Accident Analysis and Prevention 48 (2012) 254–263

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Estimation of rear-end vehicle crash frequencies in urban road tunnels

Qiang Meng[a,*], Xiaobo Qu[a,b]

[a] Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576
[b] Griffith School of Engineering, Griffith University, Gold Coast, QLD 4222, Australia

ARTICLE INFO

Article history:
Received 17 July 2011
Received in revised form 1 January 2012
Accepted 19 January 2012

Keywords:
Rear-end crash frequency
Time to collision
Inverse Gaussian regression model
Road tunnels

ABSTRACT

According to *The Handbook of Tunnel Fire Safety*, over 90% (55 out of 61 cases) of fires in road tunnels are caused by vehicle crashes (especially rear-end crashes). It is thus important to develop a proper methodology that is able to estimate the rear-end vehicle crash frequency in road tunnels. In this paper, we first analyze the time to collision (TTC) data collected from two road tunnels of Singapore and conclude that Inverse Gaussian distribution is the best-fitted distribution to the TTC data. An Inverse Gaussian regression model is hence used to establish the relationship between the TTC and its contributing factors. We then proceed to introduce a new concept of *exposure to traffic conflicts* as the mean sojourn time in a given time period that vehicles are exposed to dangerous scenarios, namely, the TTC is lower than a predetermined threshold value. We further establish the relationship between the proposed exposure to traffic conflicts and crash count by using negative binomial regression models. Based on the limited data samples used in this study, the negative binomial regression models perform well although a further study using more data is needed.

1. Introduction

Road tunnels are increasingly cost-effective infrastructures which provide underground vehicular passageways for motorists and commuters, especially in densely populated cities like Singapore. With the increasing traffic volume and urban development as well as growing needs for land use in urban areas, constructing road tunnels is becoming one alternative to enhance the capacity and accessibility for road transport systems. However, fire disasters occurred in a road tunnel would result in catastrophic consequences due to the enclosed structure nature of tunnel systems. For example, 39 people lost their lives in a fire disaster that happened in the Mont Blanc Tunnel from France to Italy in 1999; and another disaster in Tauern Tunnel of Austria resulted in 12 fatalities (PIARC, 2008). These accidents have raised the awareness among the public as well as the government on both the safety aspect of the tunnels and that of the road tunnel users. Thus, quantitative risk assessment (QRA) has been one of the requirements under the European Union (EU) Directive (2004/54/EC). In Singapore, QRA for all major urban road tunnels longer than 240 m is compulsory in accordance with the Project Safety Review (PSR) procedure manual for roads in the country (LTA, 2005).

Several QRA models for road tunnels have been developed, including TuRisMo model of Austria, TUNPRIM model of the Netherlands, Italian risk analysis model, OECD/PIARC model (PIARC, 2008), and QRAFT model of Singapore (Meng et al., 2011a,b; Qu et al., 2011). All these models acknowledge that the frequency of fire occurred in road tunnels is the most important contributing factor for the risk assessment of road tunnels. *The Handbook of Tunnel Fire Safety* (Beard and Carvel, 2005) points out that over 90% (55 out of 61 fire cases) of tunnel fires are caused by vehicle crashes (especially the rear-end crashes). In addition, according to crash statistics in Singapore's road tunnels, over 2/3 of crashes are categorized as rear-end crashes.[1] Accordingly, on one hand, rear-end crashes are the major cause for fire in road tunnels; on the other hand, rear-end crashes constitute around 70% out of all the crashes. It is, therefore, of great importance to develop a methodology that can estimate the rear-end vehicle crash frequency in road tunnels (the "rear-end crash" henceforth are referred to as "R-E crash" for short).

A number of studies have been conducted to predict/estimate frequency of various types of crashes in highways using crash-frequency data. However, identification of the cause and effect relationship is typical unavailable due to lack of microscopic traffic information (or the detailed driving data). Consequently, as pointed out by Lord and Mannering (2010), researchers have framed their analytic approaches to study the factors that affect the number of crashes occurring in some geographical spaces over some specified time periods by using various types of count-data regression models in accordance to some assumptions. These models include

* Corresponding author. Tel.: +65 6516 5494; fax: +65 6779 1635.
E-mail addresses: ceemq@nus.edu.sg (Q. Meng), x.qu@griffith.edu.au (X. Qu).

[1] 746 out of 1106 crashes (70%) are categorized as rear-end crashes in the CTE road tunnel from 2006 to 2008.

0001-4575/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2012.01.025

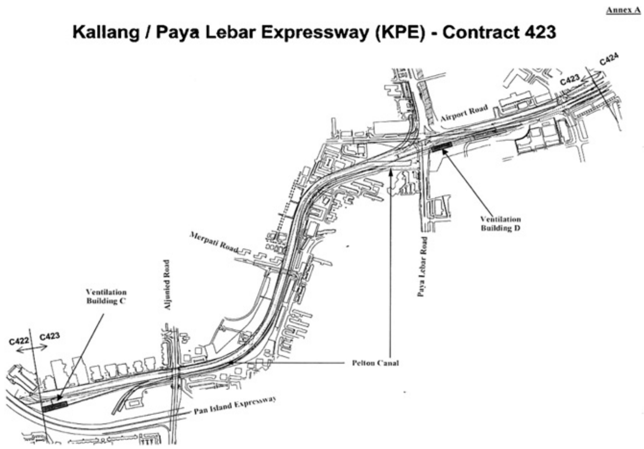

Fig. 1. General arrangement of KPE road tunnel.

Fig. 2. Traffic videos recorded from CTE road tunnel.

Poisson regression model (e.g. Miaou and Lum, 1993; Miaou, 1994; Hauer, 2001), Negative binomial/Poisson-Gamma model (e.g. Maycock and Hall, 1984; Malyshkina and Mannering, 2010a; Daniels et al., 2010), Zero-inflated Poisson and negative binomial models (e.g. Miaou, 1994; Shankar et al., 1997; Malyshkina and Mannering, 2010b; Lord et al., 2007), Conway–Maxwell–Poisson model (e.g. Lord et al., 2008, 2010), and others (e.g. Zhang and Xie, 2007; Guo et al., 2010; Haque et al., 2010). The lack of the detailed driving data on highways may make these statistical analysis models biased to reflect the fundamental cause and effect relationship. Lord and Mannering (2010) thus highlighted that the entirely new direction of research could potentially open up if the anticipated availability of the detailed driving data and crash data are available.

More detailed traffic data are obtainable in road tunnels compared to highways because most of road tunnels are equipped with the closed circuit television (CCTV) cameras and/or an operation control center (OCC). For example, each of Singapore's road tunnels has been installed 2–4 CCTV cameras every 200 m and monitored by a 24-h manned operation control center (OCC). These CCTV cameras record real time and detailed traffic information. In addition to hourly traffic volume and density, we can precisely measure/estimate the time to collision (TTC) for two consecutive vehicles moving in the same lane of a road tunnel using traffic videos. The TTC is defined as *the time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained* (Hayward, 1972). The TTC has been one of the well-recognized safety indicators for traffic conflicts on highways (Farah et al., 2009; Svensson, 1998; Vogel, 2003). Minderhoud and Bovy (2001) further pointed out that the TTC is inversely related to vehicle crash frequencies in road sections. It is widely accepted as a safety indicator in highways. A TTC threshold value is usually chosen to distinguish relatively safe situation and dangerous scenarios exposed to traffic conflicts (or critical encounters). It is acknowledged that the TTC threshold should be 2–4 s (Minderhoud and Bovy, 2001; Vogel, 2003).

