Analysis of crash frequency in motorway tunnels based on a correlated random-parameters approach

Ciro Caliendo*, Maria Luisa De Guglielmo, Isidoro Russo

Department of Civil Engineering,University of Salerno,Via Giovanni Paolo II,132-84084 Fisciano (SA),Italy

Abstract

The paper provides an analysis of the frequency of total accidents (accidents involving material damage, physical injuries and fatalities), which occurred in 226 unidirectional motorway tunnels over a four-year monitoring period, based on unrelated and correlated random-parameter models. The so-called random-intercept model, in which only the regression intercept is assumed to be random, was also developed a priori for recording the random-effects (temporal correlations among accidents occurring in the same tunnel in different years). The independent variables were: tunnel length (*L*), annual average daily traffic (*AADT*) per lane, percentage of trucks (*%Tr*), presence of a sidewalk (*SW*), longitudinal slope (*LS*), and mechanical ventilation (*MV*). The comparison among the aforementioned three models showed that the correlated random-parameters model, which takes into account the cross correlation among the random-parameters, provided a better goodness-of-fit than the corresponding uncorrelated random-parameter and intercept-random models. This means that more precise estimations of accidents can be obtained when the random-parameters are assumed to be correlated in statistical analysis. The developed model also offers additional insights into showing how different combinations of parameters affect tunnel safety. In particular, through the correlation coefficient matrix of random-parameters, we found that the non-constant longitudinal slope (*LS*) alleviates the effect of the annual average daily traffic (*AADT*) per lane on increasing crash frequency. In addition, the presence of the mechanical ventilation (*MV*) in tunnels makes less significant the influence of AADT per lane on increasing crash frequency, too. The knowledge of these correlations may be useful for future applications, for example, for road engineers in designing tunnel. The model proposed can also be used by Tunnel Management Agencies (TMAs) for estimating possible variations in accident frequency in a specific tunnel due to modifications of traffic control systems.

Keywords: *Crash frequency, Correlated random-parameters model, Poisson distribution, Binomial Negative distribution, Road tunnels.*

(*) Corresponding Author: Tel.: +39+89+964140; fax: +39+89+964045

E-mail Address: ccaliendo@unisa.it (C. Caliendo)

1. Introduction

Roads accidents throughout the world, according to the World Health Organization (WHO, 2015), cost the lives of 1.25 million people every year. Moreover, these accidents are the main cause of death among young people aged between 15 and 29 years (over 300,000 deaths). In the 28 countries of the European Union (EU28, 2016) the number of road accident fatalities was 25,720 in 2016. With regard to Italy, the Italian National Institute for Statistics (ISTAT, 2016) reports that in the same year 2016 road accidents caused 3,283 deaths and 249,175 injuries.

Road accidents cause so much pain and suffering which cannot be measured or quantified, as well as high social costs in terms of medical expenses and lost productivity. All this costs governments approximately 3% of the Gross Domestic Product (WHO, 2015).

The high human and economic toll that road accidents continue to inflict on society is no longer sustainable. Moreover, it is to be said that the sustainable development goal fixed by the WHO, which included the target of a 50% reduction in road traffic deaths and injuries by 2020, does not appear achievable.

However, so much effort has unquestionably been exerted over time to improve road safety that nowadays we can observe a drastic reduction in serious accidents compared to those of previous years. This is due more especially to: (i) the enforcement of road safety policies and road user behaviour (e.g., reducing speed, increasing helmet and seat-belt use, reducing drink- and drug-driving); (ii) safer vehicles (e.g., better standards protecting vehicle occupants, electronic stability control, new features on board to avoid collisions with pedestrians); (iii) safer roads (higher levels of infrastructure design and maintenance, also with reference to non-motorized road users). However, in contrast with other main transportation systems (such as air, water, and railway transport) where human behaviour in driving is widely controlled by safety protocols and sophisticated electronic equipment, the road transport system still remains quite dangerous. This might be attributed to different physical and mental capabilities of drivers, dissimilar perceptions of risk level, different reactions to external stimuli. In addition, distracted driving and/or talking on mobile phone while driving, are serious and ever growing threats to road safety.

The recent introduction of autonomous driving vehicles, which can potentially remove human error, might eventually lead to significant advantages in road safety. However, the large-scale spread of these vehicles is nowadays still questionable due to both the occurrence of accidents that caused the death of some drivers during tests and the interaction with other vehicles when variations in traffic flow and road characteristics are expected, as well as driving in poor weather conditions.

While the success of the aforementioned efforts in reducing road accidents and mitigating their impact cannot be denied, research based on the statistical analysis of accident data-bases is also recognized as being a useful tool for contributing in the sustainable development of road-safety policies towards saving lives and reducing the severity of injuries. However existing data-bases, which usually extract data from police accident reports do not contain all the information necessary for defining human behaviour accurately, traffic and road characteristics, weather, and in what way drivers react to continuous variations in environmental conditions while driving. In other words other elements remain unobserved by analysis. In the light of this limit of traditional accident data-base, more advanced statistical methods have been developed in order to address this relevant research question known to analysts as “unobserved heterogeneity”.

Characteristics of horizontal and vertical alignment may cause unobserved heterogeneity through observations due, for instance, to different reactions of drivers to road geometry.

Traffic also may be responsible for unobserved heterogeneity through observations attributable, for example, to variations in driving style of road users in response to traffic modifications.

Vehicle characteristics may be another example of unobserved heterogeneity caused by size, weight, structure of vehicle, as well as by vehicle speed differences.

Moreover, human stimuli linked, for instance, to gender, age, height, weight, educational level, which are generally unavailable for analysis, may also cause unobserved heterogeneity through observations.

Finally, when some of the most important explanatory variables affecting accidents are omitted, because they have not been observed (and therefore not reported in the data-base), additional unobserved heterogeneity is introduced.

Several statistical methods have been applied over the years for road accident analysis in order to develop predictive models and evaluate the significance of variables affecting crashes. In the last decade, random-parameters models have been more especially considered. These models try to record the aforementioned unobserved heterogeneity (unobserved factors that may vary through accident observations) by making it possible that the regression model parameters are random, in contrast with the assumption that parameters are fixed through observations (traditional statistical approach).

