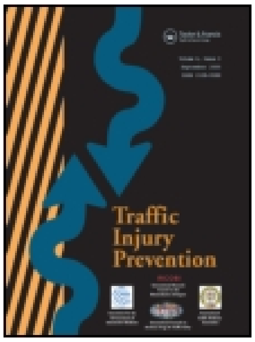

Traffic Injury Prevention

ISSN: 1538-9588 (Print) 1538-957X (Online) Journal homepage: http://www.tandfonline.com/loi/gcpi20

# Modeling Accident Occurrence at Signalized Tee Intersections with Special Emphasis on Excess Zeros

S. S. P. Kumara & H. C. Chin

Traffic Injury Prevention, 4:53–57, 2003

1538-9588/03 $12.00 + .00
DOI: 10.1080/15389580390120916

Taylor & Francis
Taylor & Francis Group

# Modeling Accident Occurrence at Signalized Tee Intersections with Special Emphasis on Excess Zeros

**S. S. P. KUMARA and H. C. CHIN**
Department of Civil Engineering, National University of Singapore, Singapore

*In implementing effective remedial treatments at hazardous intersections, it often is necessary to identify the geometric and traffic factors that lead to accident occurrence. However, one particular problem frequently encountered in accident studies is how to distinguish virtually safe intersections with little likelihood of accident occurrence from those that have happened to have no accident due to the random process. To deal with this problem, the "excess" records of zero accident, the zero-inflated negative binomial was used to assign the probability to the accident outcome. Accident data at 104 signalized tee intersections in Singapore over a period of 9 years were employed for model development. The model indicates that uncontrolled left-turn slip road, permissive right-turn phase, existence of a horizontal curve, short sight distances, large number of signal phases, total approach volume, and left-turn volume may increase accident occurrence. On the other hand, right-turn channelization, acceleration section on the left-turn lane, median railings, and more than 5% approach gradient may reduce accident occurrence. Moreover, there is a trend of reducing accidents over the years.*

**Keywords** Causal Factors; Excess Zeros; Signalized Intersection Approaches; Singapore; Traffic Safety; Zero-Inflated Negative Binomial Model

Singapore is a heavily motorized country with one of the most advanced transportation systems in Asia. Although the level of road safety is good by Asian standards, the accident rates are still unacceptable. Road accident casualties increased by 1,495 from 7,634 in 1998 to 9,129 in 1999, and fatal and injury accidents increased by 16.2% (Traffic Police Department, Singapore, 1999). Based on annual accident statistics, more than 30% of the crashes in Singapore occur at signalized intersections. In planning for suitable countermeasures, it is necessary to first identify the various factors that may influence accident occurrence at these locations. This can be done by establishing a statistical model relating accident frequencies with the traffic characteristics, traffic control measures, and geometric design elements of the intersection. However, there are few studies involving such accident models in Singapore.

Another problem in developing such statistical models is the number of occasions when zero accidents are recorded at a particular intersection. When there is a zero accident record over a period of time, it may indicate either that the site is virtually safe, or that the zero record is a chance occurrence or accidents are not reported. Since the former does not help to identify accident contributory factors, it becomes necessary to model the two states separately. Moreover if the two states are modeled as a single state, the estimated models may be biased as there may be an overrepresentation of zero accidents. Hence the presence of excess zeros in the accident count data may be mistakenly regarded as the presence of overdispersion in the data set, which arises because of an incorrectly specified model. To handle count data with excess zeros, Lambert (1992) and Land et al. (1996) have employed the zero-inflated Poisson (ZIP) regression. This article seeks to identify the factors affecting accident occurrence at signalized tee intersections in Singapore while taking into account the presence of excess zeros.