The objective of this study is to develop a novel R-E crash frequency estimation method on the basis of TTC distributions. The TTC sample data are collected from the traffic videos of Singapore's road tunnels. Based on the statistical analysis, we find that the Inverse Gaussian distribution is the best-fitted distribution model for the collected TTC data. The Inverse Gaussian regression model is thus employed to establish the relationship between TTC distributions and the corresponding traffic volume. Having had the TTC distributions, a R-E crash frequency estimation method is put up to reflect the relationship between the TTC distributions and the R-E crash frequencies.

The remainder of this paper is organized as follows. In Section 2, the TTC is defined and data collected from Singapore's road tunnels are presented. In Section 3, the Inverse Gaussian regression model is built to establish the relationship between TTC distributions and corresponding traffic volumes. In Section 4, two R-E crash frequency estimation models are developed on the basis of the derived TTC distributions. Several conjectures and recommendations for further studies are put forward in Section 5. Section 6 concludes this study.

2. TTC data collection

Assume that there are two consecutive vehicles moving in the same direction on the same lane of a road tunnel. Let L_{leader} and L_{follower} be the locations of the leading and following vehicles at a particular time, respectively. Correspondingly, let $\dot{L}_{\text{leader}}$ and $\dot{L}_{\text{follower}}$ denote the speeds of the leading and following vehicles at the particular time. According to the TTC definition, namely, the *time that remains until a collision between two vehicles would have occurred if the collision course and speed difference are maintained*, the TTC can be mathematically expressed by

$$\text{TTC} = \begin{cases} \dfrac{L_{\text{leader}} - L_{\text{follower}} - l_{\text{leader}}}{\dot{L}_{\text{follower}} - \dot{L}_{\text{leader}}} & \text{,if } \dot{L}_{\text{follower}} > \dot{L}_{\text{leader}} \\ \infty & \text{,otherwise} \end{cases} \quad (1)$$

where l_{leader} is the length of the leading vehicle. Eq. (1) implies that the TTC is measurable if we have real time traffic information.

To collect the TTC data in a road tunnel, the Kallang/Paya Lebar Expressway (KPE) and the Central Expressway (CTE) in Singapore shown in Figs. 1 and 2 are selected. KPE and CTE are two vital infrastructures in Singapore's road system. The first one has a total length of 12 km and 9 km of the expressway (Fig. 1) is built underground as a road tunnel, serving the growing traffic demand of the north-eastern sector of Singapore. The second one, a 17-km expressway, links the north and south of Singapore through the Central Business District (CBD). 2.4 km of the expressway (Fig. 2) are laid underground and these portions of the CTE form the first road tunnel in Singapore. Both road tunnels are equipped with the 24-h OOC systems.

We have requested 42-h tunnel traffic videos recorded by CCTV of these two tunnels from Land Transport Authority of Singapore, including 14 locations for 3 typical time periods – morning peak hour: 8:00 am to 9:00 am, off-peak hour: 14:00 pm to 15:00 pm, evening peak hour: 19:00 pm to 20:00 pm – in March 2011. The TTC data are generated from these traffic videos in different time periods with different traffic conditions. The procedure of measuring a TTC with respect to a particular car-following scenario

Table 1
TTC data.

	Traffic volume (vehs/(h lane))	Number of data
Location 1	894	104
Location 2	963	65
Location 3	1127	80
Location 4	1374	79
Location 5	1672	93

is summarized as follows. We first measure length of the leading vehicle (l_{leader}) in a car-following scenario. After that, the spot speeds of the vehicles ($\dot{L}_{\text{follower}}$ and $\dot{L}_{\text{leader}}$) can be estimated by measuring the time taken by the vehicle to cover two lane-markers' distance in the video. Then, the time headway (h) between the leading and following vehicles is recorded. According to Vogel (2003), the gap size ($L_{\text{leader}} - L_{\text{follower}} - l_{\text{leader}}$) can be estimated by ($\dot{L}_{\text{follower}} \times h - l_{\text{leader}}$). Finally, the TTC for the car-following scenario could be calculated according to Eq. (1).

In the measurement, we display 30 frames per second and limit the error of 0.03 s, which would yield data of comparable quality to the radar-speed measurement method. 867 car following scenarios occurred at various locations are examined, and 421 TTC data (TTC with a finite value) with respect to different traffic volumes are obtained. Statistically, the number of TTC data with a finite value should be equal to that of samples with infinite values. An infinite TTC value indicates that the following vehicle will not be possible catch up with the leading one, which is an absolutely safe situation. We would focus on the probability distributions of TTC samples with finite values accordingly.

3. Inverse Gaussian distribution for TTC

3.1. Statistical analysis for the TTC samples

A data analysis procedure is proposed in order to obtain the best-fitted TTC distributions. Five commonly used distributions are examined in this study: Inverse Gaussian, Exponential, Normal, Triangular, and Lognormal. The maximum likelihood estimation (MLE) technique is employed to estimate the parameters involved in a distribution. After obtaining the parameters for the five types of distributions, the goodness-of-fit test is conducted to select the best-fitted distribution among the given candidate distributions. Kolmogorov–Smirnov (K–S) test, a nonparametric test method, has been widely applied to compare a sample with a reference probability distribution in transportation studies (e.g. Ibeas et al., 2011; Páez et al., 2011). In this study, the K–S test is also adopted to perform the goodness-of-fit test. The K–S statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution. In this study, a distribution with the lowest K–S test statistic is regarded as the best-fitted distribution.

Following the above-mentioned data analysis procedure, we analyze five sets of the TTC data collected at different locations with respect to different traffic volumes, as shown in Table 1. Table 2 gives results of the best-fit analysis.

According to Tables 1 and 2, we can find that

(1) The Inverse Gaussian distribution is the best-fitted distribution for all the five locations.[2] Fig. 3(a) and (b) depicts the histograms and empirical cumulative distribution function (CDF)

Table 2
Statistical analysis for the TTC samples.

	Inverse Gaussian		Lognormal		Triangular		Exponential		Uniform	
	Distributions	K–S	Distributions	K–S	Distributions	K–S	Distributions	K–S	Distributions	K–S
Location 1	IG (9.26, 12.21)	0.0968[a]	Lognorm (9.27, 8.28)	0.1198	Triang (0, 2.30, 31.50)	0.2385	Expon (9.26)	0.1814	Uniform (0, 29.82)	0.3600
Location 2	IG (9.69, 12.88)	0.1003[a]	Lognorm (9.71, 8.62)	0.1138	Triang (0, 2.41, 32.80)	0.2471	Expon (9.69)	0.1764	Uniform (0, 31.39)	0.3746
Location 3	IG (11.20, 14.06)	0.1017[a]	Lognorm (11.53, 9.56)	0.1024	Triang (0, 2.10, 37.40)	0.1756	Expon (11.20)	0.2091	Uniform (0, 36.84)	0.3807
Location 4	IG (12.30, 11.01)	0.0813[a]	Lognorm (12.96, 13.86)	0.1097	Triang (0, 1.41, 40.60)	0.1768	Expon (12.30)	0.1408	Uniform (0, 39.55)	0.3449
Location 5	IG (7.26, 9.24)	0.0651[a]	Lognorm (7.24, 6.45)	0.0781	Triang (0, 1.65, 29.88)	0.3199	Expon (7.26)	0.1934	Uniform (0, 29.57)	0.4948

[a] The K–S statistics of the best-fitted distributions.

