Journal of Engineering Faculty

| cumfad.cumhuriyet.edu.tr | Founded: 2023 Available online, ISSN Publisher: Sivas Cumhuriyet Üniversitesi

Uncovering the Complex Causal Mechanisms of Road Traffic Collisions at Intersections Using Piecewise Structural Equation Modelling

Mustafa Ekmekçi[1,a,*], Nima Dadashzadeh[2,b], Lee Woods[3,c], Renan Sinanmis Balci[4,d]

[1] Assistant Professor, Civil Engineering Department, Faculty of Engineering, Çanakkale Onsekiz Mart University, 17100 Merkez, Çanakkale, Türkiye
[2]Assistant Professor, Future Mobility Centre, Huddersfield Business School, University of Huddersfield, Huddersfield HD1 3DH, United Kingdom
[3]Associate Dean for Students, Faculty of Technology, University of Portsmouth Portsmouth, PO1 3AH, United Kingdom
[4]Assistant Professor, Nevşehir Haci Bektaş Veli University 2000 Evler, Zübeyde Hanım Cd., 50300 Merkez, Nevşehir Türkiye
**Corresponding author*

Research Article

History

Received: 08/07/2024
Accepted: 10/07/2024

ABSTRACT

Understanding the causes of traffic collisions is crucial for road designers, engineers, and policymakers to improve road safety at intersections. Design standards aim to minimize the severity and frequency of collisions. However, the factors that may affect traffic collisions are extensive. Their causal mechanisms can be complex, with feedback loops between traffic flows, visibilities, speeds, risk perception, speed limits, and other geometric characteristics of intersections. Structural Equation Modelling (SEM) is commonly used in behavioural sciences to understand complex causal paths, including travel behaviour studies. However, SEMs cannot robustly represent non-normally distributed datasets and rare count events, and little literature exists on their application to road traffic collisions. To address this limitation, this paper proposes a piecewise Structural Equation Modelling (pSEM) technique, which can handle count responses (i.e. number of collisions) to represent the complex causal relationships that lead to collisions. Application of pSEM technique is compared with conventional SEM. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) values demonstrate that pSEM is a more robust approach to model collisions at unsignalized intersections than conventional SEM. In terms of prediction ability referring to explained variance, pSEM is much more robust than SEM. Piecewise Structural Equation Modelling is, therefore, recommended for policy implications.

Keywords: Road safety; Traffic collision analysis; Piecewise Structural Equation Modeling (pSEM); Intersection design; Priority three-armed intersections

[a] *mustafa.ekmekci@comu.edu.tr* *0000-0003-2636-1739*
[b] *n.dadashzadeh@hud.ac.uk* *0000-0001-5425-0572*
[c] *lee.woods@port.ac.uk* *0000-0002-8336-4939*
[d] *renansinanmis@gmail.com* *0000-0002-4054-9052*

How to Cite: Ekmekçi M, Dadashzadeh N, Woods L, Sinanmis Balci R (2024) Uncovering the Complex Causal Mechanisms of Road Traffic Collisions at Intersections Using Piecewise Structural Equation Modelling, Journal of Engineering Faculty, 2(2): 125-136

Introduction

A wide range of research has been undertaken worldwide to understand and address road traffic collisions at unsignalized intersections. In the United Kingdom, priority-controlled intersections account for 49% of urban accidents, with T or staggered intersections accounting for 59% of these [1]. Many people were killed in these collisions, of which 66% were drivers, cyclists, or motorbike riders, 19% were passengers, and 15% were pedestrians [1].

Collision prediction models (CPMs) serve a variety of purposes, including estimating the anticipated number of collisions based on traffic volume and identifying factors related to the occurrence of collisions, such as geometric, environmental, and operational variables. To accurately determine the frequency and severity of collisions and reduce their likelihood, it is essential to comprehend the causal mechanisms that connect road conditions, environmental factors, drivers, vehicles, temporal variables, and operations to collisions [2–5]. By proactively utilizing CPMs, potential collision risks can be assessed, and collision-prone locations can be identified [6,7].

CPMs have evolved over decades from an early approach with multivariate linear regression analyses of traffic collisions to the most used method; generalised linear regression modeling (GLM) and most recently, machine learning. The logic behind this evolution lies in the nature of traffic collisions that are rare and non-normally distributed count responses. GLM seems convenient; however, it cannot overcome complex hierarchical interplay between explanatory variables, such as in the case of collisions. To represent such complexity, the conventional structural equation model using multivariate linear regression can be used using AMOS SPSS or other analytical software. However, the model treats the count response as a continuous variable that will cause unreliable results due to the nature of collisions [8]. On the other hand, the machine learning approach can handle complex problems, but interpreting results are problematic, due to their "black box" nature. Therefore, there is a need to have a new analytical regression method that can combine the structural equation model for representing complex and hierarchical interplay and GLM that accounts for non-normally distributed collisions.

This paper developed a piecewise Structural Equation Model (pSEM) to road safety as a new analytical method, overcoming drawbacks of previous techniques. The new pSEM model is compared with the conventional SEM by analysing traffic collisions at 120 priority intersections in the city of Portsmouth, UK.

The remainder of this paper begins by summarising the current literature's findings regarding collision prediction models at unsignalized intersections. Commonly used regression models relevant to this paper are described, along with their advantages and disadvantages. Then, the mathematical background of path analysis as a family of structural equation modeling is explained. In the method section, after briefly describing the data for this paper, piecewise Structural Equation Modeling is formulated, and a causal diagram of the model is illustrated compared to a standard SEM. The results are discussed and concluded.

Literature Review

Studies on unsignalized intersections

Road traffic collisions (RTC) at unsignalized intersections are a major concern around the world, and several approaches have been used to better understand their causes.

The approaches on modeling traffic collision of priority three-armed intersections can be broadly categorized into three main categories: observational research, experimental research and microsimulation studies as illustrated in Figure 1.

The first category is observational research, which can be further divided into three sub-groups. The first group aims to determine the frequency or number of collisions. For example, Kitto [9] conducted a study in Wellington, New Zealand to examine the effect of physical and site characteristics on collisions. Similarly, Bonneson [10] examined unsignalized intersections in the USA (Utah, Minnesota, Illinois, etc.) to develop a model relating intersection traffic demand to collision frequency. As collisions are rare events and non-negative count data, the most commonly used approach to modeling RTCs was Generalized Linear Modeling (GLM). Other studies have developed collision prediction models based on traffic flow to assess safety at unsignalized intersections, such as Salifu [6], who studied two cities in Ghana, and Kulmala [11], who developed separate models for different types of collisions in Finland.

