Modeling animal-vehicle collisions using diagonal inflated bivariate Poisson regression

Yunteng Lao, Yao-Jan Wu, Jonathan Corey, Yinhai Wang*

Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA

ABSTRACT

Two types of animal-vehicle collision (AVC) data are commonly adopted for AVC-related risk analysis research: reported AVC data and carcass removal data. One issue with these two data sets is that they were found to have significant discrepancies by previous studies. In order to model these two types of data together and provide a better understanding of highway AVCs, this study adopts a diagonal inflated bivariate Poisson regression method, an inflated version of bivariate Poisson regression model, to fit the reported AVC and carcass removal data sets collected in Washington State during 2002–2006. The diagonal inflated bivariate Poisson model not only can model paired data with correlation, but also handle under- or over-dispersed data sets as well.

Compared with three other types of models, double Poisson, bivariate Poisson, and zero-inflated double Poisson, the diagonal inflated bivariate Poisson model demonstrates its capability of fitting two data sets with remarkable overlapping portions resulting from the same stochastic process. Therefore, the diagonal inflated bivariate Poisson model provides researchers a new approach to investigating AVCs from a different perspective involving the three distribution parameters (λ_1, λ_2 and λ_3). The modeling results show the impacts of traffic elements, geometric design and geographic characteristics on the occurrences of both reported AVC and carcass removal data. It is found that the increase of some associated factors, such as speed limit, annual average daily traffic, and shoulder width, will increase the numbers of reported AVCs and carcass removals. Conversely, the presence of some geometric factors, such as rolling and mountainous terrain, will decrease the number of reported AVCs.

Published by Elsevier Ltd.

1. Introduction

According to a report by Huijser et al. (2007), animal-vehicle collisions (AVCs) have caused significant damages to human life, property, and wildlife in the past decades. About 200 people were killed and 20,000 people were injured annually by the AVCs in the U.S. AVC-related property damage exceeds one billion dollars each year. AVCs also have a remarkable impact on wildlife because, in most cases, animals die immediately or shortly after AVCs (Allen and McCullough, 1976). Previous studies (e.g., Van der Zee et al., 1992; Huijser and Bergers, 2000) also found AVCs affect the population levels of some animal species and may even cause their survivability crisis (Proctor, 2003). Thus, effective AVC countermeasures for existing transportation facilities are critical for both traffic safety and wildlife conservation.

To identify factors contributing to traffic accidents, researchers have tried various statistical modeling techniques (e.g., Hubbard et al., 2000; Knapp and Yi, 2004). These research efforts have resulted in several widely accepted models. The Poisson regression model, one of the most classical and basic methods, has been frequently used to model collision count data (e.g., Jovanis and Chang, 1986; Miaou et al., 1992; Miaou and Lum, 1993; Miaou, 1994). A well recognized problem with the Poisson regression model is that the sample variance needs to be equal to the sample mean. In reality, however, most accident data sets do not meet this requirement. Therefore, the Poisson model is inadequate for over-dispersed (or under-dispersed) data whose variance is greater (or smaller) than the mean (Maycock and Hall, 1984; Wang et al., 2003). To address over-dispersed accident data, several modeling approaches including negative binomial (NB) regression (or Poisson-gamma) (Miaou, 1994; Shankar et al., 1995; Poch and Mannering, 1996; Maher and Summersgill, 1996; Milton and Mannering, 1998; Chin and Quddus, 2003; Wang et al., 2003; Wang and Nihan, 2004; Lord, 2006; El-Basyouny and Sayed, 2006; Donnell and Mason, 2006; Kim et al., 2007; Malyshkina and Mannering, 2010; Daniels et al., 2010) and Poisson-lognormal regression models (Miaou et al., 2005; Lord and

* Corresponding author. Tel.: +1 206 616 2696; fax: +1 206 543 5965.
E-mail addresses: laoy@u.washington.edu (Y. Lao), yaojan@u.washington.edu (Y.-J. Wu), falker@u.washington.edu (J. Corey), yinhai@u.washington.edu (Y. Wang).

Miranda-Moreno, 2008; Aguero-Valverde and Jovanis, 2008) have been developed and widely applied.

Collision data can also occasionally be under-dispersed. Under-dispersion can be caused by data sets having a very low sample mean due to the many zeros in the data set (Oh et al., 2006). In this situation, a Gamma regression model (Winkelmann and Zimmermann, 1995; Oh et al., 2006) can be used to deal with this issue. Subsequently, the Conway–Maxwell–Poisson (COM–Poisson) based models is introduced for handling either over- or under-dispersed count data (Shmueli et al., 2005; Kadane et al., 2006; Lord et al., 2008).