[2] Inverse Gaussian distribution is a two-parameter family of continuous probability distributions with support on $(0, \infty)$. Its probability density function is given

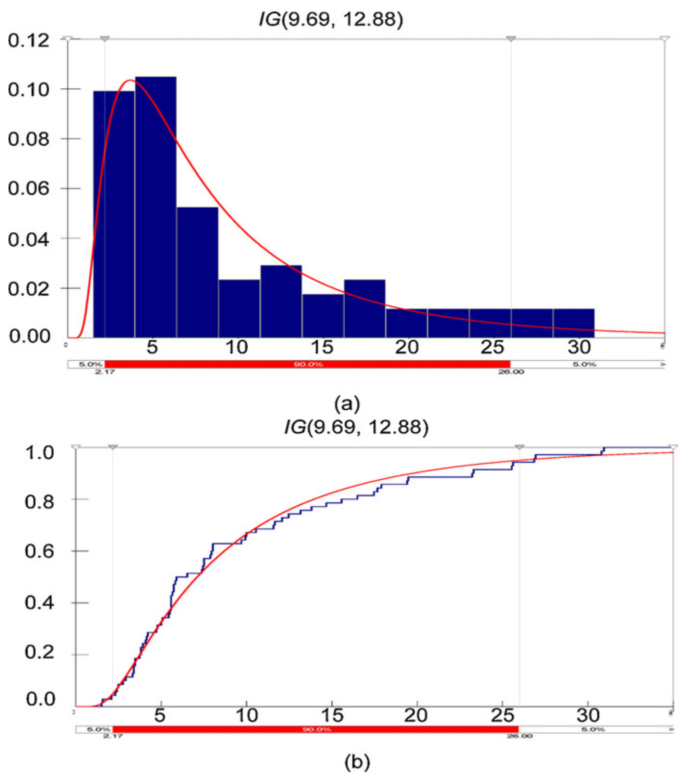

Fig. 3. The histograms and empirical CDF (traffic volume = 963 vehs/(h lane)).

for data samples with the best-fitted distributions (traffic volume = 963 vehs/(h lane)).

(2) Lognormal distribution also performs very well for the five locations (relatively small K–S values). In reality, other samples may suggest that Lognormal distributions are better. The two distributions have similar patterns. Let us take the Location 1 as an example: $P(\mathrm{IG}(9.26, 12.21) \le 3) = 0.138$ and $P(\mathrm{Lognorm}(9.27, 8.28) \le 3) = 0.139$.[3] Indeed, the differences between the two distributions are very marginal. In the following analysis, without loss of generality, we will assume the samples follow Inverse Gaussian distributions as suggested in Table 3.

(3) The TTC data collected at different locations with respect to similar traffic volume generally follows the same Inverse Gaussian distribution (e.g. Location 1 and Location 2). In other words, the traffic volume could be considered as the contributing factor for TTC distributions.

(4) The TTC sample mean and its inverse both have a parabola relationship with the traffic volume, as shown in Figs. 4 and 5. This is because the two contributing factors to TTC, distance headway and speed dispersion, are both dependent of the traffic volume. When traffic volume is low (<1000 vehs/(h lane)), the great speed dispersion could result in low TTC values. However, when traffic volume is high (>1600 vehs/(h lane)), the small distance headway would lead to low TTC values.

(5) The shape parameters (λ) of the best-fitted Inverse Gaussian distributions with respect to different traffic volumes are within a relatively small range from 9.24 to 14.06.

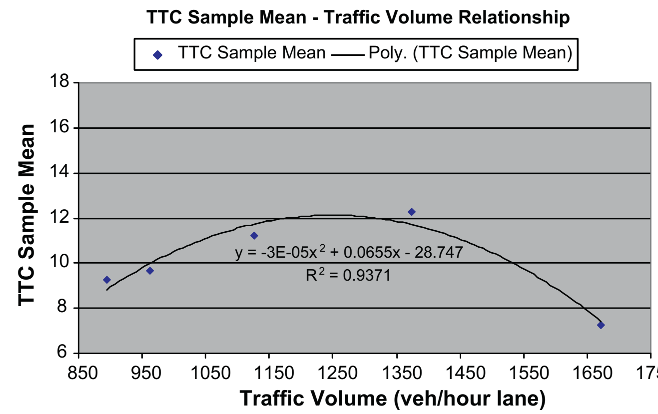

Fig. 4. TTC sample mean–traffic volume relationship.

3.2. Estimation of the parameters defining Inverse Gaussian distribution

As discussed in Section 3.1, both Inverse Gaussian distribution and Lognormal distribution fit the data very well. In the following analysis, without loss of generality, we use an Inverse Gaussian regression model to establish the relationship between TTC and traffic volume. To formulate the inverse Gaussian regression model, let y_i, $i = 1, \ldots, n$, be n independent observations (TTC samples) distributed as $\mathrm{IG}(\mu_i, \lambda)$, in which the inverse of sample mean has a parabola relationship with traffic volume, namely:

$$\frac{1}{\mu_i} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 > 0 \tag{2}$$

where x_i is traffic volume of TTC sample i. Whitmore (1983) derived the pseudo maximum likelihood estimations of β and λ as

$$\hat{\beta} = (X'YX)^{-1}X'\mathbf{1} \tag{3}$$

$$\hat{\lambda} = \frac{n}{1^T Y^{-1}\mathbf{1} - \mathbf{1}'X\hat{\beta}} \tag{4}$$

where Y is the diagonal matrix with ith diagonal elements being y_i, $\mathbf{1}$ is the n-vector of all ones and $X = (1, x_i, x_i^2)^T$. They are called the pseudo maximum likelihood estimations because the condition

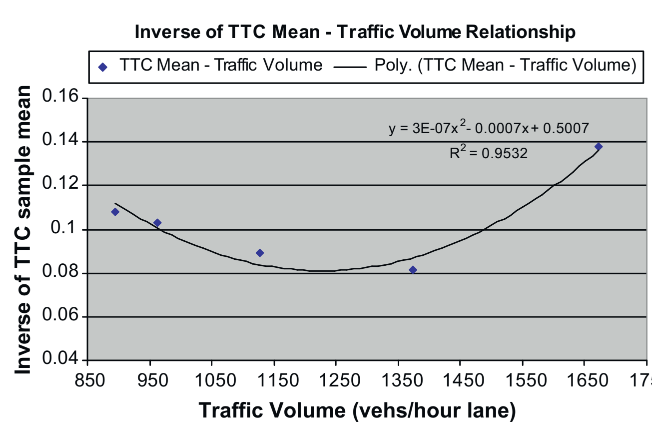

Fig. 5. Inverse of TTC mean–traffic volume relationship.

by

$$f(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{1/2} \exp\left(\frac{-\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad 0 < x < \infty,$$

where $\mu > 0$ is the mean and $\lambda > 0$ is the shape parameter. The distribution can be viewed as the distribution of first passage time of a Wiener process with an absorbing barrier, i.e., while the Gaussian describes a Brownian Motion's level at a fixed time (Wiener process), the inverse Gaussian describes the distribution of the time the Brownian Motion takes to reach a fixed positive level.