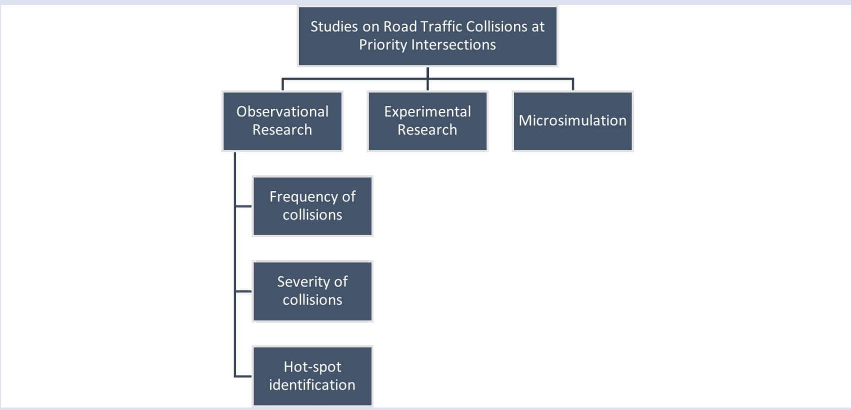

Figure 1. Categorization of studies on priority three-armed intersections.

The second approach to understanding RTCs is to determine the severity of collisions rather than the frequency. For instance, Chen et al. [12] applied logistic regression to examine the contribution of various factors to collision severity in Victoria, Australia, while Haleem and Abdel-Aty [13] examined the severity of injury collisions at unsignalized intersections in the USA. Both studies found that several variables, such as traffic volumes, the number of lanes, and driver age, were significant in determining collision severity.

Statistical models are useful for drawing conclusions and establishing relationships, but they generally have poor predictive performance due to assumptions about the collision data [14]. Researchers have used artificial neural networks (ANNs) to classify collision severity, such as Akin and Akbas [15], who predicted intersection collisions in Michigan State, USA.

The third approach to traffic safety at junctions is hot-spot identification. Sayed and Rodriguez [7] attempted to develop collision prediction models to identify collision-prone locations of urban unsignalized intersections on Vancouver Island, British Columbia, using the Empirical Bayesian (EB) method.

Category two includes experimental research, such as Polus [16] who implemented a before-and-after study to investigate whether stop signs differ from yield signs in terms of their effect on traffic collisions at unsignalized intersections.

Category three includes microsimulation studies based on traffic conflict techniques developed by General Motor Laboratory. Traffic conflicts refer to observed situations where two or more road users approach a collision course, which could result in a collision if no one acts to avoid it. Traffic conflict indicators, such as time to collision (TTC) and post-encroachment time (PET), can be used to assess traffic safety. Microsimulation software, such as VISSIM and AIMSUN, can be used with the Surrogate Safety Assessment Model (SSAM) to classify whether PET and TTC have critical conflict values.

Caliendo et al. [17] attempted to predict the number of collisions by using the number of critical traffic conflicts predicted. Pawar, Gore, and Arkatkar [18] investigated the use of accepted gaps to estimate the likelihood of a collision at unsignalized T-intersections. The best-fitting distribution was the generalised extreme value (GEV), which was used to estimate the likelihood of a collision using accepted gap and PET datasets. Accepted gap data was used for risk characterisation. Goyani [19] looked at the safety effect of mixed traffic conditions on different types of vehicles. The percentage of critical crossing conflicts (PCCC) was higher in the traffic stream with motorised two-wheelers and three-wheelers, followed by cars, buses, LCVs, and trucks. Paul [20] figured out how motorised two and three-wheelers conflicts affect crossing conflicts at unsignalized T-intersections in India. They applied Truncated Negative Binomial regression to create models of crossing conflicts. The results showed that the number of critical and non-critical conflicts is greatly affected by the proportion of two and three-wheelers in the conflicting and offending stream, the presence of a central traffic island, and the total number of conflicts.

Studies using microsimulation have also examined the effect of changes in speed limits on traffic safety at unsignalized intersections, such as Pirdavani [21], who used S-Paramics to analyse whether changes in speed limits under different traffic volumes affect traffic safety. Srinivasula, Chepuri, and Joshi [22] investigated the critical speed of conflicting vehicles and the effect of speed bumps on PET and traffic collision frequencies. They concluded that while safety has improved, other appropriate measures should be implemented to improve safety levels further.

Surrogate safety measures traditionally applied to single collision types. Gastaldi, Orsini, Gecchele, and Rossi [23] applied two bivariate extreme value theory (EVT) approaches in order to simultaneously evaluate multiple collision types at a three-leg unsignalized intersection.

Commonly used regression models in traffic safety studies

Linear regression models

In the literature, early attempts at collision prediction studies were modelled by multiple linear regression that can be formulated as follows.

$$E = B_0 + \sum_{j=1}^{n} x_j b_j \quad (1)$$

Where E is the number or frequency of collisions, B_0 is an intercept, x_j represents explanatory variables such as road width and traffic flow, n is the number of explanatory variables, and b_j are the estimated coefficients.

Kitto [9] examined the effect of various physical and site characteristics on collisions by using multiple regression techniques. Arndt and Troutbeck [24] described the relationship between the geometry of priority intersections and collision rate using multiple linear regression. However, there are many debates and controversial points regarding this modeling technique. Nambuusi [25] highlighted that multiple linear regression is not robust as the response variable is non-negative count data. The linear regression model, while appealing and simple, has stringent assumptions that are frequently violated when applied to data on road safety [26].

Generalized linear regression models (GLMs)

To overcome the limitations of conventional linear regression, GLM was used. These models allow the analysis of non-normally distributed data [27].

Poisson regression

Different generalized linear models have been established in the literature. Poison GLM regression is one of the commonly used models since collisions are unavoidably discrete and rare random events [25]. Rahman [28] indicated that another advantage of Poisson regression is that it is suitable for categorical data.

The commonly used form of Poisson regression with a log link function can be written as follows.

$$E = t_i e^{\Sigma bjXj} \quad (2)$$

Where E is the expected event, bj are the estimated coefficients, t_i is the interval as an offset variable (i.e. time, length), Xj represents an explanatory variable, and e is the natural number.

This function above specified traffic collisions by adding traffic flow as the main exposure variable. The following structure is a general form used in the literature [7,11,29].

$$E = t_i Q_1^{b1} Q_2^{b2} \quad (3)$$

Where Q_1 and Q_2 annual average daily traffic flows per 1000 vehicles b_1 and b_2 are the estimated parameters, and t_i is the offset variable (year).

What Q_1 and Q_2 depend on what kind of flow model was used. For example, in the major/minor inflow model, they are the sum of major and minor traffic flow at an intersection.

The following form was added to the model to include additional factors such as categorical and geometric variables [10,30,31].

$$E = Q_1^{b1} Q_2^{b2} e^{\sum bjXj + lnt_i} \quad (4)$$

The Quasi-Poisson and negative binomial regression

Poisson models also have some drawbacks. Turner [32] observed that the Poisson model sometimes produces inaccurate results. The limitation is that the mean must be equal to the variance [33]. When this equality is not evident, under-dispersion or over-dispersion problems occur. The pure Poisson model can be modified to deal with the over-dispersion problems through a Quasi-Poisson approach. The estimated parameters that result from the Quasi-Poisson (QP) model are the same as those from the pure Poisson model. The difference is just in their standard errors that are inflated by a factor of 'k' [30]

Another way to model an over-dispersed response variable is to use the negative binomial (NB) regression (30)The assumption of the NB model is reported to be a more realistic approach than the Poisson regression [34]. Several researchers have attempted to use NB regression to deal with dispersion problems [i.e., 6,7,11,35].

