the recommended boundary (-2 and +2) which indicates a good model fit.

Figure 8: Residuals vs. predicted value of accident



frequency

5. SUMMARY AND CONCLUSION

Safety performance of signalized intersections and overall roadway network in Duhok city is evaluated based on analyzing contributory factors to road accidents. The approach used in this study involved collecting accident data using traffic police reports and field survey to collect geometric characteristics and traffic volumes for a descriptive study and statistical analysis. A summary of study findings is presented in the following sections.

5.1 Accident distribution-general trends

Based upon accident distribution and general trends of data from (2017-2021) for all the 22 intersections under study, signalized intersections and non-intersection site types it has been found that:

1. Distribution of accidents by type indicates that rear-end, angle and side collision comprise a large proportion of accidents. It also found that large portion of right-angle accidents (37%) and rear-end accidents (31.7%) occurred at all intersections and particularly at signalized intersections with 36.8% right angle accidents and 36.4% rear-end accidents.

2. Distribution of accidents by severity level for three site types (total intersections, signalized intersections and non-intersection) shows that, the majority of

accidents are of property damage only followed by injuries with fatal ones being the least.

3. Distribution of accidents by road surface moisture condition and lamination condition for all three site types indicates that, for all data sets, most accidents occurred in dry surface conditions than wet conditions and for all data sets, most accidents occurred during day time than night time.
4. Distribution of accidents by probable cause from police accident reports reveals that driver behavior is a major contributory factor to accident occurrence. Majority of accidents involved driver inattention, followed by speeding then red light running.
5. Generally, approaches with high traffic volumes (> 10000 veh/day) are associated with higher accident frequency than other traffic categories except at very high volume approaches (> 25000 veh/day). Furthermore, accident frequency decreases as the approach width increases beyond 15m.
6. Generally, accident frequency increases with the increase in clearance distance.
7. Shared lanes are present at 50% of approaches in the sample set and these approaches are associated with 46% of accident frequency.

5.2 Statistical Analysis

Geometric characteristics, driver behavior and other factors that influence accident occurrence at signalized intersections have been identified through analysis of accident data by GEE with NB and log link function. The results indicate that:

1. Not all variables had a statistically significant impact to describe the variation in accident frequency.
2. The main effect of approach width indicates that generally wider approaches (≥15m) tend to have higher accident frequency than narrower approaches as indicated by the positive coefficient of 0.95.
3. The significant effect of surface moisture condition with negative coefficient of -2.795 implies that wet surface had lower accident frequency than dry surface. This finding is expected to be attributable to exposure with drivers generally preferring to drive during dry surface condition and avoiding accident occurrence may play some part in their decision.
4. The effect of driver behavior showed that the probability of accident occurrence is more likely when drivers are inattentive and run the red light signal as compared to speeding.
5. Negative Binomial Model evaluation confirms that the model fits the data well.

Generally, the findings of this study confirm the importance of considering the effects of geometric characteristics on safety performance of signalized intersections. It is also confirmed that regulation of road safety with effective enforcement is necessary to encourage safer behavior by road users. It can be concluded GEE with log of link function accounts for the correlation among the repeated observations and provide more accurate estimate of model parameters than Generalized Linear Model. These results also highlights the importance of considering interaction effects in accident analysis.

Acknowledgement

The authors express their appreciations to the different road agencies of Duhok city/Iraq for providing all data and relevant information for this study.

REFERENCES

ABDEL-ATY, M. & WANG, X. 2006. Crash estimation at signalized intersections along corridors: analyzing spatial effect and identifying significant factors. *Transportation Research Record,* 1953**,** 98-111.

AL-ATAWI, A. M. 2014. Characteristics of red light running violations in urban areas in Tabuk, Kingdom of Saudi Arabia. *IATSS research,* 37**,** 119-123.

AL-REESI, H., GANGULY, S. S., AL-ADAWI, S., LAFLAMME, L., HASSELBERG, M. & AL-MANIRI, A. 2013. Economic growth, motorization, and road traffic injuries in the Sultanate of Oman, 1985–2009. *Traffic injury prevention,* 14**,** 322-328.