## *METHODOLOGY*

The basic assumption in the zero-inflated models is that the population consists of two possible states (Greene, 1997). The zero-inflated models separate the true zero state process from the parent count models such as Poisson and negative binomial (NB), and let contributory factors influence the two states. In the case of accident study, this dual state system can be expressed

Received 17 May 2001; accepted 15 October 2002.
Address correspondence to S. S. P. Kumara, Department of Civil Engineering, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260. E-mail: engp0436@nus.edu.sg

as a probability function by assuming $q_i$ is the probability that intersection approach $i$ will exist in the zero accident state and $(1 - q_i)$ is the probability that a zero accident observation actually follows a suitable count distribution such as Poisson or NB. Therefore,

$$P(n_i) = \begin{cases} q_i + (1 - q_i)R_i(0) & n_i = 0 \\ (1 - q_i)R_i(n_i) & n_i > 0 \end{cases} \tag{1}$$

where $n_i$ is the annual number of recorded accidents on intersection approach, $R_i(n_i)$ is the probability of occurrence of $n_i$ accidents. In the NB formulation, this is given by

$$R_i(n_i) = \frac{\Gamma(\theta + n_i)}{[n_i!\Gamma(\theta)]} u_r^{\theta}[1 - u_i]^{n_i} \tag{2}$$

where $\theta = 1/\alpha$ and $u_i = \theta/[\theta + \lambda_i]$, in which $\alpha$ is the overdispersion parameter and $\lambda_i$ is the mean.

Zero-inflated models can be formulated in one of two ways either assuming a single parameter vector for $q_i$ and $\lambda_i$ or separate parameter vectors. For the purpose of this article, we present only the former approach. (See Lambert, 1992, for more extensive literature). The zero-inflated negative binomial ($\tau$) model, ZINB ($\tau$), can be formulated by assuming the covariates that affect the mean $\lambda_i$ of the accident state and probability $q_i$ of the zero accident state may be related to each other by a single real value parameter $\tau$ such that $\log it(q_i) = \tau\beta x_i$, where $\beta$ is a vector representing parameters to be estimated and $x_i$ is a vector of covariates describing the characteristics of the approach. However, there will be a problem of distinguishing the ZINB ($\tau$) model from an underlying NB specification as the source of the overdispersion. In the former overdispersion arises from unobserved heterogeneity, whereas in the latter overdispersion arises from excess zero accident observations. Vuong (1989) has proposed a test statistic for nonnested models, which appears to have some power to distinguish between the ovedispersion of the NB model and the force of the splitting mechanism in the ZINB ($\tau$) part of the model. That is, the Vuong statistic evaluates whether there is a regime splitting mechanism at work or not. The statistic is,

$$V = N^{1/2}\bar{m}/s_m \tag{3}$$

where $\bar{m}, s_m$, and $N$ are mean, standard deviation, and sample size respectively of $m_i$ with $m_i = \log[f_1(n_i)/f_2(n_i)]$, $f_i(.)$ and $f_2(.)$ are density functions of the ZINB ($\tau$) distribution and parent NB distribution respectively. The Vuong statistic is distributed as standard normal distribution; hence, its value may be compared to the critical value for standard normal distribution. Therefore it may be taken that $V > 1.96$ distinctly favors the zero-inflated count models while $V \leq -1.96$ distinctly favors the parent NB model otherwise the test is indecisive. The selection of the relevant model is therefore established by evaluating the Vuong statistic $V$ and the $t$ statistic for the negative binomial overdispersion parameter $\alpha$ as shown in Table I (Shankar et al., 1997). The complementary condition of selecting the ZIP ($\tau$) or ZINB ($\tau$) model is that the shape parameter $\tau$ be statistically significant.

**Table I** Decision rule for model selection using the Vuong statistic and negative binomial overdispersion parameter criteria

| | $t$ Statistic of $\alpha$ | |
|---|---|---|
| Vuong statistic | <2 | >2 |
| $V < -1.96$ | Poisson | NB |
| $V > 1.96$ | ZIP models | ZINB models |

Abbreviations: NB, negative binomial; ZINB, zero-inflated negative binomial; ZIP, zero-inflated Poisson.