Theoretically, the variance ($Var(n)$) in the negative binomial distribution is not equal to the mean (λ) as follows:

$$Var(n) = \lambda + \frac{1}{r}\lambda^2 \quad (5)$$

The term $\frac{1}{r} = \theta$, which refers to the inverse of the over-dispersion parameter (r). In the mathematical limit situation where ' r ' is close to infinity, the variance will equal the mean due to the standard Poisson model assumption as follows.

$$Var(n) = \lambda + \frac{1}{\infty}\lambda^2 = \lambda \quad (6)$$

Conventionally, Poisson regression is a particular instance of the negative binomial model where the θ parameter is zero [36]. The overall form of negative binomial regression is derived from the Poisson model by adding an independently distributed error term [37].

$$\lambda_i = e^{\sum bjXj + b_0 + \mathcal{E}_i + lnt_i} \quad (7)$$

Where exp ($\mathcal{E}_i$) is the gamma distributed error with mean 1 and variance 1/r=θ.

$$P(Y = y_i|\lambda_i, \theta) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta y_i)} \left[\frac{\theta}{\theta + y_i}\right]^{\theta} \left[\frac{\lambda_i}{\theta + \lambda_i}\right]^{y_i} \quad (8)$$

Where $y_i = 0, 1, 2, \ldots, n$ is the number of collisions, θ is the inverse dispersion parameter (1/r), and Γ (...) is the gamma distribution function.

Poch and Mannering [37] concluded that the elements related to collision frequency in an intersection approach could logically identify significant traffic and geometric phenomena by using negative binomial regression modeling.

The over-dispersion parameter as an extra variation in traffic collisions can be caused by some factors, such as misspecification in the model and an excessive number of zero collisions. In this regard, the zero-inflated Poisson or the zero-inflated negative binomial models can be used [38, 39].

Structural equation model (SEM): A path analysis

In the 1920s, path analysis was developed by Sewall Wright, a geneticist. He attempted to measure the direct influence of one variable on another along with path diagrams by taking correlations into account. The path analysis method, therefore, depends on the degree of correlations in a system [40].

Wright [41] obtained the basic formulation of path coefficients by explaining the representative path diagram illustrated in Figure 2. Variables are connected as a function of dependent relationships through the one-way arrows. Vi is a total residual determination, and the double arrows represent residual correlations between variables. Wright assumed that all relationships are linear.

Each variable from the unidirectional perspective can be formalised as follows:

$$(V_0 - V_0') = r_{01}(V_1 - V_1') + r_{02}(V_2 - V_2') + \cdots + r_{on}(V_n - V_n') \quad (8)$$

Where V_i' is the mean, $(V_i - V_i')$ represents deviations from means, and r_{oi} is a coefficient.

$$X_i = \frac{(V_i - V_i')}{\delta_i} \quad (9)$$

Where X_i is a standard z score of a variable, and δ_i is the standard deviation.

$$P_{0i} = r_{0i}\frac{\delta_i}{\delta_0} \quad (10)$$

P_{0i} is a standardised path coefficient reflecting a correlation.

Equation 9 and Equation 10; $X_i\delta_i = (V_i - V_i')$, $P_{0i}(\delta_0/\delta_i) = r_{0i}$ can be rearranged respectively. Then, they can be combined with Equation 11 as follows:

$$X_0\delta_0 = P_{01}\frac{\delta_0}{\delta_1}X_1\delta_1 + P_{02}\frac{\delta_0}{\delta_2}X_2\delta_2 + \cdots + P_{0n}\frac{\delta_0}{\delta_n}X_n\delta_n \quad (11)$$

So, if the unnecessary parameters in the equation above are eliminated, the final form of the equations will be as follows:

$$X_0 = P_{01}X_1 + P_{02}X_2 + \cdots + P_{0n}X_n \quad (12)$$

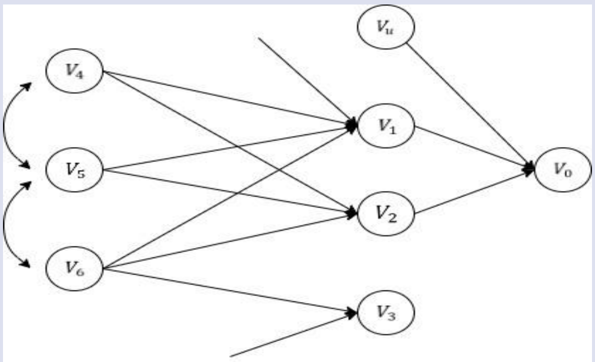

Figure 2. Path Analysis Diagram, (Wright, 1934)

According to Wright [41], the fraction of the standard deviation of the dependent variable in the above equation is measured by each path coefficient. The factor identified will be directly responsible for a change in the fraction of the dependent variable even if all other factors, including residuals are constant [41].

SEM is a multivariate technique that combines regression, factor analysis and analysis of variance in order to simultaneously estimate interrelated dependence relationships [42]. Even though conventional SEM methodology is well-known and widely applied in many different fields of research, nowadays, there are no robust, practical applications for collision prediction.

SEM used multivariate linear regression that does not account for the nature of collision data, which are rare, non-negative counts and may not be normally distributed with respect to explanatory variables. The linear regression of a count response variable may not present meaningful results. As explained above, generalised linear modeling (GLM) approaches taking account of discrete responses would be more logical. Regarding traffic collisions at unsignalized intersections, such approaches have been applied in the UK [30,43] and overseas [i.e., 7, 10, 11, 29, 31, 34, 37] However, GLM may not be able to represent the complex inter-relationships between explanatory variables such as speeds, visibilities, speed limits and collisions. Therefore, in this paper, a piecewise structural equation model (pSEM) adapting GLM into a structural equation model will be introduced and compared with conventional SEM by analysing collisions at 120 priority intersections in the city of Portsmouth, UK.

Method and Data

Data collection in Portsmouth, UK

Portsmouth is a city located in Southeast of the UK. It is about 19 miles south of Southampton and 70 miles south of London. The city has the highest population density in the United Kingdom, with 5,100 people per square kilometres, significantly above the national average [44]. In 2007, Portsmouth was the first city in the UK to introduce a city-wide 20mph speed limit [45].

The locations where data collection was conducted shown in Figure 3 were collected from 120 locations in Portsmouth and Gosport. Locations were reviewed between Sep 2020 and Aug 2021.

The data collection process summarized in Figure 4. Speed data from approaching vehicles at 360 junction arms were measured by using an SL700 spot speed camera. Classified 15-minute traffic counts were undertaken in the am, pm or both. These were converted to Average Annual Daily Traffic (AADT) using the standard conversion factors suggested by the Chartered Institution of Highways and Transportation [46].

As input variables, geometric factors (such as visibility, road width, and turning radii) are measured. To measure the geometric factors, the DIGIMAP service in the United Kingdom provides AutoCAD-compatible ".dwg" files. However, actual conditions, such as parking situations, could not be visualized in the AutoCAD drawing. As a result, visibility measurements were drawn in 3D using Google Earth. All junctions were visited to see if there were any discrepancies between Google Earth views and current junctions.