A special case in the class of random-parameters models, is the so-called random-intercept model in which only the regression intercept is assumed to be random. Therefore, this type of model records only an unobserved specific heterogeneity which is unrelated to explanatory variables, for example the so-called unobserved temporal heterogeneity. This unobserved heterogeneity, varying over time on a specific segment of a road, might include weather-related factors and the reaction of drivers in response to weather variations that cannot be observed by the analyst. In keeping with the international literature, this model is known to be equivalent to the random-effects one. Random-effects models are necessary when temporal correlations (data collected on the same road section over successive time periods) or spatial correlations (data collected from the same geographic area) are thought to be present.

Generally speaking, it is to be said that in the field of accident analysis it is assumed that the effects of unobserved heterogeneity are independent one from another. In other words, it is often assumed that there is no correlation among parameters that are found to be random in the regression model developed. However, this might not be true, and ignoring the interaction among random-parameters might lead to less precise accident predictions. This is an issue that researchers have little investigated over time, and so very few studies are available in the literature (some references are reported in the section background of the present paper) and indicate a gap in our knowledge. Accounting for correlation among random-parameters can be achieved by means of applying the so-called correlated random-parameters approach, which is the main scope of the present paper.

Various studies exist in the literature on unobserved heterogeneity associated with accidents occurring on open roads, while the occurrence of accidents inside road tunnels have been much less investigated. Driving behaviour in road tunnels is different when compared to that on open roads. Driving in tunnels may cause anxiety because tunnels are dark, narrow and monotonous. Moreover, drivers may also be frightened of hitting other vehicles and the tunnel wall, or of dangerous scenarios such as fires. Moreover, the sunshine reflected from the tunnel portal might cause ocular blinding of drivers before entering the tunnel. Therefore, there is evidence that accidents in road tunnels need to be better examined; which is another intent of this paper. In this respect it is to be mentioned that, according to an Italian study (Caliendo and De Guglielmo, 2012), the average rate of severe accidents (accidents involving only injuries and fatalities) was computed to be of 12 severe accidents /10^8veh./km in Italian motorway tunnels against 9 severe accidents /10^8 veh./km on the corresponding motorways containing the tunnels investigated. Also Zhuang-lin et al. (2009) showed that the severity of injuries is higher in Chinese freeway tunnels than on open roads. Yeung et al. (2013) revealed, on the other hand, that traffic accident rates in Singapore expressways tunnels were higher in the transition zones rather than in the interior zones of tunnels. The small number of studies on the parameters that influence traffic safety in road tunnels represents, according to the authors of this article, a further lack of knowledge.

As a result of the above considerations, there are at least three main reasons for justifying this paper. The first is motivated by the need to investigate in greater depth on crash frequency (accidents per year) without ignoring possible interactions among random-parameters. In this respect, predictive models based on a correlated random-parameters approach should be developed so that new insights might be provided for increasing our knowledge. The second, given the lacuna of studies on accidents occurring in road tunnels, is to have a better understanding of the relationships between crash frequency in road tunnels and geometric-traffic variables; as well as providing additional insights for road engineers in designing tunnel. In fact there is a priori reason for believing that the effects of unobserved heterogeneity, associated with accidents in road tunnels, might be different from those of unobserved heterogeneity on open roads. Finally, since the statistical methodology generally used in accident analysis is based on the Negative Binomial (NB) distribution, it is interesting to examine in greater depth whether the NB model may converge to the Poisson model in presence of random-parameters assumed to be unrelated or correlated.

Therefore, the present paper may make a significant contribution to the state-of-the-art by showing the effects of combined parameters on the occurrence of accidents, and by making proposals in the design of tunnel geometry in order to improve safety.

The paper is organized as follows: the next section deals with the statistical methodology applied. Then the data set used and the process of preparing it in statistical analysis are described. Subsequently the results are presented and discussed and appropriate comparisons are made among the predictive models developed. Finally comments for practical applications, conclusions and addresses for additional research are made.

2. Background

Considerable literature exists on the merits and demerits both of traditional statistical models and new ones. Some international references can be found more especially in Washington et al. (2011), Mannering and Bhat (2014), Mannering et al. (2016); as well as with reference to Italian studies in Caliendo et al. (2007, 2013), Caliendo and Guida (2014). However, in the last decade, the potentiality of random-parameter models compared to the fixed-parameters one has been particularly investigated. Random-parameter models are considered in Milton et al. (2008), Anastasopoulos and Mannering (2009), Anastasopoulos and Mannering (2011), Caliendo et al. (2016), Caliendo and De Guglielmo (2017). It is to be remembered that the model in which only the regression intercept is random, in the literature, is known to be equivalent to the random-effects model (see Greene (2007); and Hilbe (2007) for in-depth knowledge).

However, it is to be stressed that in the aforementioned random-parameters models the distribution of random-parameters is assumed to be independent. In other words, the potential effects of the cross correlation among random-parameters cannot be captured. From the viewpoint of statistical modelling, according to Conway and Kniesner (1991) ignoring the correlations among random-parameters might lead to biased estimations. As a result, researchers have recently been trying to expand methodological frontiers by taking into account the issue mentioned above.

Yu et al. (2015), for example, adopted a correlated random-parameter approach in order to investigate the effects of weather conditions on the risk of crashes on freeways. In their study they found that the correlated random-parameters Tobit model was statistically superior to the corresponding unrelated-parameters model. By investigating the crash frequency of many cities, Coruh et al. (2015) showed that the correlated random-parameters Negative Binomial (NB) model is more appropriate than the unrelated random-parameters NB model. In analysing crash frequency in the tunnels of China, Hou et al. (2018) also proved that the correlated random-parameters NB model provided a better goodness-of-fit compared to the unrelated one.