## *THE DATA*

Tee Intersections from 1992 to 2000 contributed to 15.7% of total road accidents. A total of 104 three-legged signalized intersections were used in the study and annual accident data from 1992 to 2000 were extracted from National Road Accident Data Base. Table II shows the distribution of number of accidents per approach per year for the intersections considered. Traffic volume data were obtained from loop detectors at the sites maintained by the Land Transport Authority (LTA). Data relating to traffic control measures and geometric design elements at the site were obtained either from LTA records or through site measurements. Table III presents summary statistics of these data for intersection approaches examined. Some of the variables are defined in Table IV.

## *MODEL ESTIMATION*

Equations (1) and (2) can be combined to estimate model parameters $\beta$, $\alpha$, and $\tau$. The probability density function for the observed random variable, $n_i$, is

$$p(n_i) = (1 - q_i)R_i(n_i) + z_i q_i \tag{4}$$

where $z_i = 1$ when $n_i$ is observed to be zero and $z_i = 0$ for all other values of $n_i$. The indicator variable $z_i$ eases the

**Table II** Distribution of accident counts from 1992 to 2000 at the selected signalized tee intersection approaches

| Number of accidents | Frequency |
|---|---|
| 0 | 2,238 |
| 1 | 368 |
| 2 | 107 |
| 3 | 51 |
| 4 | 12 |
| 5 | 2 |
| 6 or greater | 2 |

**Table III** Summary statistics of variables for the signalized tee intersection approaches

| Variables | Min | Max | *M* | *SD* |
|---|---|---|---|---|
| Annual accident frequency | 0 | 6 | 0.2896 | 0.6933 |
| Traffic data | | | | |
| Total approach volume in AADT | 1,093 | 66,378 | 12,922.3 | 9,497.2 |
| Total left-turn volume in AADT | 0 | 21,312 | 2,742.0 | 3,713.2 |
| Right-turn volume in AADT | 0 | 26,514 | 2,068.36 | 3,222.07 |
| Geometric data | | | | |
| Sight distance (m) | 49.15 | 515.63 | 244.117 | 154.260 |
| Existence of 5% gradient[a] | 0 | 1 | 0.2051 | 0.4039 |
| Number of approach lanes | 1 | 6 | 2.8109 | 0.9506 |
| Median width (m) | 0 | 29 | 1.9540 | 2.7813 |
| Horizontal curve[a] | 0 | 1 | 0.3590 | 0.4798 |
| Right-turn channelization[a] | 0 | 1 | 0.3237 | 0.4690 |
| Left-turn channelization[a] | 0 | 1 | 0.2058 | 0.4043 |
| Left-turn slip road[a] | 0 | 1 | 0.2981 | 0.4575 |
| Length of left-turn slip road (m) | 0 | 103.46 | 8.6615 | 15.7726 |
| Acceleration section on left-turn lane[a] | 0 | 1 | 0.0128 | 0.1125 |
| Traffic devices | | | | |
| Surveillance camera[a] | 0 | 1 | 0.1154 | 0.3195 |
| Median rails[a] | 0 | 1 | 0.2788 | 0.4485 |
| Number of signal phases | 2 | 4 | 2.9423 | 0.3348 |
| Permissive right turn[a] | 0 | 1 | 0.7356 | 0.4411 |

AADT—Average Annual Daily Traffic.
[a]Indicator variables.

maximization of the log likelihood function given by

$$L(\beta, \alpha, \tau) = \Sigma_{i=1}^{N} \log p(n_i). \qquad (5)$$

It should be noted that the derivation of zero-inflated models assumes that annual approach accident frequencies are independent. As the data in our model are for consecutive years, there may be some correlation in time and this may affect the efficiency of coefficient estimates. To investigate the potential serial correlation, the data set was divided into several sets and a comparison of coefficient estimates using a likelihood ratio test was conducted, as suggested by Shankar et al. (1997).