ALGHAFLI, A., MOHAMAD, E. & AHMED, A. Z. 2021. The effect of geometric road conditions on safety performance of Abu Dhabi road intersections. *Safety,* 7**,** 73.

BAGULEY, C. 1988. 'Running the red'at signals on high-speed roads. *Traffic engineering and control,* 29**,** 415-420.

CHIN, H. C. & QUDDUS, M. A. 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *accident analysis & prevention,* 35**,** 253-259.

DONG, C., CLARKE, D. B., YAN, X., KHATTAK, A. & HUANG, B. 2014. Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections. *Accident Analysis & Prevention,* 70**,** 320-329.

DONNELL, E. T., PORTER, R. J. & SHANKAR, V. N. 2010. A framework for estimating the safety effects of roadway lighting at intersections. *Safety science,* 48**,** 1436-1444.

FINCH, D., KOMPFNER, P., LOCKWOOD, C. & MAYCOCK, G. 1994. Speed, speed limits and accidents. *TRL project report*.

FITZPATRICK, K., SCHNEIDER IV, W. H. & PARK, E. S. 2006. Operation and safety of right-turn lane designs. *Transportation research record,* 1961**,** 55-64.

FRAWLEY, W. E. & EISELE, W. L. 2004. Investigation of access point density and raised medians: Crash analysis and micro-simulation.

HILBE, J. M. 2011. *Negative binomial regression*, Cambridge University Press.

KAMYAB, A. 2000. Red light running in Iowa: the scope, impact, and possible implications.

KEAY, K. & SIMMONDS, I. 2005. The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident analysis & prevention,* 37**,** 109-124.

KLAUER, C., DINGUS, T. A., NEALE, V. L., SUDWEEKS, J. D. & RAMSEY, D. J. 2006. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.

LIANG, K.-Y. & ZEGER, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika,* 73**,** 13-22.

MCDONALD, J. W. 1953. Relation between number of accidents and traffic volume at divided-highway intersections. *Highway Research Board Bulletin,* 74**,** 7-17.

MILTON, J. & MANNERING, F. 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation,* 25**,** 395-413.

OGDEN, K., NEWSTEAD, S., RYAN, P. & GANTZER, S. 1994. Factors affecting crashes at signalised intersections. *Monash University Accident Research Center Report*.

OTHMAN, S., THOMSON, R. & LANNÉR, G. Identifying critical road geometry parameters affecting crash rate and crash type. Annals of advances in automotive medicine/annual scientific conference, 2009. Association for the Advancement of Automotive Medicine, 155.

PERSAUD, B., LORD, D. & PALMISANO, J. 2002. Calibration and transferability of accident prediction models for urban intersections. *Transportation Research Record,* 1784**,** 57-64.

RETALLACK, A. E. & OSTENDORF, B. 2020. Relationship between traffic volume and accident frequency at intersections. *International journal of environmental research and public health,* 17**,** 1393.

SAWALHA, Z. & SAYED, T. Statistical issues in traffic accident modeling. 82 nd Annual Meeting of the Transportation Research Board, 2003.

SCHATTLER, K. L. & DATTA, T. K. 2004. Driver behavior characteristics at urban signalized intersections. *Transportation research record,* 1862**,** 17-23.

SHAMS, Z., NADERI, H. & NASSIRI, H. 2021. Assessing the effect of inattention-related error and anger in driving on road accidents among Iranian heavy vehicle drivers. *IATSS research,* 45**,** 210-217.

TREAT, J., TUMBAS, N., MCDONALD, S., SHINAR, D. & HUME, R. 1979. Tri-level study of the causes of traffic accidents. executive summary.

TURNER, S., SINGH, R. & NATES, G. 2012. *Crash prediction models for signalised intersections: Signal phasing and geometry*.

WANG, X. & ABDEL-ATY, M. 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis & Prevention,* 38**,** 1137-1150.

YAN, X., RADWAN, E. & ABDEL-ATY, M. 2005. Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis & Prevention,* 37**,** 983-995.