It is to be stressed that the present paper is different from our previous studies mentioned above: (i) for introducing two additional independent variables such as the longitudinal slope and mechanical ventilation; (ii) for non-assigning a priori a functional form for describing the dependence of the expected number of accidents on traffic flow (i.e., we did not use the Hauer and Bamfo approach (1997)); (iii) for taking into account possible correlations among the variables investigated.

However, it is evident that very few studies are hitherto available in the literature concerning the correlated random-parameters models. Moreover, only one study deals with accidents occurring in tunnels. In addition, it is to be stressed that the results mentioned above, which are based only on Chinese tunnels, might not be directly transferable to other geographic areas where traffic, tunnel geometry, and driving style are different. This may be the case of Italy. Therefore, given the lacuna of studies on the aforementioned issues and doubts on transferability of results to other countries, the potential of considering combined random-parameters should be investigated in greater depth, and the frontier methodological should be expanded in applying this statistical approach, which appears to be innovative in the field of accident analysis, also to Italian tunnels. This might make additional knowledge possible and help in tunnel design.

3. Methodology

The statistical methodology which is generally applied in crash analysis is based on the assumption that the fluctuation of accident counts, say Y_i, which occur on a road section i (or within a tunnel) during the observational time interval, is a Negative Binomial (NB) random variable. Unlike the Poisson model, the NB model allows for the variance of accident counts to be greater than the mean, provided by the so-called "overdispersion parameter" (α). When the "overdispersion parameter" α of the NB model converges to zero (i.e., the "inverse dispersion" parameter $\varphi = 1/\alpha$ diverges to infinity), it suggests that the corresponding Poisson model is still statistically appropriate.

A regression model of the expected number of accidents is defined by typically using the log-linear function,

$$\lambda_i = exp\left(\boldsymbol{X}_i^T \boldsymbol{\beta}\right) \tag{1}$$

where $\boldsymbol{\beta}$ is a vector of fixed, even if unknown, coefficients and $\boldsymbol{X}$ is a vector of k covariates.

In order to take into account unobserved heterogeneity which may vary across road sections (or tunnels), a suitable approach is, however, considering the $\boldsymbol{\beta}$ parameters as random variables, by assuming:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \delta_i \tag{2}$$

where δ_i is a random variable with some probability density function $f(\delta)$, for example a normal variable with zero mean and constant standard deviation. If the variance of the chosen distribution is not significantly different from zero, it means that a conventional fixed-parameters model is still statistically appropriate.

The aforementioned approach assumes that the possible unobserved heterogeneity can be captured by considering that the random-parameters are independent (i.e., there is no correlation among random-parameters in the model). Indeed, there might be correlation among random-parameters so that a correlated approach is needed.

In order to take into account the potential correlation among the random parameters, $\boldsymbol{\beta}_i$ is assumed to follow a multivariate normal distribution written as follows:

$$\boldsymbol{\beta}_i = \boldsymbol{b} + \boldsymbol{C}\boldsymbol{\omega}_i \tag{3}$$

with $\boldsymbol{b}$ and $\boldsymbol{C}$:

$$\boldsymbol{b} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_{j-1} \\ \beta_j \end{bmatrix} and \boldsymbol{C} = \begin{bmatrix} (\sigma_1)^2 & \sigma_{1,2} & \cdots & \sigma_{1,j-1} & \sigma_{1,j} \\ \sigma_{2,1} & (\sigma_2)^2 & \cdots & \sigma_{2,j-2} & \sigma_{2,j} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{j-1,1} & \sigma_{j-1,2} & \cdots & (\sigma_{j-1})^2 & \sigma_{j-1,j} \\ \sigma_{j,1} & \sigma_{j,2} & \cdots & \sigma_{j,j-1} & (\sigma_j)^2 \end{bmatrix} \tag{4}$$

In equation (3) $\boldsymbol{\beta}_i$ is own the random-parameters vector for observation i, $\boldsymbol{b}$ is the mean vector, C is the variance-covariance matrix **(**$\boldsymbol{C}$ engenders correlation among the elements of the parameters vector $\boldsymbol{\beta}_i$), j is the number of random parameters, and $\boldsymbol{\omega}_i$is the randomly and independently distributed uncorrelated vector term. Obviously in the unrelated random-parameters models, the off-diagonal elements in the variance-covariance matrix are equal to zero.
The model parameters were estimated by a maximization procedure of the likelihood function. It is to be noted, however, that traditional procedures based on numerical maximization of log-likelihood are, in most cases, unfeasible when the number of unknown parameters is high. Thus, Green (2007) developed an efficient simulation procedure, based on the Halton draws (see in detail Halton (1960)), which allows estimating the model parameters even in the presence of a high number of random parameters. In this respect, an approach based on 2000 Halton draws is used in the present paper in order to obtain an efficient convergence for numerical integration in estimation.

4. Data description

A 4-year monitoring period, extending from 2006 to 2009, was considered for Italian motorway tunnels. A number of 226 unidirectional tunnels with two lanes were in particular investigated. The data-base consists of 1930 total accidents (i.e., accidents involving material damage, injuries and fatalities). The number of total accidents observed for each year was as follows: 664 (2006), 545(2007), 384 (2008), 337 (2009). The width of each lane is 3.75 m, the presence of a sidewalk (with a width approximately of 1.0 m) was recorded in 143 tunnels, while the emergency lane was always absent in all tunnels considered. In this respect, it is to be said that an emergency lane is not generally present in the existing Italian tunnels that were designed and built several years ago, in contrast with the corresponding open road sections. This is attributable to the decision to reduce tunnel construction costs. However, nowadays for the design of new motorway tunnels, according to the more recent Italian standards (MIT, 2002), the emergency lane must be introduced also in tunnels (as on open roads) in order to achieve higher safety levels.
The length of the tunnels investigated varies between 0.39 km and 3.25 km. The longitudinal slope is constant in 144 tunnels (93 with ascending gradients, and 51 with descending gradients), while varies in the remaining 82 tunnels with positive and negative gradients. The mechanical ventilation system, which is longitudinal with jet-fans, is present in 55 tunnels and absent in the remaining 171 tunnels.
Unidirectional traffic flow in each tunnel, which is expressed in terms of the annual average daily traffic (AADT), was found to be between 2250 and 20,380 vehicles/day per lane. The percentage of trucks ranged from 15% to 31%. Summary statistic of tunnel length, AADT per lane and percentage of trucks are given in Table 1.