[3] In this example, the TTC threshold is assumed to be as 3 s. Similar results are obtainable if we assume the threshold is 2 s or 4 s.

Table 3
K–S tests.

Traffic volume (vehs/(h lane))	Number of samples (n)	Distributions	K–S values (D_n)	Critical value ($K_{0.05}$)	Test results
894	104	IG (9.02, 12.17)	0.0977	1.36	1.00 < 1.36
963	65	IG (10.26, 12.17)	0.1086	1.36	0.88 < 1.36
1127	80	IG (12.33, 12.17)	0.1496	1.36	1.34 < 1.36
1374	79	IG (12.83, 12.17)	0.0821	1.36	0.73 < 1.36
1672	93	IG (7.30, 12.17)	0.1026	1.36	0.99 < 1.36

$\hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 > 0$ for all i is not guaranteed to be satisfied.[4] According to the collected 421 TTC data with different traffic volumes, the estimated coefficients are

$$\hat{\beta}_0 = 5.606 \times 10^{-1} \quad (5)$$

$$\hat{\beta}_1 = -7.900 \times 10^{-4} \quad (6)$$

$$\hat{\beta}_2 = 3.21 \times 10^{-7} \quad (7)$$

$$\hat{\lambda} = 12.17 \quad (8)$$

After obtaining the estimated coefficients, the TTC distributions could be determined for different traffic conditions reflected by their traffic volumes. In order to evaluate how well the Inverse Gaussian regression model estimates the TTC distributions, we compare the derived TTC distributions with the TTC samples at different traffic volumes – 894 vehs/(h lane), 963 vehs/(h lane), 1127 vehs/(h lane), 1374 vehs/(h lane), and 1672 vehs/(h lane) – by using the hypothesis test. The K–S test is applied to conduct the hypothesis test. The null hypothesis is rejected at level α if

$$\sqrt{n} D_n > K_\alpha \quad (9)$$

where n is the number of samples, D_n is the K–S statistic, and K_α is the critical value ($\alpha = 0.05$ in this study). The results of K–S tests are reported in Table 3.

Table 3 shows that the regression model performs well. Fig. 6(a) and (b) depicts the CDF of the best-fitted Inverse Gaussian distribution and the CDF generated by Inverse Gaussian regression for a TTC sample (traffic volume = 1672 vehs/(h lane)), respectively. As can be seen in the figures, the samples could be well represented by either of the two Inverse Gaussian distributions with different parameters.

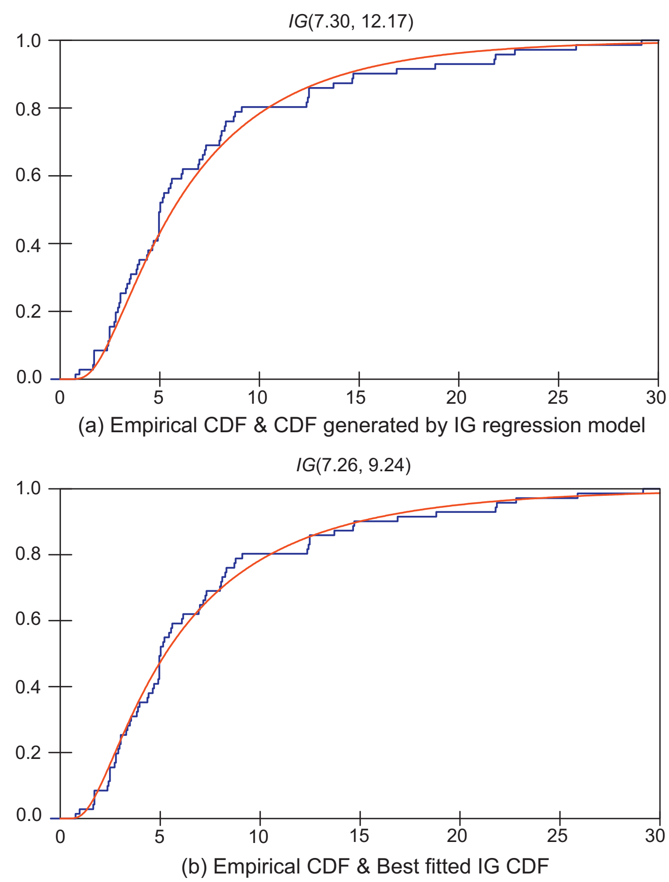

Fig. 6. Empirical CDF with Inverse Gaussian distributions (traffic volume = 1672 vehs/(h lane)).

4. Accident frequency estimation

4.1. TTC threshold value and exposure to traffic conflicts

As mentioned in Section 1, a TTC threshold value is usually chosen to distinguish relatively safe situation and dangerous scenarios exposed to traffic conflicts (or critical encounters). Various opinions can be found from the literature as to which value should be used as the threshold value. Hirst and Graham (1997) reported that a time-to-collision measure of 4 s could be used to discriminate between cases where drivers unintentionally find themselves in a dangerous situation from cases where drivers remain in control. Hogema and Janssen (1996) presented a minimum TTC value of 3.5 s for non-supported drivers and 2.6 s for supported drivers. It is widely acknowledged that the TTC threshold should be 2 s to 4 s (Minderhoud and Bovy, 2001; Vogel, 2003).

We define the *exposure to traffic conflicts* as the mean sojourn time in a given time period (e.g. an hour) that vehicles are exposed to dangerous scenarios (or critical encounters), i.e. the TTCs are lower than a predetermined threshold value τ. Having had the TTC distributions for road tunnel sections (Section 3), the *exposure to traffic conflicts* in an hour can be quantified by.

$$N_{\text{conflict}}(\tau) = (K \times L - 1) \times Pr(\text{TTC}(x) \le \tau) \times 0.5 \quad (10)$$

where K denote the traffic density; L is the length of a road tunnel section; $(K \times L - 1)$ indicates number of gaps in the section; $P(\text{TTC}(x) \le \tau) \times 0.5$ represents the probability of *TTC* less than the threshold value τ[5]; x is the traffic volume of the time period in the road tunnel section. Note that only half of car following scenarios will result in finite TTCs and the other half is considered as absolutely safe situations (infinite TTCs).

4.2. Historical crash-damage database

Historical crash-damage (HCD) database of Singapore is used to examine the relationship between *exposure to traffic conflicts* and

[4] The condition is guaranteed in this study since the traffic volume is with the range from 800 to 1700 vehs/(h lane).

[5] The infinite TTC samples are considered as absolutely safe situations. The coefficient 0.5 is adequate only in stable traffic conditions; the coefficient would be a little greater than 0.5 for unstable traffic flows. For simplicity, the coefficient is assumed to be 0.5 in this study.

Table 4
Traffic volumes, density, length, and crash records for different time periods.