**Table IV** Definitions of some of the unfamiliar variables for signalized tee intersection approaches

| Variable | Definition |
|---|---|
| Right-turn channelization/ left-turn channelization | This is the exclusive lane provided to complete the maneuver. Hence, the right-turning or left-turning vehicles need not slow down in a lane shared with speed-maintaining vehicles. |
| Left-turn slip road | This is the uncontrolled lane which allows left-turning vehicles to merge into the cross-traffic stream. This is normally not under signal control in Singapore. |
| Acceleration section on left-turn slip road | This is the exclusive road section that facilitates the left-turning vehicle from the slip road to proceed without stopping or slowing down to give right of way for the oncoming vehicles from the other approach. |
| Permissive right-turn phase | This is the unprotected right-turning phase which makes right-turning maneuver to proceed with through vehicles when there is a gap in the cross-traffic. |

## *RESULTS AND DISCUSSION*

Based on the NB specification, the estimated parameters of the model are computed and shown in Table V. The overdispersion parameter $\alpha$ was found to be statistically significant ($t = 7.755$, $p < .0001$), which seems to imply at first glance that the NB distribution is appropriate. However, when tested against the alternative ZINB ($\tau$), the Vuong statistic ($V = 4.98 > 1.96$)

**Table V** Negative binomial accident frequency estimation for signalized tee intersection approaches

| Explanatory variable | Estimated coefficients | *p*-value |
|---|---|---|
| Constant | −9.1938 | .0000 |
| Total approach volume in thousand AADT | 0.6260 | .0000 |
| Left-turn volume in thousand AADT | 0.1786 | .0242 |
| Sight distance (1 if >100 m, 0 otherwise) | −0.2347 | .1435 |
| Time trend (1 = 1992, 2 = 1993, and so on) | −0.0249 | .1609 |
| Right-turn channelization (1 if yes, 0 otherwise) | −0.4771 | .0000 |
| Uncontrolled left-turn slip road (1 if yes, 0 otherwise) | 0.2310 | .0603 |
| Acceleration section on left-turning lane (1 if yes, 0 otherwise) | −0.4111 | .0053 |
| Signal phases per cycle | 0.2705 | .1529 |
| Median railing (1 if yes, 0 otherwise) | −0.2059 | .0742 |
| Horizontal curve (1 if yes, 0 otherwise) | 0.2114 | .0682 |
| Permissive right-turn phase (1 if yes, 0 otherwise) | 0.4774 | .0000 |
| Existence of 5% gradient (1 if yes, 0 otherwise) | −0.4509 | .0006 |
| $\alpha$ ($t = 7.755$) | 1.6968 | .0000 |
| Log likelihood at convergence | −1808.40 | |

AADT—Average Annual Daily Traffic.

**Table VI** Zero-inflated negative binomial ($\tau$) accident frequency estimation for signalized tee intersection approaches

| Explanatory variable | Estimated coefficients | $p$-value |
|---|---|---|
| Constant | −5.6637 | .0000 |
| Total approach volume in thousand AADT | 0.4133 | .0000 |
| Left-turn volume in thousand AADT | 0.1223 | .0348 |
| Sight distance (1 if $\geqslant$ 100 m, 0 otherwise) | −0.1433 | .2123 |
| Time trend (1 = 1992, 2 = 1993, and so on) | −0.0164 | .1819 |
| Right-turn channelization (1 if yes, 0 otherwise) | −0.3192 | .0003 |
| Uncontrolled left-turn slip road (1 if yes, 0 otherwise) | 0.1622 | .0592 |
| Acceleration section on left-turning lane (1 if yes, 0 otherwise) | −0.2855 | .0087 |
| Signal phases per cycle | 0.1752 | .1767 |
| Median railing (1 if yes, 0 otherwise) | −0.1287 | .1165 |
| Horizontal curve (1 if yes, 0 otherwise) | 0.1382 | .0896 |
| Permissive right-turn phase (1 if yes, 0 otherwise) | 0.3194 | .0002 |
| Existence of 5% gradient (1 if yes, 0 otherwise) | −0.3028 | .0032 |
| $\alpha$ | 0.1496 | .0751 |
| $\tau$ | −0.8124 | .0419 |
| Voung statistic | 4.98 | |
| Log likelihood at convergence | −1807.02 | |