Table 1 Summary statistic of variables

Variables	Mean	Mode	Standard deviation	Minimum	Maximum
Length [km]	1.20	0.53	0.64	0.39	3.25
AADT /10,000 per lane [veh./day]	0.7620	0.3063	0.4224	0.2250	2.0380
Percentage of trucks [%]	21%	16%	4%	15%	31%

It is to be mentioned that the aforementioned data-base of accidents is based on official data available at the Italian Ministry of Infrastructures and Transport. As far as the authors are aware, an update of these data (i.e. accidents occurring in tunnels in more recent years) was not yet available for a significant number of tunnels at the time of the writing of this paper. However, the main intent of this paper is not to show the trend of accidents occurring in tunnels over time (even if according to the Italian National Institute of Statistics (ISTAT, 2016) accidents continue to decrease on open roads; the same thing which might be expected also in tunnels); but to individualise a model which was statistically superior for taking into account the unobserved heterogeneity in accident data, and which provided more precise predictive estimations of accidents in tunnels. Therefore, the authors are confident that the following analysis, which is based on the aforementioned data-base, is appropriate in relation to the designated objective.

5. Statistical analysis of data

Set of variables

The dependent variable is assumed to be the frequency of total accidents (number of total accidents per year). The independent variables are: length (*L*), average annual daily traffic (*AADT*) per lane, percentage of trucks (*Tr%*), sidewalk (*SW*), longitudinal slope (*LS*), and mechanical ventilation (*MV*). The variables SW, LS, and MV are assumed to be dummy variables: SW = 1 if there is a sidewalk, 0 otherwise; LS = 1 if the longitudinal slope of tunnel varies with ascending and descending gradient, 0 otherwise; MV = 1 if there is the mechanical ventilation, 0 otherwise.

Since the intent of this paper is also to record (within the aforementioned monitoring period) the year effect on the occurrence of total accidents in tunnels, the dummy variables year 2007, year 2008, and year 2009 were also considered, by assuming 2006 as year of reference.

Procedure for choosing the statistically more significant variables

Since some of the potential covariates could actually have very little or even no effect on accident counts, a procedure based on the Likelihood Ratio Test (LRT) was used in order to decide which subset of the full set of potentially explanatory variables should be included in the regression model. The likelihood-ratio statistic

$$\Lambda = 2\left[l(\hat{\boldsymbol{\beta}}, \hat{\varphi}) - l(\hat{\boldsymbol{\beta}}', \beta_k = 0, \hat{\varphi}) \right] \tag{5}$$

where $l(\hat{\boldsymbol{\beta}}, \hat{\varphi})$ is the log-likelihood of the regression model containing all the covariates, and $l(\hat{\boldsymbol{\beta}}', \beta_k = 0, \hat{\varphi})$ is the log-likelihood of the regression model with the *k-th* covariate out, is asymptotically distributed as a chi-squared distribution (χ^2) with 1 degree of freedom. Variables were assumed to be significant in each model at a significance level of much less than 0.05 (this reflects at least a 95% confidence).

The aforementioned LRT statistic was also used between two competing models in order to test the statistical superiority of the target model. In this case, the χ^2 statistic is with the degrees of freedom depending on the difference between two competing models. For example, in the present paper the LRT was also used to decide if the unrelated random-parameters model were statistically superior to the corresponding random-intercept model; in this case the number of degrees of freedom (dof) is equal to the number of parameters that are effectively random.

The LRT was also applied for understanding if the correlated random-parameters model were statistically better than the unrelated random-parameters one. In the comparison between these two models the number of degrees of freedom is $k = m((m-1)/2)$, where m is the number of the random parameters that are effectively correlated.

However, first of all, on the basis of the aforementioned data-base, a random-intercept Negative Binomial (RINB) model was developed. Moreover, a random-parameters Negative Binomial model in which both the intercept and parameters are allowed to vary randomly (RPINB) was also developed. It was found that the "inverse dispersion" parameter (φ) both of the RINB model and RPINB one diverged respectively to infinity (i.e., the over-dispersion parameter α converged to zero), thus indicating that the outcomes of RINB and RPINB models converged to the corresponding Poisson models. For these reasons only the Poisson model with the random intercept, as well as the Poisson model where both the intercept and regression parameters are allowed to vary randomly, are more especially commented in this paper.

Panel data

The statistical analysis presented in this paper involves the number of total accidents reported for each tunnel in each year from 2006 to 2009. It is to be remembered that the number of accidents occurring in the same tunnel in different years are not stochastically independent, but that they share at least the effects of modifications and/or improvements introduced in that specific road tunnel over time. Thereby it is logic to consider that they form a cluster of observations sharing common, even if unknown, random-effects. Random-effects models analyse data by clusters, thus they may take into account the possible temporal correlation among observations relating to the same tunnel in different years. It is known that random-effects are the same within each panel of cluster, but they differ across clusters. However, the regression parameters are always the same within and across clusters; in other words the effect-random models do not consider that the regression parameters can vary across clusters. Therefore, other types of models have been introduced as the random-parameters models. In order to resolve the issue, we used the LIMDEP statistical package (Greene, 2007) in which a panel-data vector is contained for taking into account also the fact that the random-parameters can be fixed within clusters and vary between them.

6. Analysis of results

In order to analyse the effects of the aforementioned set of independent variables and of year covariate on total accident counts observed in unidirectional motorway tunnels, the following three predictive models were developed.

Random-intercept Poisson (RIP) model

The results of the random-intercept Poisson (RIP) model (which is equivalent to the random-effects Poisson (REP) model) are reported in Table 2. Table 2 confirms, coherently with the previous studies of the authors of this paper, that each variable has the sign expected: the frequency of total accidents is positively associated in a non-linear way with the tunnel length (*L*), the annual average daily traffic (*AADT*) per lane, and the percentage of trucks (*%Tr*). Possible explanations may be given. By increasing *L* higher crash frequency, attributable to the drivers' diminishing concentration, is expected. As *AADT* per lane and *%Tr* increase, the frequency of accidents is also expected to increase because in order to maintain the desired speeds, the lane changing and overtaking movements, more especially of cars, become more numerous.