Another issue in modeling collision data is the phenomena of an apparent excess of zeros. In this situation, zero-inflated models, such as zero-inflated Poisson and zero-inflated NB, have been used for modeling such collision data sets (Shankar et al., 1997; Garber and Wu, 2001; Lee and Mannering, 2002; Kumara and Chin, 2003; Miaou and Lord, 2003; Rodriguez et al., 2003; Shankar et al., 2003; Noland and Quddus, 2004; Qin et al., 2004; Lord et al., 2005a,b). However, Lord et al. (2005a,b) and Warton (2005) have argued that considerable caution should be exercised when applying zero-inflated models to crash data because a true two-state process may not exist. Recently, several innovative accident models, including random parameter models, finite-mixture/Markov switching models, neural networks, Bayesian neural networks, and support vector machines, have been used in accident analysis research. A comprehensive review of the accident models mentioned above can be found in the paper by Lord and Mannering (2010).

Most of the regression models described above are univariate Poisson- (or gamma-) based models designed for modeling general count problems. These univariate models are capable of estimating only one distribution parameter and would be limited in modeling multivariate issues. Thus, most previous AVC studies only adopted univariate models that separately considered the two types of AVC data: reported AVC data (Hubbard et al., 2000; Malo et al., 2004; Seiler, 2005) and carcass removal data (Reilley and Green, 1974; Allen and McCullough, 1976; Knapp and Yi, 2004).

According to a survey conducted among the Washington State DOT (WSDOT) carcass removal professionals, a carcass found on road is likely an animal hit by a vehicle (Lao et al., 2010). Thus, these two sets of data should overlap to a large extent. However, previous studies (Romin and Bissonette, 1996; Knapp et al., 2007; Huijser et al., 2007) found that these two datasets are significantly different. This implies that not all animals involved in AVCs died on the road, not all animal carcasses were removed and reported by transportation agencies, and/or not all AVCs were properly reported and recorded.

In order to model the two datasets simultaneously, a new approach that can consider the two data sets in one modeling process is needed. However, most count models do not allow modeling two datasets simultaneously except that a data merging process is implemented beforehand (Lao et al., 2010). Recently, multivariate Poisson regression models (Miaou and Song, 2005; Ma and Kockelman, 2006; Park and Lord, 2007), multivariate zero-inflated Poisson regression models (Li et al., 1999), or multivariate Poisson-lognormal regression models (Karim and Tarek, 2009) have been used for modeling different but correlated count data sets. As a special case of multivariate Poisson regression models, bivariate Poisson regression model can be used for paired count data sets. However, bivariate Poisson and other multivariate Poisson regression models cannot handle over- or under-dispersed count data. In order to concurrently utilize the reported AVC data and carcass removal data even when they are dispersed, a diagonal inflated bivariate Poisson regression model (Karlis and Ntzoufras, 2005) is developed and applied to AVC modeling in this research.

The remaining parts of this paper are organized as follows. The next section details the bivariate Poisson model, the diago-
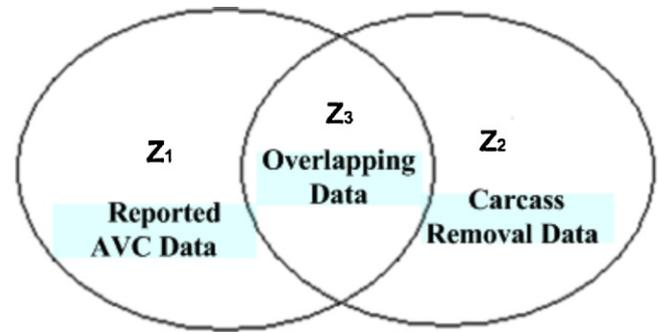

Fig. 1. Relationship between the reported AVC and CR data sets.

nal inflated bivariate Poisson model, and the estimation approach for these models. This is followed by the data description and the model estimation using 5-years (2002–2006) of AVC data collected from 10 study routes in Washington State. The diagonal inflated bivariate Poisson model is also compared with three other models, double Poisson, bivariate Poisson, and zero-inflated double Poisson. Then, the model results are interpreted and discussed. Finally, conclusions and recommendations are provided at the end of this paper.

2. Methodology

The bivariate Poisson model and its diagonal inflated version, diagonal inflated bivariate Poisson regression model, have been used in the analysis of health care and sports data (Karlis and Ntzoufras, 2005). The diagonal inflated bivariate Poisson regression model is chosen in this study for two reasons. First, both the bivariate Poisson and diagonal inflated bivariate Poisson models are appropriate for modeling paired count data with correlation. The two data sets for AVC analysis, reported AVC data and carcass removal data, are different but should be correlated. Second, the diagonal inflated bivariate Poisson regression model is capable of handling both over- and under-dispersed data (Karlis and Ntzoufras, 2005).

2.1. Bivariate Poisson regression model

Fig. 1 shows the relationships between two types of data, the reported AVC data (left circle) and the carcass removal data (right circle) typically collected by transportation agencies. There are three regions of interest: Z_1, Z_2, and Z_3. Z_1 represents the AVC reports with no corresponding carcass removal data. Z_2 represents the carcass removal data with no counterparts in the reported AVC data. The records contained in both the reported AVC data and the carcass removal data are represented by Z_3. This area is the overlapping portion of the two data sets.