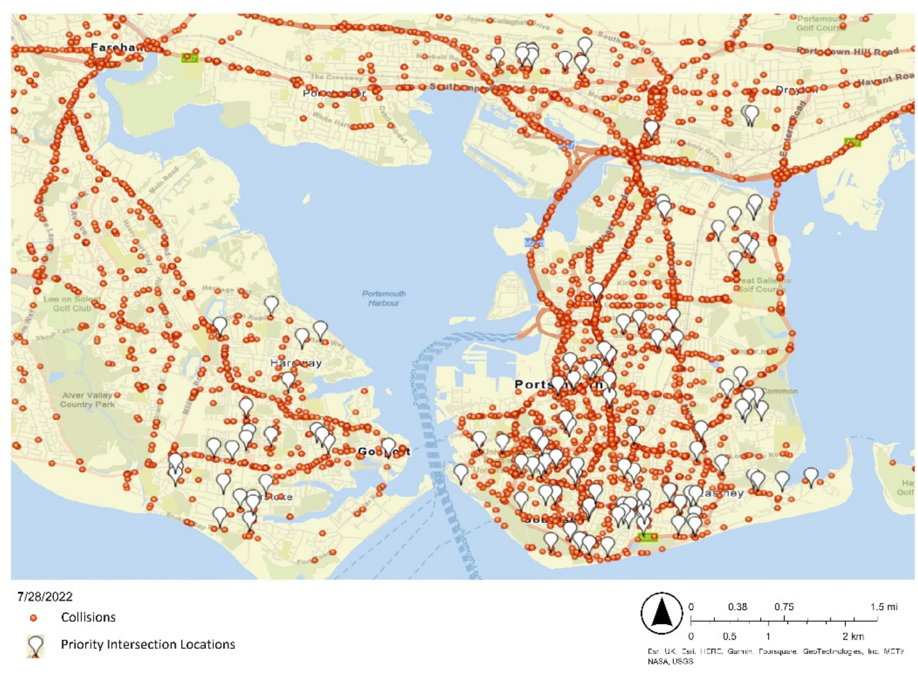

Figure 3. Locations for the data collection [47]

As output variables, collision scenarios collected from STATS19 data provided by Department of Transport. In the UK, each accident has specific index number linking vehicle and collision information. These two massive datasets combined by R Studio, STATS19 package. Then it was put into ArcGIS online to see the collisions in the target locations (Figure 3).

Piecewise Structural Equation Modeling (pSEM)

pSEM was developed by Shipley's studies in the 2000s [48]. The relationships between variables in the conventional SEM are simultaneously estimated in a single variance-covariance matrix [49]. However, in pSEM, each set of relationships are estimated locally (or, in other words, independently) [48]. The conventional SEM uses chi-square tests to compare observed and predicted covariance matrix, while the goodness-of-fit of a pSEM is measured as a “test of directed separation” [50]. Shipley [48] highlighted that this test does not depend on asymptotic methods, so it can be used with smaller sample sizes. Secondly, many problems involving non-normality and non-linearity can be overcome. Another aspect of the “test of directed separation” is that it does not conflict with some problems of covariance matrices involving a maximized loss function. Finally, it allows each piece of the local causal model to determine how much it contributes to the lack of fit [51]. Therefore, pSEM appears to be a robust technique for estimating collisions. R Studio was selected as the analytical package, as it can run with the “piecewiseSEM” package.

The d-separation test of pSEM

Shipley’s method of testing the path model is based on graph theory notation called directed separation (d-separation) and its relation to conditional independence claims in the probability distribution developed by Pearl and Verma [52]. They explained the logic and mechanics of the d-separation test. Shipley [48] generalised and adapted this to deal with data with a hierarchical structure.