A reduction in crash frequency over years from 2006 to 2009 is also confirmed, which agrees with accident data observed. This decrease might be attributable to the positive effects of more stringent road-safety policies over time and/or to the implementation of certain safety measures in tunnels after the date of the coming into force in Italy (October 2006) of the European Directive 2004/54/EC (European Parliament and Council, 2004).

All variables are statistically significant at a level of much less than 0.05, except: the presence of sidewalk (*SW*), the longitudinal slope (*LS*), and the mechanical ventilation (*MV*) (χ^2 with 1 dof < 3.84).

Table 2 Estimations results of the random-intercept Poisson (RIP) model (or random-effects Poisson (REP) model).

Variables	Random Intercept Poisson (RIP) Model		
	Point Estimate	Standard Error	LRT Statistic
Fixed Parameters			
Log of length [km]	1.28366	0.15622	28.84
Log of AADT per lane /10,000	2.85643	0.12685	159.15
Log of percentage of trucks/100	1.82361	0.34981	11.22
SW (Sidewalk: 1 if present; 0 otherwise)	-0.08198	0.07290	0.53
LS (Longitudinal slope: 1 if not constant; 0 otherwise)	0.11396	0.04823	2.13
MV (Mechanical ventilation: 1 if present; 0 otherwise)	0.09152	0.07692	0.60
Y2007	-0.21811	0.07852	14.29
Y2008	-0.53578	0.07039	72.09
Y2009	-0.61406	0.07363	80.52
Random Parameters			
Constant	2.50738	0.26053	-
standard deviation of parameter distribution: 0.36861			
standard error of standard deviation: 0.02365			
Other statistical information			
Number of observations		904	
Log likelihood function LL($\beta_{\text{RIP Model}}$)		-1454.20814	

Unrelated random-parameters Poisson (URPP) model

Table 3 shows the results of the unrelated random-parameters Poisson (URPP) model in which both intercept and parameters are assumed to vary randomly.

Table 3 Estimation results of uncorrelated random-parameters Poisson (URPP) model.

Variables	Uncorrelated Random-Parameters Poisson (URPP) Model		
	Point Estimate	Standard Error	LRT Statistic
Fixed Parameters			
Log of length [km]	1.33190	0.15999	31.97
Log of percentage of trucks/100	1.54582	0.35773	8.74
SW (Sidewalk: 1 if present; 0 otherwise)	-0.07537	0.07321	0.51
Y2007	-0.21902	0.07875	14.40
Random Parameters			
Constant	2.35492	0.26505	-
standard deviation of parameter distribution: 0.19244			
standard error of standard deviation: 0.02257			
Log of AADT per lane /10,000	2.87971	0.13293	158.63
standard deviation of parameter distribution: 0.58764			
standard error of standard deviation: 0.10062			
LS (Longitudinal slope: 1 if not constant; 0 otherwise)	0.06161	0.05066	7.80
standard deviation of parameter distribution: 0.35332			
standard error of standard deviation: 0.03462			
MV (Mechanical ventilation: 1 if present; 0 otherwise)	0.03655	0.08095	3.07
standard deviation of parameter distribution: 0.31135			
standard error of standard deviation: 0.04945			
Y2008	-0.54294	0.06974	72.75
standard deviation of parameter distribution: 0.08684			
standard error of standard deviation: 0.05154			
Y2009	-0.66468	0.07754	85.39
standard deviation of parameter distribution: 0.26375			
standard error of standard deviation: 0.05746			
Other statistical information			
Number of observations		904	
Log likelihood function LL($\beta_{\text{URP Model}}$)		-1449.20704	
Comparison between random-intercept (RIP) and uncorrelated random parameters models (URPP)			
χ^2= -2[LL($\beta_{\text{RIP Model}}$)-LL($\beta_{\text{URPP Model}}$)]		10.00	

One may note that each regression coefficient has the sign coherent with the one corresponding in the aforementioned RIP model; moreover, all variables are statistically significant, except: the sidewalk (*SW*), and the mechanical ventilation (*MV*). A reader may see that, in addition to the regression intercept assumed to be a priori random, only 5 parameters were found to be effectively random. These are: AADT, LS, MV, year 2008, and year 2009.

In particular, AADT per lane is a random parameter that is normally distributed with mean 2.88 and standard deviation 0.588, as a result 99.99% of its distribution is greater than 0. This indicates that all the investigated motorway tunnels present an increase in crash frequency as AADT per lane increases.

In a similar way, the longitudinal slope (LS) resulted to be a random parameter that is normally distributed with mean 0.062 and standard deviation 0.353, as a consequence 57.14% of its distribution is greater than 0 and 42.86% less than 0. This indicates that the majority of tunnels present an increase in crash frequency when the longitudinal slope is non-constant, but also that a significant minority results in a decrease.

The application of the LRT, between the two competing models (RIP and URPP), shows that the unrelated random-parameters model is not statistically superior to the corresponding random-intercept model. In fact, the χ^2 statistic with 5 degrees of freedom is 10.00 < 11.07. It is to be stressed that this finding might have been conditioned by the a priori assumption that there is no correlation among random-parameters.

Correlated random-parameters Poisson (CRPP) model

Table 4 shows that the estimation results of the correlated random-parameters Poisson (CRPP) model are consistent in signs with those of the corresponding unrelated random-parameters Poisson (URPP) model, but with differences in the magnitude of estimates.