Time period	R-E crash records (2006–2008)	Estimated Traffic volume (vehs/(h lane))	Density (vehs/(km lane))	Average speed (km/h)	Exposure to traffic conflicts		
					2 s	3 s	4 s
7:00 am–8:00 am	11	1600	25	62	657	2024	3548
1:00 pm–2:00 pm	5	1200	16	73	263	829	1502
5:00 pm–6:00 pm	8	1400	20	70	364	1155	2070
8:00 pm–9:00 pm	20	1700	45	39	1566	4673	7998
9:00 pm–10:00 pm	17	1600	50	34	1341	4131	7243
11:00 pm–12:00 am	4	900	11	78	252	777	1374

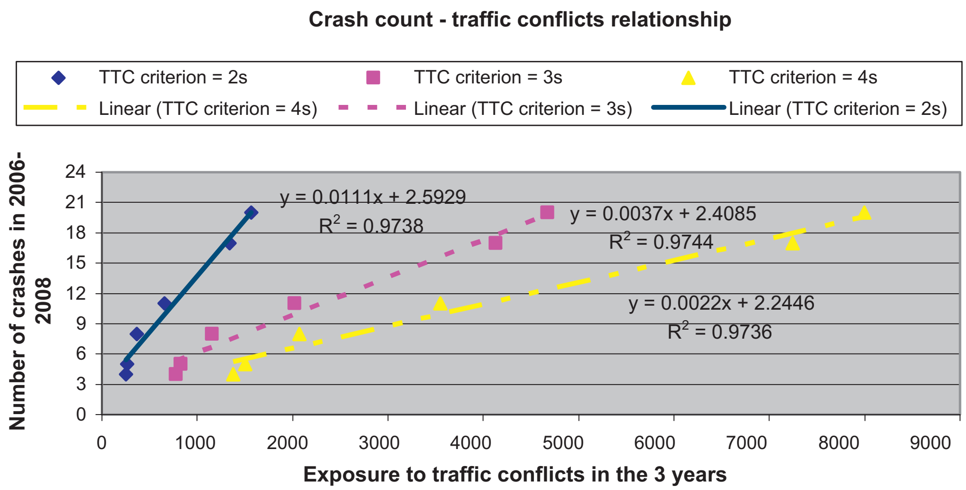

Fig. 7. R-E crash count–traffic conflicts relationship with linear fit.

crash frequencies. According to the Motor Claims Framework (MCF) introduced by the General Insurance Association of Singapore (GIA), in the event of a crash in expressways, everyone involved must inform the insurance company within 1 day using the GIA Motor Accident Report form. In addition, according to Road Traffic Act in Singapore, another report must be made within 24 h of a crash if an injury has occurred. The HCD database (2006–2008) has all the reported crash records, by means of either ways, occurred at Singapore expressways from 2006 to 2008, which includes the time of crash, location of crash, crash type (e.g. rear-end, skidded, etc.), vehicle type (e.g. car–car, car–truck, etc.), number of slight injuries, number of serious injuries, and number of fatalities.[6] To sum up, there are 746 rear-end crashes in the CTE road tunnel from 2006 to 2008, causing 0 fatalities, 45 severe injuries, and 458 slight injuries.

4.3. Relationship between exposure to traffic conflict and crash frequency

2 s, 3 s, and 4 s are considered as the TTC threshold values. From the HCD database we get the crash frequencies in a 1-km road tunnel section in CTE road tunnel are 11, 5, 8, 20, 17, and 4 for he time period 7:00 am to 8:00 am, 1:00 pm to 2:00 pm, 5:00 pm to 6:00 pm, 8:00 pm to 9:00 pm, 9:00 am to 10:00 am, and 11:00 pm to 12:00 am from 2006 to 2008, respectively. In the 1-km tunnel section, there is a 2.4 m wide shoulder and three 3.6 m traffic lanes in each carriageway with a tunnel structural height of approximately 6 m high. Both the curvature and gradient are very gentle in this section. In the current study, we just measure the TTC for vehicles in the mid lane to represent the traffic state. According to the HCD database, only the longitudinal positions of crashes are obtainable and the latitude positions (shoulder lane, mid lane, or median lane) are not reported. Therefore, it would be unlikely for us to disaggregate the three lanes.

We assume that the traffic volumes in the road tunnel section in a specific time period would not have significant daily variations. In accordance with the LTA policy, only the latest 2 years traffic volume data are obtainable. Therefore, the traffic volumes and densities from 2006 to 2008 are not obtainable for this study. The average traffic volumes are estimated by LTA tunnel operators on the basis of the 2010–2011 traffic volume and the summary traffic data in 2006–2008. Accordingly, due to the data unavailability, we just use accurate R-E crash data and estimated average traffic volume to illustrate the methodology in the case study. According to Eq. (9), the *exposure to traffic conflicts* could be calculated. The estimated traffic volumes, densities, lengths, and number of R-E crashes are summarized in Table 4.

4.3.1. Preliminary analysis by using linear regression method

In this section, we analyze the relationship between *exposure to dangerous encounter* and the crash frequency in a linear manner illustrated in Fig. 7. The cumulative residual (CURE) method is a well-recognized method to examine the goodness-of-fit of models in transportation studies (AASHTO, 2010; Hauer, 2004; Hauer and Banfo, 1997). Fig. 8 depicts the cumulative residuals for the linear regression models. As can be seen from Fig. 8, the linear regression models perform well.

[6] According to *the Cost of Road Traffic Accidents in Singapore*, a serious injury is one who has suffer injuries such as fractures or a concussion and/or internal lesions, crushed body parts or organs, severe cuts, or severe general shock requiring medical treatment or hospitalization that prevents the person from performing ordinary tasks for at least 7 days; a slight injury refers to one who is transported to a hospital from the scene in an ambulance or, otherwise, one who requires subsequent medical treatment entailing hospitalization and medical leave of no less than 3 days.