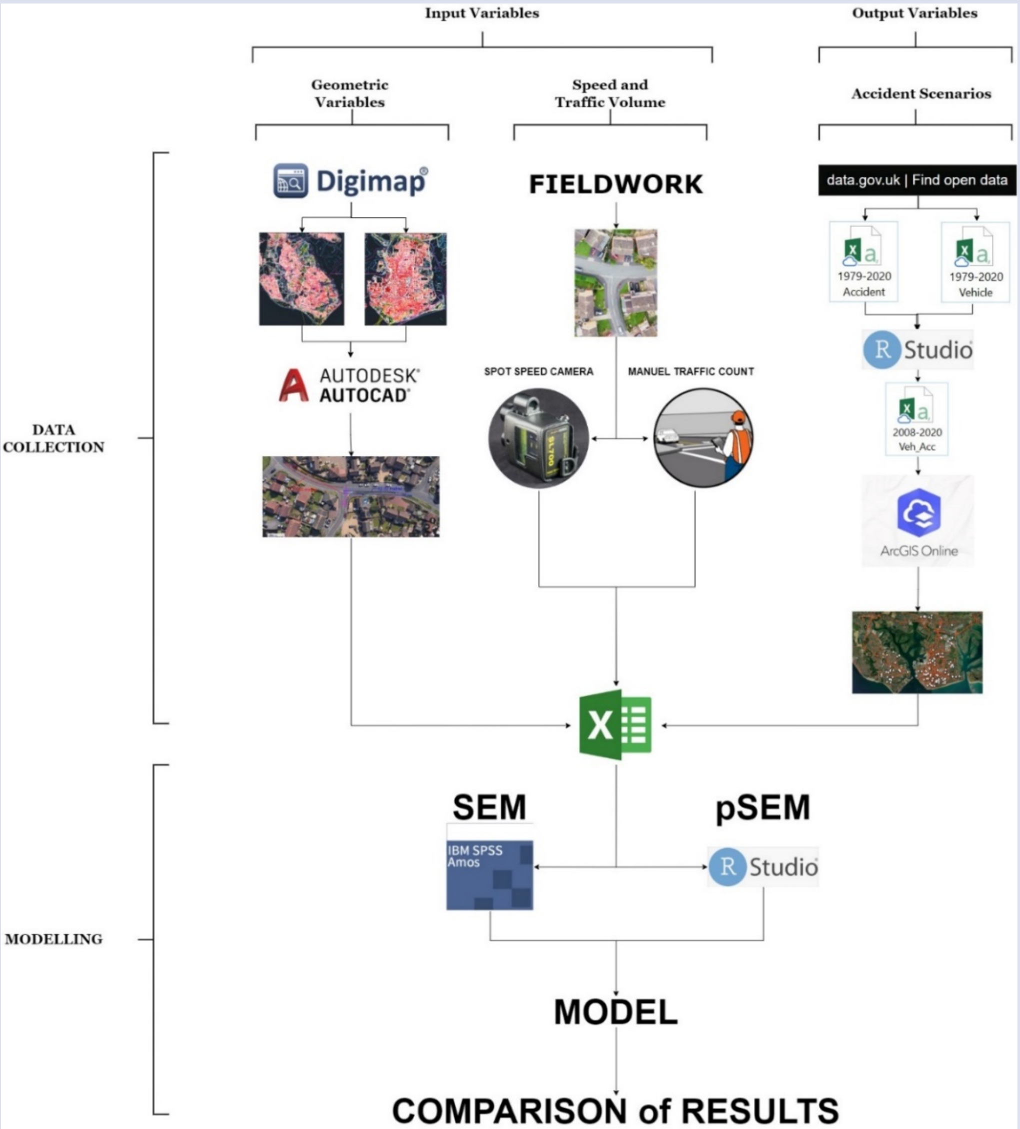

Figure 4. Data collection and analysis process

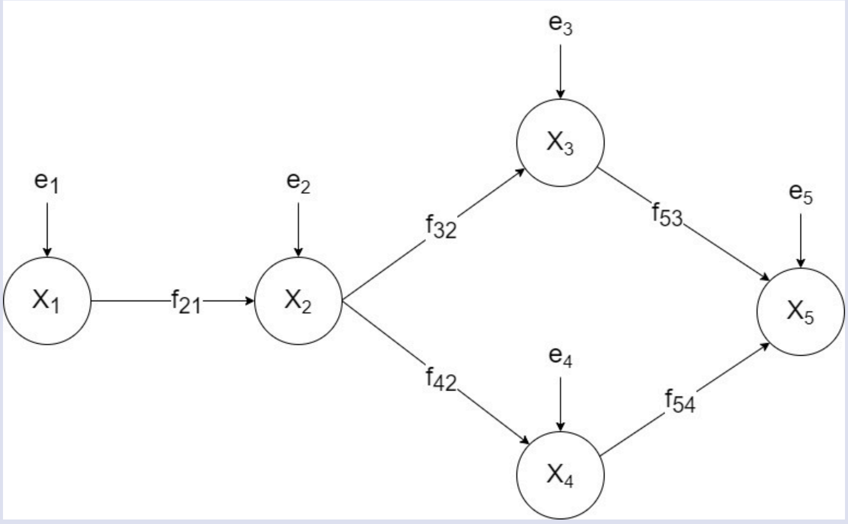

Figure 5. An illustration of a causal path model (directed acyclic graph).

Table 1. d-separation for independence claims

Claim No	d-separation claim
1	X1 ⊥ X3 | {X2}
2	X1 ⊥ X4 | {X2}
3	X1 ⊥ X5 | {X3, X4}
4	X2 ⊥ X5 | { X1, X2, X4}
5	X3 ⊥ X4 | {X2}

The mathematical background of the d-separation test was introduced in Shipley [51] and detailed in Shipley [48,53]The process can be summarised in 6 steps as follows.

1. First, the causal hypothesis should be formed in the directed acyclic graph showing the variables involved in the hypothesis and directed causal links between each other as arrows. It can be seen in Figure 5 as an example where Shipley [48] indicates the functional form of the equations (“f_{ij}”) that link variables between i and j.
2. Identify each pair of variables (X, Y) in the graph for variables that do not have an arrow from one direction to another, such as X1→ X3 or X3←X1.
3. Find the causal parents that are mediating the links. For instance, the causal parent of the relationship between X1 and X3 is X2. Because there is no direct arrow between X1 and X3, this set of variables (X1 →X2 ←X3) is called the conditioning set.
4. Create an independence claim by converting each unique pair and conditioning set. The independent claims of the directed acyclic graph in Table 1 are listed below as an example.
5. Xi ⊥ Xj | Q, means that Xi and Xj variables are a probabilistically independent conditional set of variables in Q.
6. Calculating the null probability (pi) for each predicted independence claim.
7. In the final step, a global test of the model requires simultaneous testing of these independence relationships given by Fisher [54]. Fisher’s C test can be written as follows.

$$C = -2\sum_{j=1}^{k} ln(p_j) \quad (12)$$

The test follows the chi-square distribution with 2k degrees of freedom. Where k is the number of independent claims. This statistic also can be used to compute an AIC score for comparisons in the model selection process [53].

$$AIC_c = C + 2K \quad (13)$$

Where C is Fisher’s C statistics, K is the likelihood degrees of freedom.

The pSEM model in Figure 6 was specially designed in line with the literature. Arrows from any variable into the response variables (4 types of crashes) were modelled by the GLM approach of Poisson regression with log link functions. In contrast, multiple linear regression under the pSEM model simultaneously modelled other connections with normal distributions [55,56].

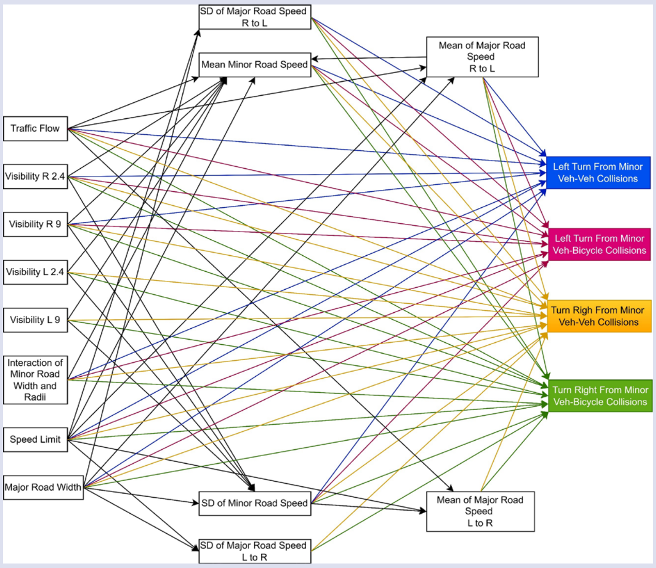

Figure 6. Directed acyclic graph of the fully tested model.

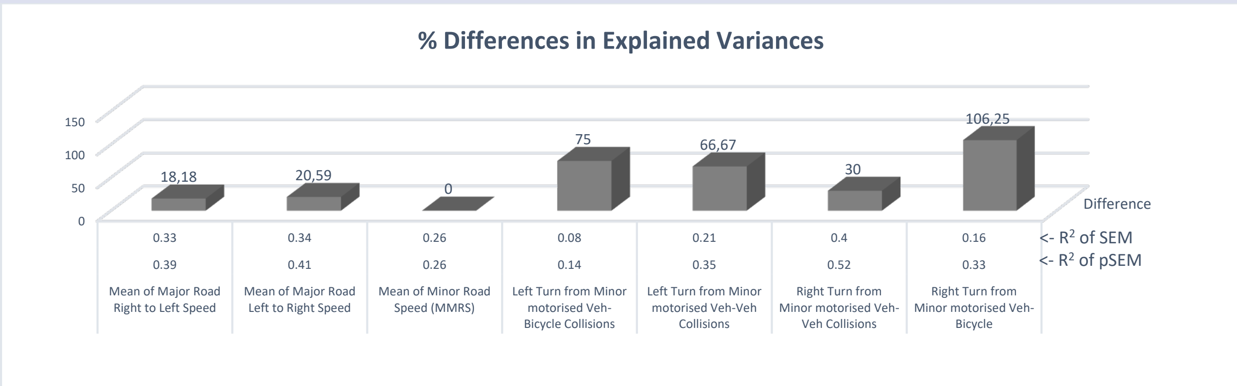

Figure 7. Comparing R-squared values of response variables for piecewise SEM and conventional SEM.

The associations between speed variables (mean and standard deviation) and other factors were modelled with pSEM through multiple linear regression, as speeds are dependent variables. Kolmogorov-Smirnov and Shapiro-Wilk tests were used to check the normal distribution assumptions.