Table 4 Estimation results of correlated random-parameters Poisson (CRPP) model

Variables	**Correlated Random-Parameters Poisson (CRPP) Model**		
	Point Estimate	Standard Error	LRT Statistic
Fixed Parameters			
Log of length [km]	1.22926	0.16540	28.83
Log of percentage of trucks/100	1.14033	0.37252	4.76
SW (Sidewalk: 1 if present; 0 otherwise)	-0.05781	0.07751	0.26
Y2007	-0.21966	0.08029	14.49
Random Parameters			
Constant	2.07605	0.27454	-
standard deviation of parameter distribution: 0.249881			
Log of AADT per lane /10,000	2.88809	0.14210	155.66
standard deviation of parameter distribution: 0.586767			
LS (Longitudinal slope: 1 if not constant; 0 otherwise)	0.03907	0.05340	10.03
standard deviation of parameter distribution: 0.301248			
MV (Mechanical ventilation: 1 if present; 0 otherwise)	0.01542	0.08817	8.54
standard deviation of parameter distribution: 0.530294			
Y2008	-0.62102	0.07782	95.20
standard deviation of parameter distribution: 0.374315			
Y2009	-0.75360	0.08806	111.07
standard deviation of parameter distribution: 0.444747			
Other statistical information			
Number of observations		904	
Log likelihood function LL($b_{\text{URPP Model}}$)		-1432.88366	
Comparison between uncorrelated (URPP) and correlated random-parameters (CRPP) models			
χ^2= -2[LL($\beta_{\text{URPP Model}}$)-LL($\beta_{\text{CRPP Model}}$)]		32.65	
Comparison between random-intercept (RIP) and correlated random-parameters (CRPP) models			
χ^2=-2[LL($\beta_{\text{RIP Model}}$)-LL($\beta_{\text{CRPP Model}}$)]		42.65	

In the CRPP model all variables are statistically significant including also the mechanical ventilation (*MV*), except the sidewalk (*SW*). The LRT, applied between the two competing models URPP and CRPP, shows that the correlated random-parameters model is statistically superior to the unrelated random-parameters one. In particular, the $\chi2$ statistic with 15 degrees of freedom is 32.65 > 24.99. This proves that there is a cross correlation among the random-intercept and the 5 random-parameters (i.e., AADT, LS, MV, year 2008, and year 2009) that cannot be ignored. In this respect, in our case, the effects of longitudinal slope (*LS*) and mechanical ventilation (*MV*) are needed to be commented more especially together with annual average daily traffic (AADT) per lane.

For completeness of information, one may also note that the CRPP model is better than the random-intercept Poisson (RIP) model since the χ^2 with 20 degrees of freedom is 42.65 > 31.41.

Table 5 shows the variance-covariance matrix for the random-parameters (also including the intercept). Table 6 reports, instead, the correlation coefficient matrix of random parameters. Table 6, in particular, shows: (i) non-constant longitudinal slope (*LS*) is negatively associated with the annual average daily traffic (*AADT*) per lane, this indicates that under non-constant LS conditions the effect of the AADT on increasing crash frequency could be alleviated; (ii) also mechanical ventilation (MV) presents a negative sign with AADT per lane, which shows that the presence of the mechanical ventilation in tunnels makes less relevant the influence of AADT per lane on increasing crash frequency.

Table 5 Variance-covariance matrix of random parameters

Variables	Constant	Log of AADT per lane /10,000	LS	MV	Y2008	Y2009
Constant	0.0624					
Log of AADT per lane /10,000	0.1255	0.3443				
LS (Longitudinal slope: 1 if not constant; 0 otherwise)	0.0181	-0.0480	0.0907			
MV (Mechanical ventilation: 1 if present; 0 otherwise)	-0.1250	-0.2526	-0.0293	0.2812		
Y2008	-0.0519	-0.0264	-0.0851	0.1322	0.1401	
Y2009	0.0561	0.0233	0.0960	-0.1522	-0.1656	0.1978

Table 6 Correlation coefficient matrix of random parameters

Variables	Constant	Log of AADT per lane /10,000	LS	MV	Y2008	Y2009
Constant	1.00000	0.85601	0.23987	-0.94369	-0.55470	0.50468
Log of AADT per lane /10,000	0.85601	1.00000	-0.27167	-0.81165	-0.12012	0.08915
LS (Longitudinal slope: 1 if not constant; 0 otherwise)	0.23987	-0.27167	1.00000	-0.18365	-0.75502	0.71674
MV (Mechanical ventilation: 1 if present; 0 otherwise)	-0.94369	-0.81165	-0.18365	1.00000	0.66594	-0.64527
Y2008	-0.55470	-0.12012	-0.75502	0.66594	1.00000	-0.99492
Y2009	0.50468	0.08915	0.71674	-0.64527	-0.99492	1.00000

One possible explanation of these results might be as follows. When the longitudinal slope (*LS*) of tunnel is non-constant, changes in the speed of trucks along the tunnel length are expected. For example, on ascending gradients trucks slow down, while on descending gradients they travel faster. However, since grades of the tunnels investigated are not more than ± 4%, the effect of non-constant LS on speed of trucks might be limited, and as a result the average speed of trucks might be approximately the same as in flat tunnels. If the longitudinal slope is, instead, constant for example with a positive gradient (upgrade), the speed of trucks continues to decrease along the tunnel length for the grade resistance force that obstacles the truck motion. Negative constant grade also has a significant impact on the speed of trucks, in fact these vehicles in general slow down considerably on downgrades in order to avoid the brakes failure. In the light of the above considerations, the non-constant longitudinal slope (*LS*) of a tunnel has a less effect on the reduction of trucks speed than upgrade or downgrade. As a result, smaller speed differences between trucks and passenger cars might be expected. This means that for maintaining their desirable speed, cars do not have to change lane and overtake trucks more frequently with an increase in traffic flow; as a consequence, for non-constant LS conditions of tunnels, the influence of the AADT per lane on increasing crash frequency could be mitigated.

Moreover, when the AADT per lane increases, higher level of pollution due to traffic flow can be expected in tunnels. But if the mechanical ventilation (*MV*) is present, it provides a minimum level of ventilation in order to ensure adequate air quality. This might reduce the negative effects on driving behaviour due, for example, to respiratory defects, eye irritation, as well as to a reduction in visibility in discerning the presence of other vehicles; and as a result a minor influence of AADT per lane on increasing accidents could be expected.