indicated that the latter was a better representative model than its parent NB model in detecting excess zeros after controlling for overdispersion. The parameter $\alpha$ indicates the residual overdispersion in the nonzero accident state. This may be due to the existence of a reasonable proportion of high frequency counts. The ZINB ($\tau$) shape parameter $\tau$ was also significant ($p = .0151$) indicating that ZINB ($\tau$) with constrained coefficients were appropriate. This suggests that the covariates of the zero accident state and the covariates that affect the NB mean $\lambda_i$ have an implied relationship to each other.

It should be noted that in Singapore vehicles are driven on the left side of the road when referring to the following explanations. The results in Table VI show that accident occurrence is significantly affected by the total approach volume ($p < .0001$) and left-turn volume ($p = .0348$). Increased volume implies greater interaction between vehicles and possibly more conflicts. Accidents will therefore increase because of greater exposure. Furthermore, as volume increases, there may also be fewer gaps in the traffic for the right-turning opposing maneuver as well as left-turning merging maneuver.

The model indicates that the provision of right-turn slip road may reduce accident occurrence and increase the zero accident state ($p = .0003$). This exclusive lane lowers the differential in speed among approaching vehicles; hence, a possible reduction in conflicts. In their study of intersection accident frequencies Poch and Mannering (1996) concluded that, without an exclusive lane, right-turning vehicles that are required to slow in a lane shared with speed-maintaining vehicles may cause a speed differential that tends to cause rear-end accidents.

According to the model, the uncontrolled left-turn slip road increases the accident occurrence ($p = .0603$). The uncontrolled lane, which allows left-turning vehicles to merge into the cross-traffic stream, may also increase the likelihood of sideswipe and head-to-side accidents where drivers fail to yield to oncoming traffic. However, if an acceleration section is provided in the left-turn lane, drivers may be able to merge more easily. This explains the reduction in accidents ($\beta = -0.4111$, $p = .0053$).

One of the most important traffic control devices is the median railing. Median railings are installed to prevent pedestrians from crossing the road where there are no designated pedestrian crossings. There is some evidence that the presence of a median railing has reduced accident occurrence ($p = .0742$).

Accident occurrence seems to increase if the intersection approach is on a horizontal curve ($p = .0682$). Poch and Mannering (1996) have also indicated that the horizontal curve on an approach would increase all types of accidents. In another study, Kulmala (1995) pointed out that drivers cannot make premature observations of the other approaching traffic and decisions concerning the negotiation of the junctions before dealing with the curve.

The model indicates that the unprotected right-turn phase may increase accident occurrence ($p < .0001$). Unprotected right-turn phasing makes the right-turning maneuver proceed with through vehicles when there is a gap. This may increase the likelihood of collisions resulting from the right-turning vehicles failing to yield with oncoming vehicles.

Surprisingly, the model indicates that approaches on gradients of 5% or more may reduce accident occurrence ($p = .0006$). Almost all of the approach gradients at the intersections considered were less than 8%. Although grades with 8% and above are known to be more hazardous (Thagesen, 1996), grades from 5% to 8% may be safer because of reduced speed. In addition, there likely are mitigation factors such as signing and markings at these gradients.

Following explanations should be treated as general trends, as the significance level of following variables is lower as compared to other variables. Accident frequency seems to increase when the sight distance is smaller than 100 m. Short sight distances may put drivers in critical situations and reduce the drivers' abilities to judge the traffic conditions at the intersection. In his study about safety at rural three- and four-arm junctions, Kulmala (1995) also concluded that the short sight distances might increase accident occurrence.

Over the years many engineering advances in vehicles, roads, and lighting systems have been made, and more advanced driver training and vehicle checking systems have been introduced in Singapore. These may reduce the possibility of accident occurrence at intersection approaches as indicated by the negative model parameter of the time trend.