Results and Discussion

The results of this study indicate that the pSEM model showed better performance, especially for predicting collision frequencies. Overall, the collision frequencies of pSEM explained 57% more variance compared to those in the conventional SEM, as the differences were summarised in Figure 7.

Differences in explained variance vary through the responses. Much greater differences have been observed where the responses are count variables (collisions) compared to the response variables that are continuous. This demonstrates the limited capability of linear regression to model traffic collisions. Even where the responses (mean speed) were suitable (continuous) for multivariate linear regression through SEM and pSEM models, the piecewise structural equation model generally shows better performances with 18.18% and 20.59% differences. All of these differences seem to depend on how estimations are done, and other factors as listed below:

- pSEM estimates locally while SEM estimates simultaneously [48].
- Whether it is an asymptotic method or not; pSEM does not depend on asymptotic method and gives better estimation performance with smaller sample sizes [48].
- The distribution of response variables

Table 2 shows the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) values. The BIC and AIC are used to assess model performance. Model performance is measured by taking complexity into account. They include a term that assesses how well the model fits the data with a term that penalises the model based on how well it fits the data, dependent on the number of parameters in the model [57]. The same model structure in Figure 6 was used to compare the model fit. The SEM model performed better in terms of the AIC and BIC values than SEM.

Table 2. Comparison of model fit values

Model	AIC	BIC
pSEM	176.291	254.341
SEM	267.642	387.505

Table 3. The results of the directed separation test

PREDICTORS	Mean of Minor Road Speed (MRMS)	Left Turn from Minor Veh- Bicycle Coll.	Right Turn from Minor Veh- Veh Coll.	Right Turn from Minor Veh-Bicycle Coll.	Mean of Major Road Left to Right Speed	Left Turn from Minor Veh- Veh Coll.	Mean of Major Road Right to Left Speed
Traffic Flow	0.287	0.234	-	-	-	-	-
Speed Limit	-	0.286	0.323	0.942	-	-	-
Minor Rd. Width	-	0.751	0.883	0.316	0.513	0.973	-
SD of Minor Rd. Speed	-	0.917	0.743	0.523	0.504	-	0.727
Vis. Right 2.4	0.234	0.348	-	0.882	0.090	0.086	0.109
Vis. Left 2.4	0.454	0.511	-	-	-	0.353	-
Vis. Right 9	0.705	0.141	0.765	-	0.093	0.850	0.070
Vis. Left 9	0.591	0.539	0.931	-	0.011	0.792	0.044
Mean of Major Road Right to Left Speed	0.253	0.136	0.349	0.946	-	0.497	-
Mean of Major Road Left to Right Speed	0.068	0.319	-	0.599	-	0.85	-
MRMS	-	-	0.017	-	-	0.752	-
Left Turn from Minor Veh-Veh Collisions	-	0.595	0.463	0.253	-	-	-
Right Turn from Minor Veh-Veh Coll.	-	0.45	-	0.857	0.101	-	-
Left Turn from Minor Veh-Bicycle Coll.	-	-	0.469	-	-	-	-

In terms of global goodness of fit, the SEM uses chi-square tests to compare observed and predicted covariance matrices. The null hypothesis of the chi-square test is that the covariance matrix of observation (∑) is equal to the predicted covariance matrix (∑(θ)). The P-value was less than 0.001 with a chi-square value of 181.642 and 48 degrees of freedom. Predicted-observed covariance matrices significantly differ from each other. Thus, the alternative hypothesis (∑ ≠ ∑(θ)) is accepted. In contrast, in the piecewise SEM, the goodness-of-fit is tested using a test of directed separation. Fisher’s C statistic is 120.3 with P-value = 0.279 and on 112 degrees of freedom. The model indicates that there is no significant need to improve model fit.

Lack of fit values of the directed separation test are shown in Table 3. The analysis revealed a few associations with high probability where p values are less than 0.05. For example, the relationship between the Mean of major road speed and left-hand visibility at 9 metres back (Vis. Left 9) is significant with a p-value = 0.012. However, one cannot reasonably infer a causal relationship between them. Minor road visibilities may be correlated to major road visibilities, which may moderate the mean speed on major roads. This connection and others were not added to the model because two reasons. Firstly, incorporating these relationships into the model is completely unnecessary. Secondly, Fisher's C statistic demonstrated that the model is adequate without the inclusion of new relationships.

The connections between non-significant predictors and response variable in the initial tested model (Figure 6) were step by step removed using the backward elimination method. Then, the results of the final pSEM and SEM models (remining connections) were presented in Table 4 including the estimated coefficients, standard errors and corresponding p-values of the models. “NA” represents that connections between predictor and response variables were removed due to the backward elimination. However, the latest model forms for pSEM and SEM do not match each other. Some variables were kept in one of the models, while these variables were removed from the other. On the other hand, some variables were removed entirely from both models. For the “right turn from minor motorised vehicle-bicycle collision” scenario in Table 4, SEM lost many response variables while pSEM kept them.

Table 4. Comparison of pSEM and SEM

			pSEM			SEM	
Response	**Predictor**	**Estimate**	**S.E.**	**P. Value**	**Estimate**	**S.E.**	**P. Value**
Mean of Major Road Right to Left Speed	Traffic Flow	0.0005	0.0001	0.00001	0.0010	0.0001	0.0001
Mean of Major Road Right to Left Speed	Speed Limit	3.9902	0.6910	0.00001	3.8610	0.6940	0.0001
Mean of Major Road Left to Right Speed	Traffic Flow	0.0005	0.0001	0.00001	0.0100	0.0001	0.0001
Mean of Major Road Left to Right Speed	Speed Limit	3.1690	0.6077	0.0001	3.0340	0.6060	0.0001
Mean of Minor Road Speed (MMRS)	Speed Limit	2.2128	0.5160	0.00001	2.2130	0.5120	0.0001
Mean of Minor Road Speed (MMRS)	Minor Rd. Width (MRW)	0.6378	0.2111	0.0031	0.6380	0.2090	0.0020
Left Turn from Minor Veh-Bicycle Collisions	MMRS	0.2116	0.0569	0.0002	0.0440	0.00001	0.0130
Left Turn from Minor Veh-Veh Collisions	MMRS	NA	NA	NA	0.0440	0.0130	0.0001
Left Turn from Minor Veh-Veh Collisions	SD of minor Rd. Speed	-0.9104	0.3819	0.0171	NA	NA	NA
Left Turn from Minor Veh-Veh Collisions	Vis. Right 2.4	NA	NA	NA	0.0010	0.0010	0.0250
Left Turn from Minor Veh-Veh Collisions	Traffic Flow	0.0002	0.0001	0.00001	0.0001	0.0001	0.0001
Left Turn from Minor Veh-Veh Collisions	Vis. Right 9	NA	NA	NA	-0.0020	0.0010	0.0310
Right Turn from Minor Veh-Veh Collisions	Traffic Flow	0.0002	0.0001	0.00001	0.0001	0.0001	0.0001
Right Turn from Minor Veh-Veh Collisions	MRW x Left Radius	0.0113	0.0031	0.0003	NA	NA	NA
Right Turn from Minor Veh-Veh Collisions	Vis. Left 2.4	-0.0035	0.0016	0.0293	-0.0010	0.0001	0.0160
Right Turn from Minor Veh-Veh Collisions	Vis. Right 2.4	0.0026	0.0013	0.0364	0.0020	0.0010	0.0180
Right Turn from Minor Veh-Bicycle Collisions	Traffic Flow	0.0002	0.0001	0.0685	NA	NA	NA
Right Turn from Minor Veh-Bicycle Collisions	Vis. Right 9	0.0174	0.0058	0.0024	0.0050	0.0010	0.0001
Right Turn from Minor Veh-Bicycle Collisions	Vis. Left 2.4	-0.0183	0.0061	0.0027	NA	NA	NA
Right Turn from Minor Veh-Bicycle Collisions	Vis. Left 9	0.0174	0.0058	0.0024	NA	NA	NA
Right Turn from Minor Veh-Bicycle Collisions	MMRS	0.1222	0.0558	0.0285	NA	NA	NA