On the other hand, we note that the continuous variables enter the model in logarithmic form. Therefore, the estimated coefficient can also be interpreted as the percentage increase in the expected number of accidents corresponding to 1% increase in the variable. For example, from Table 4 one can see that a 1% increase in the AADT per lane produces a 2.888% increase in the expected number of accidents.

7. Comments for practical applications and discussion for future investigations.

The results of this paper can have potential applications in the design of tunnels. With reference more especially to geometry, road engineers should avoid designing tunnels which are too long. An emergency lane should also be introduced for reducing the probability of accidents. This would also provide an escape lane for vehicles that have broken down and/or emergency services in the case of more relevant events such as accidents followed by fire and/or single burning vehicles caused by electrical-mechanical faults.

Generally speaking, the choice of the number of lanes is motivated by the need to relieve traffic congestion. But it is often believed that decreased congestion resulting from a greater number of lanes is also associated with improved safety. This is not always true because by adding another lane, for example passing from two to three lanes, if the traffic flow remains almost the same the opportunities of lane changing would prove to have increased; as a result more traffic conflicts and consequently more accidents might be expected. In this respect, a compromise in the choice of the more appropriate number of lanes would be reached by tunnel designers. With regard to a possible optimal solution, it is to be recorded that specified levels of service (LOS) must be, however, guaranteed in tunnels in relation to entity and composition of traffic flow.

It is also the opinion of the authors that the use of safety barriers having a form characterized by a re-directive profile, even if it has not been investigated in this study, should be encouraged in order to reduce the severity of vehicle collisions against the tunnel wall. In addition, the use of porous asphalt in pavement in the transition zones of tunnels (i.e., near the entrance and exit portals), where a higher crash frequency may be expected compared to that of the interior zones, might produce a reduction in accidents caused by adverse weather conditions (e.g., rain, snow, and ice). However, designers should pay attention because the porous asphalt is not advisable in the interior zones of tunnels. In fact, it might not guarantee the drainage in safety conditions of toxic and/or inflammable fluids that might be released by vehicles carrying dangerous goods (DGVs).

The predictive model developed can also be used by Tunnel Management Agencies (TMAs) for estimating more accurately variations in accident frequency caused by modifications in the annual average daily traffic (*AADT*) per lane, in relation to the longitudinal slope (*LS*) of tunnel and mechanical ventilation (*MV*), in order to decide more appropriate safety policies.

Although the writers of this paper are confident that they have carried out a suitable analysis, there still remain one point of interest that is worthy of discussion. Tunnel lighting is also a factor that can influence accidents. The quality of lighting, for instance, expressed in terms of density, uniformity, and change in colour, affects the behaviour of drivers within tunnels and consequently their response to traffic and geometry modifications. In our study, the lighting was present in all the tunnels investigated so that the influence of this factor has not been possible to investigate. Since this matter should be dealt with in greater depth, a possible address for future investigations might be an ad-hoc study based on a before-after analysis, in which to compare the count of accidents occurring in tunnels with and without lighting.

8. Summary and Conclusions

The article explores the use of a correlated random-parameters model to account for the cross correlation among random-parameters, in contrast with the corresponding unrelated random-parameters model and the fixed-parameters one where only the intercept is assumed to vary randomly.

The dependent variable is the frequency of total accidents (accidents involving material damage, injuries and fatalities), which occurred in 226 Italian unidirectional motorway tunnels over a four-year monitoring period.

The group of dependent variables is: tunnel length (*L*), annual average daily traffic (*AADT*) per lane,

percentage of trucks (*%Tr*), presence of sidewalk (*SW*), longitudinal slope (*LS*), and mechanical ventilation (*MV*).
The so-called random-intercept model (RIP), which was developed a priori for showing the random-effects (temporal correlations among accidents occurring in the same tunnel in different years) confirmed that each variable had the sign expected. The frequency of accidents was positively associated to: *L*, *AADT*, and *% Tr*. A reduction in crash frequency over the years was also confirmed, coherently with accident data observed.
All variables were found to be statistically significant except: the presence of sidewalk (*SW*), the longitudinal slope (*LS*), and the mechanical ventilation (*MV*).
The unrelated random-parameters model (URPP), in which both the intercept and parameters are assumed to vary randomly, showed results consistent with the aforementioned random-intercept model. All variables were found to be statistically significant, except: the sidewalk (*SW*), and the mechanical ventilation (*MV*). In addition to the regression intercept (assumed a priori to be random) the following 5 parameters were found to be effectively random: *AADT*, *LS*, *MV*, *year 2008*, and *year 2009*. However, the application of the Likelihood Ratio Test (LRT) showed that the unrelated random-parameters model was not statistically superior to the random-intercept model. This finding might have been conditioned by assuming that there is no correlation among random-parameters.
The correlated random-parameters model (CRPP) showed results consistent in signs with those of the corresponding unrelated random-parameters model, but with differences in the magnitude of estimates. All variables were found to be statistically significant except the sidewalk (*SW*). The LRT proved that the correlated random-parameters model was statistically superior to the corresponding uncorrelated random-parameter and intercept-random models. In other words, a cross correlation exists among the random-intercept and the 5 parameters that were found to be effectively random (*AADT*, *LS*, *MV*, *year 2008*, and *year 2009*). This finding cannot be ignored in statistical analysis. In this respect, in particular, it was found that the non-constant longitudinal slope (*LS*) of tunnels was negatively associated with the annual average daily traffic (*AADT*) per lane. This indicates that under non-constant LS conditions the effect of the AADT on increasing crash frequency could be mitigated. Also the mechanical ventilation (MV) was found for having a negative sign with AADT per lane. This means that the presence of the mechanical ventilation in tunnels makes less significant the influence of AADT per lane on increasing crash frequency.
The correlations found can provide additional insights for future applications to road engineers in the field of tunnel design. Moreover, the model proposed can also be used by Tunnel Management Agencies (TMAs) for the prediction of variations in accident frequency in a specific tunnel attributable to modifications in traffic control systems.
Despite the potential of the correlated random-parameters models, it is to be said that the present paper is focused more especially on total accidents. Therefore, a possible direction of expansion of this study could be an analysis based on random-parameters bivariate models in which also severe accidents occurring in tunnels are simultaneously considered. However, it is to be stressed that in this respect a different statistical method should be used. For example, the empirical and/or full Bayes approaches, which are not within the scope of this paper, appear to be more appropriate. In particular, the authors believe that future studies should be addressed towards the development of predictive models for the joint analysis of total and severe accidents occurring in tunnels by applying, for example, the Bayesian random-bivariate method.
Another possible extension of this study, which the lack of available information in the data-base has not permitted to be activated in the present paper, should be in making an in-depth investigation of accidents occurring more especially in the transition zones of tunnels, which in combination with different weather and lighting conditions, might occur with greater frequency if compared to those taking place inside tunnels. Also the effectiveness of lighting on the occurrence of accidents in tunnels should be investigated in greater depth, for example by means of an ad-hoc study based on a before-after analysis.
Therefore, further research is required for making further development possible.