The model indicates that accident occurrence increases with more phases per cycle. Normally the number of phases is higher for busy intersections with more conflicting demands on the intersection (Poch & Mannering, 1996). The result is not surprising since most accidents occur during the phase-change period.

## CONCLUSION

The purpose of this study was to identify the factors affecting road accident occurrence at signalized tee intersection approaches in Singapore, taking into account cases of records of no accidents. The results of the fitted ZINB ($\tau$) model indicate the significance of several highway geometric characteristics, traffic control measures, and traffic characteristics and also the effects of excess zeros on accident modeling. We found that the presence of right-turn channelization, an acceleration section on the left-turning lane, median railing, and existence of a more than 5% gradient may reduce accident occurrence while total and left-turn volumes, an uncontrolled left-turn slip road, signal phases per cycle, existence of a horizontal curve, and a permissive right-turn phase would increase accident occurrence. Further, there is a trend of reducing accidents over the years.

## NOMENCLATURE

| | |
|---|---|
| $\alpha$ | Overdispersion parameter of negative binomial distribution |
| $\beta$ | A vector of parameters to be estimated representing the effects of covariates |
| $\lambda_i$ | Expected number of accidents for the intersection approach $i$ |
| $\tau$ | The real value parameter of zero-inflated Poisson, ZIP ($\tau$) or zero-inflated negative binomial, ZINB ($\tau$) models |
| $f_1(.)$ | Density functions of the zero-inflated distributions |
| $f_2(.)$ | Density functions of the parent Poisson or negative binomial (NB) distribution |
| $L(.)$ | Likelihood value of a given function |
| $m_i$ | Logarithmic value of the division between density functions of the zero-inflated and parent NB distribution for approach $i$ |
| $\overline{m}$ | Mean of $m_i$ |
| $n_i$ | A random variable representing the annual number of accidents for approach $i$ |
| $N$ | Sample size of $m_i$ |
| $p(n_i)$ | Probability of $n$ accidents occurring on approach $i$ |
| $q_i$ | Ancillary state probability in the zero-inflated models |
| $R(n_i)$ | Poisson or negative binomial probability with $n_i$ |
| $s_m$ | Standard deviation of $m_i$ |
| $V$ | Vuong statistic |
| $x_i$ | A vector of covariates representing geometric, traffic, and traffic control measures |
| $z_i$ | A binary variable that determines the probability states of $n_i$ |
| $\Gamma(.)$ | A gamma function |

## REFERENCES

Greene, W. H. (1997) *Econometric Analysis*, 3rd ed., Prentice-Hall, Upper Saddle River, NJ.

Kulmala, R. (1995) *Safety at Rural Three- and Four-Arm Junctions: Development and Application of Accident Prediction Models*, Technical Research Centre of Finland, Vuorimiehentie, Finland.

Lambert, D. (1992) Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing, *Technometrics*, Vol. 34, pp. 1–14.

Land, K. C., McCall, P. L., Nagin, D. S. (1996) A Comparison of Poisson, Negative Binomial and Semi-parametric Mixed Poisson Regressive Models with Empirical Applications to Criminal Careers Data, *Sociol. Methods Res*., Vol. 24, pp. 387–442.

Poch, M., Mannering, F. (1996) Negative Binomial Analysis of Intersection Accident Frequencies, *J. Transportation Eng.*, Vol. 122, pp. 105–113.

Shankar, V., Milton, J., Mannering, F. (1997) Modeling Accident Frequencies as Zero-Altered Probability Process: An Empirical Inquiry, *Accid. Anal. Prev.*, Vol. 29, pp. 829–837.

Thagesen, B. (1996) *Highway and Traffic Engineering in Developing Countries*, 1st ed., E & FN Spon, London.

Traffic Police Department. (1999) Available at http://www.spinet.gov.sg/tp_report/stat99.html.

Vuong, Q. (1989) Likelihood Ratio Tests for Model Selection and Non-nested Hypothesis, *Econometrica*, Vol. 57, pp. 307–334.