Summary and Conclusion

The present study was designed to propose piecewise Structural Equation Modeling and compare it with the conventional SEM for understanding the causality behind road traffic collisions. Two analytical models were run for the same model structure. The global model fit value of conventional SEM could not pass the chi-square test leading to improper results, while the goodness of fit for pSEM did with Fisher's C statistic. Test of directed separation of pSEM also informed about other lack of fit for paths not included in the model. The pSEM model performed better, particularly in forecasting collision frequencies. Values for the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) used to evaluate model performance. When the same model structure was employed to examine model fit, the pSEM model outperformed the SEM model in terms of AIC and BIC values.

SEM employed multivariate linear regression, which ignores the nature of collision data, which are infrequent, non-negative counts that may not be normally distributed in relation to explanatory factors. So, the results of SEM show that it seems not to be a proper approach to model traffic collisions. As previously stated, GLM techniques that take the discrete nature of responses into account would be a more reasonable approach. GLM, on the other hand, seems to be incapable of representing the complicated interrelationships between explanatory factors such as speeds, visibility, speed limits, and collisions. Application of pSEM extends the generalised linear modeling approach one step further by allowing it a hierarchical model structure including intermediation and provides a new inferential test called the test of directed separation to measure model fits. This paper proposes that pSEM is a more robust approach for modeling collisions than conventional SEM.

The collision frequencies of pSEM explained an average of 57% more variation when compared to those of the conventional SEM. Differences in explained variance differ between responses. Where the responses are count variables, much greater variation has been seen (collisions). This reveals linear regression's poor capacity to model traffic collisions. However, even when the response variable (in this case mean speed) is suitable for multivariate linear regression, due to it being a non-count variable, the piecewise structural equation model consistently outperforms SEM with 18.18% and 20.59% differences.

The scope of this research is confined to a model structure developed to comprehend certain types of road traffic collisions at 120 priority three-arm intersections. More studies could be done to assess the efficiency of the piecewise Structural Equation Model at various types of intersection and collisions.

Acknowledgements

This work is fully funded by the Republic of Turkey Ministry of National Education, Çanakkale Onsekiz Mart University: Study Abroad Program.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Department for Transport. Road Safety Data - data.gov.uk [Internet]. 2023 [cited 2024 Jan 15]. Available from: https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

[2] Eenink R, Reurings M, Elvik R, Cardoso J, Wichert S, Stefan C. Accident Prediction Models and Road Safety Impact Assessment: recommendations for using these tools. Vol. 506184, RiPCORD iSEREST. 2005.

[3] Usman T, Fu L, Miranda-moreno LF. Quantifying safety benefit of winter road maintenance: Accident frequency modeling. Accid Anal Prev [Internet]. 2010;42(6):1878-87. Available from: http://dx.doi.org/10.1016/j.aap.2010.05. 008

[4] Sarkar A, Sahoo UC, Sahoo G. Accident prediction models for urban roads. Int J Veh Saf. 2012;6(2):149-61.

[5] Çelik AK, Oktay E. A multinomial logit analysis of risk factors influencing road traffic injury severities in the Erzurum and Kars Provinces of Turkey. Accid Anal Prev. 2014 Nov 1;72:66-77.

[6] Salifu M. Analysis of accident potential at unsignalised urban junctions in Ghana. 2002.

[7] Sayed T, Rodriguez F. Accident prediction models for urban unsignalized intersections in British Columbia. Transp Res Rec. 1999;(1665):93-9.

[8] Muthen LK. Mplus Discussion >> Count Data Within Path Models [Internet]. 2012 [cited 2022 Apr 26]. Available from: http://www.statmodel.com/discussion/messages/11/802.html?1562474031

[9] Kitto HJ. Accident rate at urban right-angle intersections. 1980.

[10] Bonneson JA, Mccoy PT. Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways. 1993.

[11] Kulmala R. Safety at rural three- and four-arm junctions. Development and application of accident prediction models. VTT Publications [Internet]. 1996 [cited 2021 Mar 25];(233). Available from: https://trid.trb.org/view/684994

[12] Chen H, Cao L, Logan DB. Analysis of Risk Factors Affecting the Severity of Intersection Crashes by Logistic Regression. http://dx.doi.org/101080/153895882011653841 [Internet]. 2012 May [cited 2022 Aug 19];13(3):300-7. Available from: https://www.tandfonline.com/doi/abs/10.1080/15389588.2011.653841

[13] Haleem K, Abdel-Aty M. Examining traffic crash injury severity at unsignalized intersections. J Safety Res. 2010 Aug;41(4):347-57.

[14] Zhou D, Gayah VV, Wood JS. Integration of machine learning and statistical models for crash frequency modeling. Transportation Letters. 2023;15(10):1408-19.

[15] Akin D, Akbas B. A neural network (NN) model to predict intersection crashes based upon driver, vehicle and roadway surface characteristics. 2010; Available from: https://www.researchgate.net/publication/228367341

[16] Polus A. Driver behaviour and accident records at unsignalized urban intersections. Accid Anal Prev. 1985 Feb 1;17(1):25-32.

[17] Caliendo C, Guida M. Microsimulation approach for predicting crashes at unsignalized intersections using traffic conflicts. J Transp Eng. 2012;138(12):1453-67.

[18] Mukund Pawar N, Ninad Gore;, Arkatkar S. Examining Crossing Conflicts by Vehicle Type at Unsignalized T-Intersections Using Accepted Gaps: A Perspective from Emerging Countries. J Transp Eng A Syst. 2022 Jun;148(6).

[19] Goyani J, Asce SM, Aninda;, Paul B, Gore N, Arkatkar S, et al. Investigation of Crossing Conflicts by Vehicle Type at Unsignalized T-Intersections under Varying Roadway and Traffic Conditions in India. J Transp Eng A Syst. 2021 Feb;147(2).

[20] Paul AB, Goyani J, Arkatkar S, Joshi G. Modeling the Effect of Motorized Two-Wheelers and Autorickshaws on Crossing Conflicts at Urban Unsignalized T-Intersections in India using Surrogate Safety Measures. Transportation Research Procedia. 2022;62:774-81.