References

Anastasopoulos, P.C., Mannering, F.L., 2009. A note on modelling vehicle accidents frequencies with random-parameters count models. Accident Analysis and Prevention, 4, pp. 153-159.

Anastasopoulos, P.C., Mannering, F.L., 2011. An empirical assessment of fixed and random parameter logit modes using crash- and non-crash specific injury data. Accident Analysis and Prevention, 43, pp.1140-1147.

Caliendo, C., De Guglielmo, M.L., 2012. Accident rates in road tunnels and social costs evaluation. SIIV – 5th International Congress-Sustainability of Road Infrastructures. Procedia - Social and Behavioral Sciences, Vol. 53, pp. 166-177.

Caliendo, C., De Guglielmo, M.L., Guida, M., 2013. A crash-prediction model for road tunnels. Accident Analysis and Prevention, Vol. 55, pp. 107-115.

Caliendo, C., De Guglielmo, M.L., Guida, M., 2016. Comparison and analysis of road tunnel traffic accident frequencies and rates using random-parameter model. Journal of Transportation Safety & Security (Taylor & Francis). Vol. 8, Issue.2, pp.177-195.

Caliendo, C., De Guglielmo, M.L., 2017. Accident analysis based on fixed and random-parameters models. International Journal of Civil Engineering and Technology. Vol 8, Issue 9, pp. 1150-1160.

Caliendo, C., Guida, M., 2014. A New Bivariate Regression Model for the Simultaneous Analysis of Total and Severe Crashes Occurrence. Journal of Transportation Safety & Security (Taylor & Francis). Vol.6, Issue 1, pp. 78-92.

Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accident Analysis and Prevention. Vol.39, No.4, pp. 657-670.

Conway, K., Kniesner, T. , 1991. The important econometric features of a linear regression model with cross-correlated random coefficients. Economics Letters. Let. 35, pp. 143.147.

Coruh, E., Bilgic, A., Tortum, A. Accident analysis with aggregated data: The random parameters negative binomial panel count data model. Analytic Methods in Accident Research. Vol. 7(2015), pp. 37-49.

European Commission (EU28). Road Safety 2016. https://ec.europa.eu/transport/road_safety/specialist/statistics_en

European Parliament and Council, 2004. Directive 2004/54/EC. Official Journal of the European Uninion. L. 167, Bruxelles, 30 April.

Greene, W. , 2007. LIMDEP Version 9.0. Plainview, NY: Econometric Software Inc.

Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik, 2, pp. 84-90.

Hauer, E., Bamfo, J. (1997).Two tools for finding what function links the dependent variable to the explanatory variables. In Proceedings of the International Cooperation on Theories and Concepts in Traffic Conference.Retrieved from: http://www.oocities.org/hauer@rogers.com/Pubs/ICTCT97TwoTools.pdf

Hilbe, E., 2007. Negative Binomial regression. New York, NY: Cambridge University Press. Retrieved from www.cambridge.org

Hou, Q., Tarko,A.P., Meng, Z., 2018. Analyzing crash frequency in freeway tunnels: A correlated random parameters approach. Accident Analysis and Prevention. Vol. 111, pp. 94-100.

Italian Ministry of Infrastructures and Transport (MIT), 2002. Functional and geometric standards for roads construction. G.U. n. 3, 4th of January, 2002.

Italian National Institute for Statistics, 2016. Road Accidents, year 2016. http://www.aci.it/fileadmin/documenti/studi_e_ricerche/dati_statistiche/incidenti/Sintesi_dello_studio_2016__English_version_.pdf

Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic Methods in Accident Research. Vol.1, pp. 1-12.

Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved and the statistical analysis of highway accident data. Analytic Methods in Accident Research. Vol.11, pp. 1-16.

Milton, J., Shankar, V., Mannering, F., 2008. Highway accident severities and mixed logit model: an exploratory empirical analysis. Accident Analysis and Prevention 40 (1), pp. 318-221.

Rongjie Y., Xiong, Y., Abdel-Aty, M., 2017.A correlated random parameter approach to investigate the effects of weather conditions on crash risk for a mountainous freeway. Transportation Research Part C. Vol. 50 (2017), pp 68-77.

Washington, S.P., Karlafitis, M.G., Mannering, F.L., 2011. Statistical and econometric methods of transportation data analysis. New York, NY: Chapman & Hall/CRC.

World Health Organization (WHO), 2015. Global Status Report on Road Safety 2015. Geneva, Switzerland.https://books.google.it/books?id=wV40DgAAQBAJ&printsec=frontcover&hl=it#v=onepage&q&f=false

Yeung, J.S., Wong, Y., D., 2013. Road traffic accidents in Singapore expressway tunnels. Tunnelling and Underground Space Technology. Vol. 38, pp 534-541.

Zhuang-lin, Ma, Chun-fu, S., Sheng-rui, Z., 2009. Characteristics of traffic accidents in Chinese freeways tunnels. Tunnelling and Underground Space Technology Vol. 24, pp. 350-355.