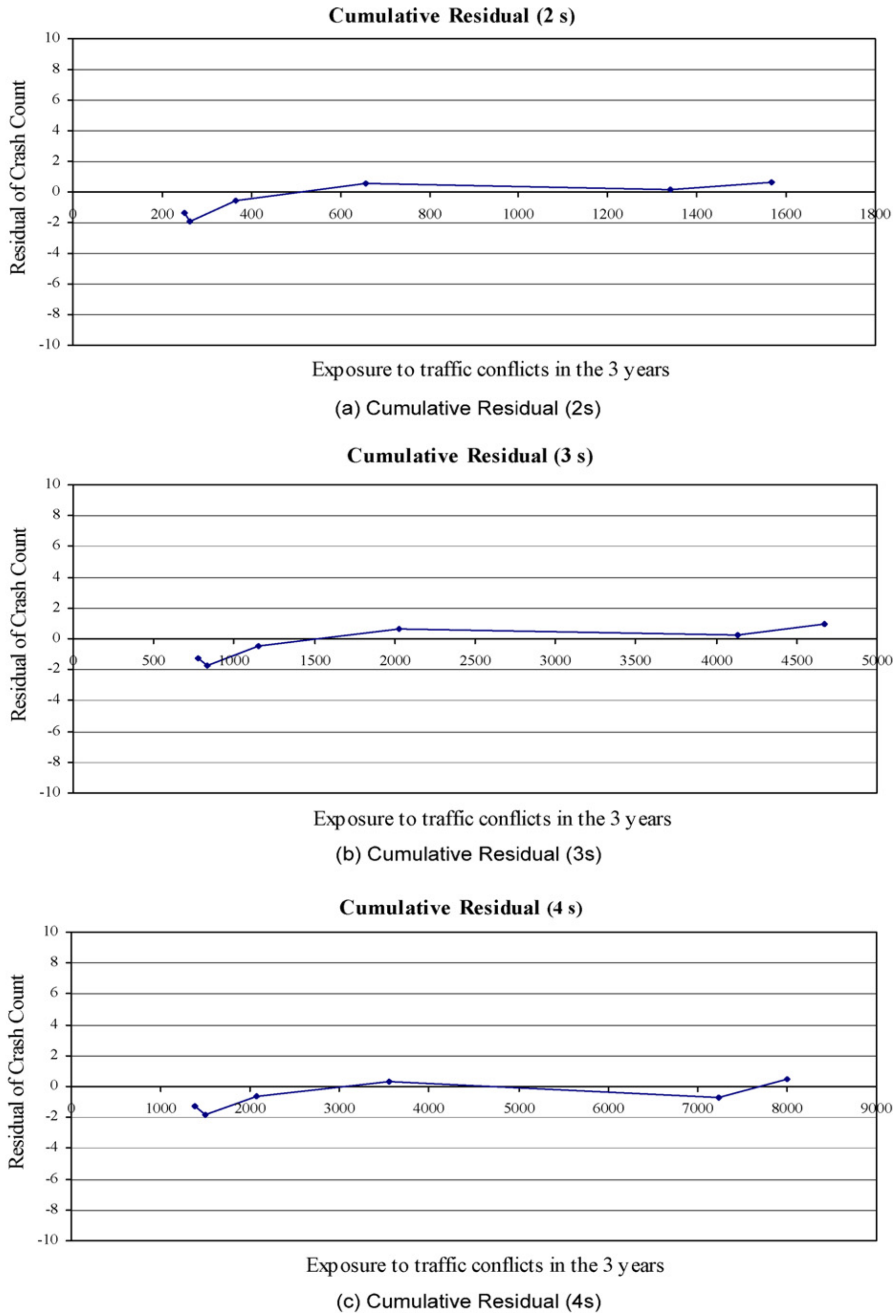

Fig. 8. Plot of cumulative residuals against the proposed safety indicator.

The statistical results for the linear regression models are reported in Table 5. Surprisingly, the *P*-values of the coefficients with respect to constant for the three linear regression models are all greater than 0.035. By contrast, the *P*-values of coefficients with respect to R-E crash frequency are all close to 0. That is to say, the coefficients with respect to intercept are very significant. We further conduct another linear regression model with 0 intercept, as shown in Fig. 9. This indicates that the crash rate may have a proportional linear relationship with the proposed *exposure to traffic conflicts*. The corresponding proportional coefficient is defined as causation factor ($P(t)$) in a linear manner, which could be considered as the conditional probability that vehicle crashes could not be avoided under dangerous encounters for 1 h.

Table 5
Statistical results of linear regression models.

	Constant		Crash frequency		*R*-Sq
	Coefficient	*P*-value	Coefficient	*P*-value	
2 s	2.5929	0.035	0.0111	0.000	0.9738
3 s	2.4085	0.044	0.0037	0.000	0.9744
4 s	2.2446	0.058	0.0022	0.000	0.9736

4.3.2. Negative binomial regression models

As suggested by Hauer et al. (1988), Lord and Mannering (2010), and Miaou and Lum (1993), it is theoretically inappropriate to model discrete and non-negative crash count data using the conventional linear regression method. Generalized linear modelling techniques (GLIM) have the advantages of overcoming the shortcomings associated with linear models. Therefore, the GLIM is

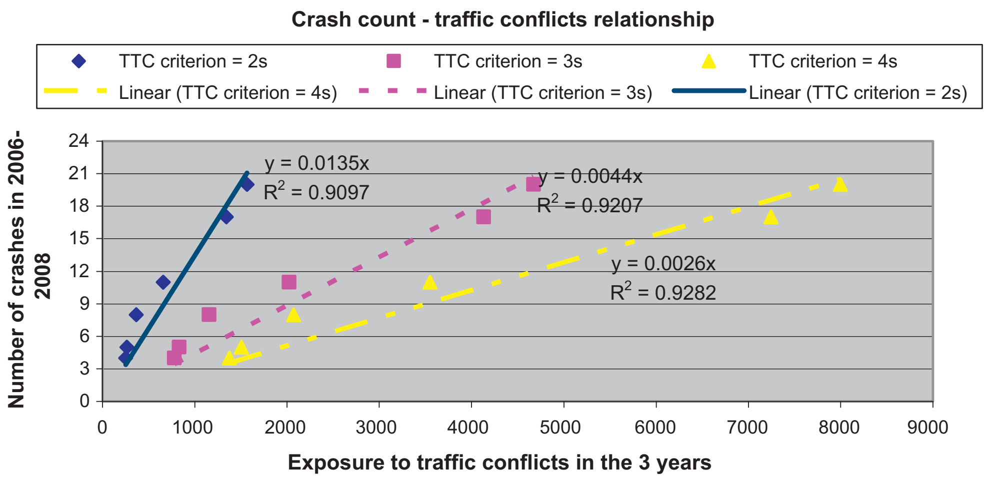

Fig. 9. R-E crash count–traffic conflicts relationship with linear fit (0 intercept).

applied to fit the model using a negative binomial distributed error structure.

The Negative Binomial regression model considered in this study has the form as presented by Miaou (1994).

$$p(Y_i = y_i) = \frac{\Gamma(y_i + (1/\alpha))}{\Gamma(y_i+1)\Gamma(1/\alpha)} \left(\frac{1}{1+\alpha\mu_i}\right)^{1/\alpha} \left(\frac{\alpha\mu_i}{1+\alpha\mu_i}\right)^{y_i} \tag{11}$$

$$y_i = 0, 1, 2, \ldots \tag{12}$$

$$\mu_i = E(Y_i) = N_{\text{conflict},i} \exp(\beta) \tag{13}$$

and the variance of Y_i is

$$Var(Y_i) = \mu_i + \alpha{\mu_i}^2 \tag{14}$$

where Y_i is a random variable representing the number of crashes in time period i; y_i is the actual number of crash count in the time period; $N_{conflict,i}$ is the exposure to traffic conflicts in the time period; $\alpha \geq 0$ and is referred to as dispersion parameter. According to the analysis in Section 4.3.1, it is reasonable to assume the mean value of crash count μ_i or $E(Y_i)$ to be proportional to the exposure to traffic conflicts. This model assumes an exponential rate function $\exp(\beta)$, which ensures that the crash rate is always non-negative.

The parameters (α and β) are estimated by the three approaches (hybrid, fisher, and Newton Raphson methods) using the SPSS software. The three approaches deliver the same estimators for the two parameters, presented in Table 6.

Table 6
Results of negative binomial regression models.