[21] Pirdavani A, Brijs T, Bellemans T, Pirdavani A, Brijs T, Bellemans T, et al. Evaluation of Traffic Safety at Un-signalized Intersections Using Microsimulation: A Utilization of Proximal Safety Indicators [Internet]. Vol. 22, Advances in Transportation Studies an international Journal Section A. 2010. Available from: https://www.researchgate.net/ publication/228660289

[22] Srinivasula SR, Chepuri A, Arkatkar SS, Joshi G. Developing proximal safety indicators for assessment of un-signalized intersection – a case study in Surat city. Transportation Letters [Internet]. 2020 May 27 [cited 2023 Feb 22];12(5):303-15. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=ytrl20

[23] Orsini F, Gastaldi M, Rossi R. Conflict-Based Real-Time Road Safety Analysis: Sensitivity to Data Collection Duration and its Implications for Model Resilience. Transp Res Rec [Internet]. 2024;2678(1):460-72. Available from: https://doi.org/10.1177/03611981231171151

[24] Arndt O, Troutbeck R. Relationship between unsignalised intersection geometry and accident rates - A literature review. Vol. 10, Road and Transport Research. 2003.

[25] Nambuusi BB, Brijs T, Hermans E. A Review of Accident Prediction Models for Road Intersections. Infrastructuur en ruimte. 2008. 69 p.

[26] Yannis G, Antoniou C, Papadimitriou E. Road casualties and enforcement: Distributional assumptions of serially correlated count data. Traffic Inj Prev [Internet]. 2007 Sep [cited 2023 Feb 22];8(3):300-8. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=gcpi20

[27] Dinarcan GN. Count Data Regression Models 1. Hacettepe Üniversitesi; 2018.

[28] Rahman S. Development of an Accident Prediction Model for Intersections of Dhaka City, Bangladesh. Int J Comput Appl. 2012;47(16):10-6.

[29] Hauer E, Ng JCN, Lovell J. Estimation of safety at signalized intersections. Transp Res Rec. 1988;(1185):48-61.

[30] Mahel M, Summersgill I. A comprehensive methodology for the fitting of predictive accident models. Science (1979). 1996;28(3):281-96.

[31] Salifu M. Accident prediction models for unsignalised urban junctions in Ghana. International Association of Traffic and Safety Sciences [Internet]. 2004;28(1):68-81. Available from: http://dx.doi.org/10.1016/S0386-1112(14)60093-5

[32] Turner S, Nicholson A, Miaou SP, Lord D. Intersection accident estimation: The role of intersection location and non-collision flows. Accid Anal Prev. 2003;30(1840):31-40.

[33] Chin HC, Quddus MA. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid Anal Prev. 2003;35(2):253-9.

[34] Sawalha Z, Sayed T. Transferability of accident prediction models. Saf Sci. 2006;44(3):209-19.

[35] Vogt A. Crash models for rural intersections: Four-lane by two-lane stop-controlled and two-lane by two-lane signalized [Internet]. Vol. Jan, Transportation Research Record: Journal of the Transportation Research Board. 1999 [cited 2021 Jun 1]. 18-29 p. Available from: https://trid.trb.org/view/648036

[36] Hilbe JM. Negative binomial regression [Internet]. Negative Binomial Regression. 2007 [cited 2021 Jun 7]. 1-251 p. Available from: https://books.google.co.uk/books?hl=tr&lr=&id=0Q_ijxOEBjMC&oi=fnd&pg=PR11&dq=Poisson+regression+is+a+special+situation+of+Negative+Binomial+regression.&ots=IK2_V

[37] Poch M, Mannering F. Negative binomial analysis of intersection-accident frequencies. J Transp Eng. 1996;122(2):105-13.

[38] Kumara, S. S. P., and Chin HC. Modeling Accident Occurrence at Signalized Tee-Intersections with Special Emphasis on Excess Zeros. 2003.

[39] Prasetijo J, Zahidah Musa W. Modeling Zero Inflated Regression of Road Accidents at Johor Federal Road F001. 2016.

[40] Wright S. Correlation and Causation. 1921.

[41] Wright S. The method of path coefficients. 1934.

[42] Fornell C, Larcker DF. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Vol. 18, Journal of Marketing Research. JSTOR; 1981. 39 p.

[43] Pickering D, Hall RD, Grimmer M. Accidents at rural T-junctions. Vol. P270, Research Report - Transport and Road Research Laboratory. 1986.

[44] World Population Review. Portsmouth Population (Demographics, Maps, Graphs) [Internet]. 2022 [cited 2022 Nov 4]. Available from: https://worldpopulationreview.com/world-cities/portsmouth-population

[45] Bunce A. Portsmouth's 20mph Scheme [Internet]. 2009 [cited 2022 May 17]. Available from: https://www.roadsafetyknowledgecentre.org.uk/rskc-54/

[46] CIHT. Transport in the urban environment (TUE) [Internet]. 1997 [cited 2022 May 19]. Available from: https://www.thenbs.com/PublicationIndex/documents/details?Pub=IHT&DocID=305156

[47] Ekmekci M. An investigation of piecewise Structural Equation Modelling to understand the effect on collisions of shorter visibility designs at urban three arm junctions. [Portsmouth]: University of Portsmouth; 2023.

[48] Shipley B. Confirmatory path analysis in a generalized multilevel context. Vol. 90, Ecology. 2009.

[49] Stein CM, Morris NJ, Nock NL. Structural Equation Modeling. Methods in Molecular Biology [Internet]. 2012 [cited 2022 Apr 27];850:495-512. Available from: https://link.springer.com/protocol/10.1007/978-1-61779-555-8_27

[50] Lefcheck JS. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. Methods Ecol Evol [Internet]. 2016 May 1 [cited 2022 Aug 9];7(5):573-9. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.12512

[51] Shipley B. A New Inferential Test for Path Models Based on Directed Acyclic Graphs. http://dx.doi.org/101207/S15328007SEM0702_4 [Internet]. 2000 [cited 2022 Apr 26];7(2):206-18. Available from: https://www.tandfonline.com/doi/abs/10.1207/S15328007SEM0702_4

[52] Pearl J, Verma T. The logic of representing dependencies by directed graphs. 1987.

[53] Shipley B. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology [Internet]. 2013 Mar 1 [cited 2022 Apr 27];94(3):560-4. Available from: https://onlinelibrary.wiley.com/doi/full/10.1890/12-0976.1

[54] Fisher RA. Statistical methods for research workers. Statistical methods for research workers. 1954;(10th. ed.).

[55] Ekmekci M, Dadashzadeh N, Woods L. Assessing the impact of low-speed limit zones' policy implications on cyclist safety: Evidence from the UK. Transp Policy (Oxf) [Internet]. 2024 Jun; 152:29-39. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0967070X24001173

[56] Ekmekci M, Woods L, Dadashzadeh N. Effects of road width, radii and speeds on collisions at three-arm priority intersections. Accid Anal Prev. 2024 May 1;199.

[57] Chakrabarti A, Ghosh JK. AIC, BIC and Recent Advances in Model Selection. Philosophy of Statistics. 2011 Jan 1;583-605.