	α	β	*P*-value	Log-likelihood	AIC
Model 1 (2 s)	0.044	−4.114	0.000	−19.777	41.554
Model 2 (3 s)	0.036	−5.244	0.000	−19.760	41.519
Model 3 (4 s)	0.031	−5.814	0.000	−19.767	41.493

From Table 6, we can see the *P*-values are close to 0, indicating that the three models perform well. The Log-likelihood value and the Akaike Information Criterion (AIC) value for each model are also given in the table. Note that estimated models with high Log-likelihood and low AIC values are preferred. Accordingly, the performances of the three models with respect to different TTC thresholds do not have significant differences. Table 7 depicts the estimated expected values of crash counts by the three models and the actual crash count for the six data points in this study, which also indicate that the models perform very well. The proportional coefficients of the expected values of crash counts (with negative binomial assumption) over the exposure to traffic conflicts are $e^{-4.114} = 0.0163$, $e^{-5.244} = 0.0053$, and $e^{-5.814} = 0.0030$, respectively.

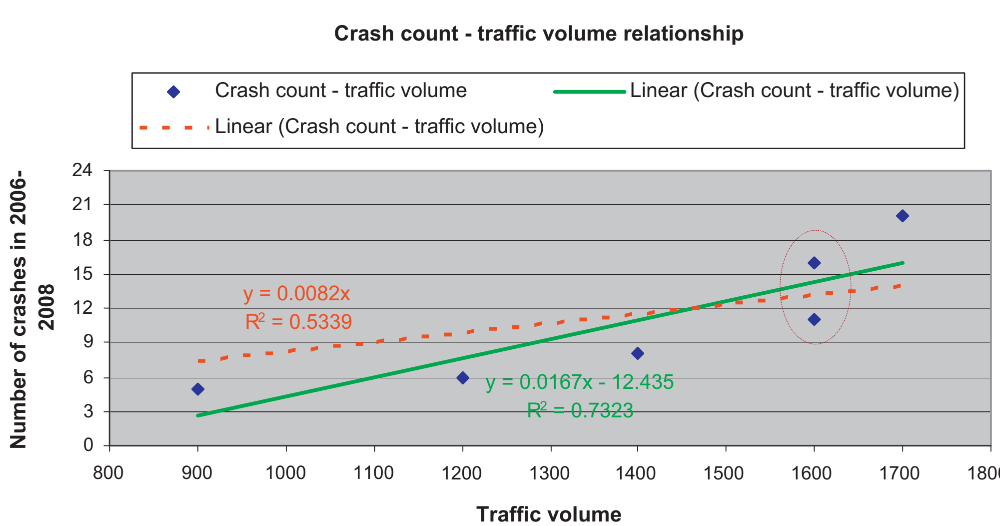

Fig. 10. Crash count–traffic volume relationship.

Table 7
Estimated expected values of crash counts.

R-E crash records (2006–2008)	Estimated by model 1 (2 s)	Estimated by model 2 (3 s)	Estimated by model 3 (4 s)
11	10.74	10.68	10.59
5	4.30	4.38	4.48
8	5.95	6.10	6.18
20	25.59	24.67	23.87
17	21.91	21.81	21.62
4	4.12	4.10	4.10

These coefficients are the causation factors in regression models with generalized linear manner.

4.4. Remarks on the crash count–traffic flow relationship

As for the crash count–traffic flow relationship, a series of studies were carried out by several scholars on the basis of the actual data. Jovanis and Chang (1986) proposed a method to model the relationship between miles travelled and crash count in a time-space domain. Hauer et al. (1988) developed a model to estimate the safety of a signalized intersection on the basis of the traffic flow and crash count. Miaou and Lum (1993) compared four regression models – two conventional linear regression models and two Poisson regression models – in terms of their ability to model the relationship among traffic flow, geometry design, and crash count in highways. All the above-mentioned studies acknowledged that it was not appropriate to apply the conventional linear regression to model crash count and traffic flow. Fig. 10 depicts the crash count–traffic volume relationship for the present study. As can be seen in the figure, neither of the linear regression ($R^2 = 0.7323$) and proportional regressions ($R^2 = 0.5339$) performs well. Since traffic volume equals to the product of density and speed (Lieu et al., 1999) the densities and speeds for two road sections with the same traffic volume may not be the same (e.g. the two points circled in Fig. 9). In reality, the crash frequency is also closed related to the speed and density of a road section (Aarts and van Schagen, 2006). Hence, it is not appropriate to assume the linear relationship between crash count and traffic volume.

Comparatively, as can be seen from Figs. 7 and 9, both linear regression and proportional regression perform well in this study. This is because not only traffic volume but also density is taken into account in the proposed exposure to traffic conflicts. The results shows that, in Singapore's road tunnels, the exposure to traffic conflicts based method outperforms the traffic volume based approach.

5. Discussions

Theoretically, linear regression models are not appropriate to model discrete and non-negative crash count data. GLIM is proven to be more effective to formulate the rare events such as crash count. However, as illustrated in Section 4.3.1, the linear regression models also perform well according to the CURE method and correlation coefficients. Therefore, it is also acceptable to formulate the relationship between crash count and proposed index in this study. The coefficients between crash counts (or expected values of crash counts) and the exposure to traffic conflicts are defined as the causation factor in this study. The proposed causation factor $P(t)$ reflects the conditional probability that vehicle crashes have occurred when the vehicle are exposed to dangerous scenarios for 1 h. The probability would be dependent on vehicle conditions, drivers' abilities, and the road geometries. We conjecture that this factor could be a constant for a given road tunnel section in the long run with a given TTC threshold value. The TTC values would generally have a parabola relationship with traffic volume because they will be affected by not only speed dispersion but also distance headways. For non-interrupt traffic flows with traffic volume from 900 vehs/(h lane) to 1700 vehs/(h lane), the TTC distributions may follow the Inverse Gaussian distributions (lognormal distributions are also a good approximation) and traffic volume could be considered as the contributing factor to the distribution parameters. It should be pointed out that these perspectives need to be validated using more actual data from other expressways and/or urban road tunnels.

The crash data from Singapore's road tunnels shows that linear or proportional relationship may not be good enough to reflect the relations between crash count and traffic volume. Instead, the linear and proportional relationships perform very well between crash count and exposure to traffic conflicts. This may be because not only traffic volume but also density is taken into consideration in the proposed exposure to traffic conflicts.

Other than the TTC, the deceleration rate to avoid the crash (DRAC) and the post encroachment time (PET) have also been considered as good safety indicators to measure the safety level in roads (Meng and Weng, 2011; Cunto and Saccomanno, 2008). Further study may be conducted to establish the relationship between crash frequency and the above-mentioned two safety indicators. The comparative analysis of these three safety indicators could also be studied accordingly. In addition, the model can also be applied to identify the hotspots in the urban road tunnels and/or expressways (Cheng and Washington, 2005; Montella, 2010).

6. Conclusions

In this study, a novel approach is proposed to estimate the R-E crash frequency in road tunnels. We first conclude that Inverse Gaussian distribution is the best-fitted distribution to the TTC data based on the best-fit analysis. Accordingly, an Inverse Gaussian regression model is applied to establish the relationship between the TTC and the corresponding traffic volume. A new concept of *exposure to traffic conflicts* is defined as the mean sojourn time in a given time period that vehicles are exposed to dangerous scenarios, namely, the TTCs are lower than a predetermined threshold value. A R-E crash frequency estimation model is then proposed on the basis of the accident records provided by the HCD database for Singapore's road tunnels. We find that the expected value of crash frequency has a proportional linear relationship with the proposed *exposure to traffic conflicts*. Finally, several conjectures and recommendations are proposed.

It should be pointed out that the analysis with limited data samples in this study may not be adequate to validate the relationship between crash count and the proposed indicator. A further study with more data is needed to further valid and calibrate the relationship.

Acknowledgments

We are really grateful to the two anonymous referees and the editors whose comments improved the presentation and the content of the earlier version. Special thanks will also be expressed to Ms Soh Ling Tim from Land Transport Authority of Singapore on the data collection for this project. This study is supported by the innovation fund of Land Transport Authority of Singapore.

References

Aarts, L., van Schagen, I., 2006. Driving speed and the risk of road crashes: a review. Accident Analysis and Prevention 38, 215–224.
AASHTO, 2010. Highway Safety Manual, 1st edition.
Beard, A., Carvel, C., January 1, 2005. The Handbook of Tunnel Fire Safety. Thomas Telford Publishing, London.
Cheng, W., Washington, S., 2005. Experimental evaluation of hotspot identification methods. Accident Analysis and Prevention 37, 870–881.

Cunto, F., Saccomanno, F.F., 2008. Calibration and validation of simulated vehicle safety performance at signalized intersections. Accident Analysis and Prevention 40, 1171–1179.
Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accident Analysis and Prevention 42, 393–402.
Farah, H., Bekhor, S., Polus, A., 2009. Risk evaluation by modeling of passing behavior on two-lane rural highways. Accident Analysis and Prevention 41, 887–894.
Guo, F., Wang, X., Abdel-Aty, M., 2010. Modeling signalized intersection safety with corridor spatial correlations. Accident Analysis and Prevention 42, 84–92.
Haque, M.M., Chin, H.C., Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis and Prevention 42, 203–212.
Hauer, E., 2004. Statistical road safety modeling. Transportation Research Record 1897, 81–87.
Hauer, E., Banfo, J., 1997. Two tools for finding what function links the dependent variable to explanatory variables. In: Proceedings ICTCT 97 (International Cooperation on Theories and Concepts in Traffic Safety), Lund, Sweden, pp. 1–7.
Hauer, E., Ng, J.C.N., Lovell, J., 1988. Estimation of safety at signalized intersections. Transportation Research Record 1185, 48–61.
Hauer, E., 2001. Overdispersion in modelling accidents on road sections and in Empirical Bayes estimation. Accident Analysis and Prevention 33 (6), 799–808.
Hayward, J.C., 1972. Near Miss Determination Through Use of a Scale of Danger (Traffic Records 384). Highway Research Board, Washington, DC.
Hirst, S., Graham, R., 1997. The format and presentation of collision warnings. In: Noy, N.I. (Ed.), Ergonomics and Safety of Intelligent Driver Interfaces.
Hogema, J.H., Janssen, W.H., 1996. Effects of Intelligent Cruise Control on Driving Behavior. TNO Human Factors, Report TM-1996-C-12, Soesterberg, The Netherlands.
Ibeas, Á., Cordera, R., dell'Olio, L., Moura, J.L., 2011. Modelling demand in restricted parking zones. Transportation Research Part A 45, 485–498.
Jovanis, P.P., Chang, H.L., 1986. Modeling the relationship of accidents to miles traveled. Transportation Research Record 1068, 42–51.
Land Transport Authority (LTA), 2005. Design Safety Submission for KPE [internal report], Singapore.
Lieu, H., Gartner, N., Messer, C.J., Rathi, A.K., 1999. Traffic Flow Theory. U.S. Department of Transportation, Federal Highway Administration.
Lord, D., Geedipally, S.R., Guikema, S., 2010. Extension of the application of Conway–Maxwell–Poisson models: analyzing traffic crash data exhibiting under dispersion. Risk Analysis 30, 1268–1276.
Lord, D., Guikema, S., Geedipally, S.R., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. Accident Analysis and Prevention 40, 1123–1134.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation Research Part A 44, 291–305.
Lord, D., Washington, S.P., Ivan, J.N., 2007. Further notes on the application of zero inflated models in highway safety. Accident Analysis and Prevention 39, 53–57.
Malyshkina, N., Mannering, F., 2010a. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis and Prevention 42, 131–139.
Malyshkina, N., Mannering, F., 2010b. Zero-state Markov switching count-data models: an empirical assessment. Accident Analysis and Prevention 25, 77–84.
Maycock, G., Hall, R.D., 1984. Accidents at 4-Arm Roundabouts. TRRL Laboratory Report 1120, Transportation and Road Research Laboratory, Crowthorne, UK.
Meng, Q., Weng, J., 2011. Evaluation of rear-end crash risk at work zone using work zone traffic data. Accident Analysis and Prevention 43, 1291–1300.
Meng, Q., Qu, X., Wang, X., Yuanita, V., Wong, S.C., 2011a. Quantitative risk assessment modeling for non-homogeneous urban road tunnels. Risk Analysis 31, 382–403.
Meng, Q., Qu, X., Yong, K.T., Wong, Y.K., 2011b. QRA model based risk impact analysis of traffic flow in urban road tunnels. Risk Analysis 31 (12), 1872–1882.
Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accident Analysis and Prevention 25, 689–709.
Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accident analysis and Prevention 26, 471–482.
Minderhoud, M.M., Bovy, P.H.L., 2001. Extended time-to-collision measures for road traffic safety assessment. Accident Analysis and Prevention 33, 89–97.
Montella, A., 2010. A comparative analysis of hotspot identification methods. Accident Analysis and Prevention 42, 571–581.
Páez, A., Trépanier, M., Morency, C., 2011. Geodemographic analysis and identification of potential business partnerships enabled by transit smart cards. Transportation Research Part A 45, 640–652.
PIARC Technical Committee C3.3 Road Tunnel Operation, 2008. Risk Analysis for Road Tunnels, May 2008, http://publications.piarc.org/ressources/publications_files/4/2234,TM2008R02-WEB.pdf, accessed 19 May 2011.
Qu, X., Meng, Q., Yuanita, V., Wong, Y.K., 2011. Design and implementation of a quantitative risk assessment software tool for Singapore's road tunnels. Expert Systems with Applications 38 (11), 13827–13834.
Shankar, V., Milton, J., Mannering, F.L., 1997. Modeling accident frequency as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention 29, 829–837.
Svensson, A., 1998. A method for analyzing the traffic process in a safety perspective. Doctoral Dissertation, University of Lund, Lund, Sweden.
The European Parliament and the Council of the European Union, 2004. Directive 2004/54/EC of the European parliament and of the council of 29 April 2004 on minimum safety requirements for tunnels in the Trans-European Road Network. Official Journal of the European Union (April), L 167/39-91.
Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accident Analysis and Prevention 35, 427–433.
Whitmore, G.A., 1983. A regression method for censored Inverse Gaussian data. Canadian Journal of Statistics 11, 305–315.
Zhang, Y., Xie, Y., 2007. Forecasting of short-term freeway volume with v-support vector machines. Transportation Research Record 2024, 92–99.