Let us assume that the count data sets Z_1, Z_2, and Z_3 follow independent Poisson distributions with parameters (means) λ_1, λ_2, and λ_3, respectively. Then the reported AVC data set $X = Z_1 + Z_3$ and the carcass removal data set $Y = Z_2 + Z_3$ follow a bivariate Poisson distribution $(\lambda_1, \lambda_2, \lambda_3)$, with a joint probability mass function defined as (Karlis and Ntzoufras, 2005):

$$f_{BP}(x, y | \lambda_1, \lambda_2, \lambda_3)$$

$$= e^{-(\lambda_1+\lambda_2+\lambda_3)} \frac{\lambda_1^x}{x!} \frac{\lambda_2^y}{y!} \sum_{i=0}^{\min(x,y)} \binom{x}{i} \binom{y}{i} i! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right) \quad (1)$$

where x and y are the values that X and Y take on. The bivariate Poisson distribution is appropriate for modeling two random variables with positive dependence, which is the case for the reported AVC

and carcass removal data sets. Its marginal distributions of X and Y follow Poisson distributions with $E(X) = \lambda_1 + \lambda_3$ and $E(Y) = \lambda_2 + \lambda_3$, respectively. Moreover, $COV(X, Y) = \lambda_3$, and hence λ_3 is a measure of dependence between the reported AVC data set and the carcass removal data set.

In the bivariate Poisson model, λ_k with $k = 1, 2$, and 3 can be related to various explanatory variables using the classical exponential link functions. Therefore, the bivariate Poisson regression model can take the following form:

$$(X_i, Y_i) \sim BP(\lambda_{1i}, \lambda_{2i}, \lambda_{3i})$$
$$\ln(\lambda_{ki}) = \omega_{ki}^T \beta_k \qquad (2)$$

where $i = 1, \ldots, n$, is the roadway segment number, ω_{ki} is the vector of explanatory variables for roadway segment i, β_k is the corresponding coefficient vector for Z_k. In this study, the roadway segments are separated by consistent geometric factors. It should be noted that the double Poisson model is a special case of the bivariate Poisson model when $\lambda_3 = 0$.

2.2. Diagonal inflated bivariate Poisson regression model

A major disadvantage of the bivariate Poisson model is that its marginal distributions cannot handle over- or under-dispersed data since its marginal distributions are Poisson distributions that require the mean and the variance to be equal (Karlis and Ntzoufras, 2005). The diagonal inflated bivariate Poisson model proposed by Karlis and Ntzoufras (2005) can be used to fix this problem. This model uses a more general form developed on the basis of zero-inflated models and the probabilities of the diagonal elements are inflated in the probability table. The diagonal inflated bivariate Poisson model can be defined on the basis of the bivariate Poisson regression model as follows:

$$f_{IBP}(x, y) = \begin{cases} (1 - p_m)f_{BP}(x, y | \lambda_1, \lambda_2, \lambda_3), & x \neq y \\ (1 - p_m)f_{BP}(x, y | \lambda_1, \lambda_2, \lambda_3) + p_m f_D(x | \theta, J), & x = y \end{cases}$$
$$(3)$$

where p_m is the mixing Proportion. $f_D(x | \theta, J)$ is the probability mass function of a discrete distribution $D(x; \theta)$. $D(x; \theta)$ can be a Poisson, geometric, or a simple discrete distribution. That is, the data process has a probability of $1 - p_m$ to follow a bivariate Poisson distribution and a probability of p_m to follow $D(x; \theta)$. Note that the bivariate Poisson (when $p_m = 0$) and the zero-inflated double Poisson model (when $\lambda_3 = 0$ and $J = 0$) are special cases of diagonal inflated model (Karlis and Ntzoufras, 2005). $f_D(x | \theta, J)$ can be defined as

$$f_D(x | \theta, J) = \begin{cases} \theta_x & \text{for } x = 0, 1, \ldots, J \\ 0 & \text{for } x \neq 0, 1, \ldots, J \end{cases} \qquad (4)$$

where $\sum_{x=0}^{J} \theta_x = 1$. J is a parameter that controls the number in the diagonal cells in the cross-tabulation for the paired datasets considered by the model (cross-tabulation will be introduced in Section 3). If $J = 0$, only $x = y = 0$ contributes to the inflated part ($f_D(x | \theta, J)$), then the model in Eq. (3) become a zero-inflated model. If $J = 1$, both $x = y = 0$ and $x = y = 1$ contributes to the inflated part. In this case, cell $(0, 0)$ and cell $(1, 1)$ of the cross-tabulation are considered in the inflated part.

The marginal distributions of a diagonal inflated bivariate Poisson regression model are mixtures of distributions with one Poisson component. For example, the marginal distribution of X is:

$$f_{IBP}(x) = (1 - p)f_{P_0}(x | \lambda_1 + \lambda_3) + p f_D(x | \theta) \qquad (5)$$

where $f_{P_0}(x | \lambda)$ is the Poisson probability mass function with parameter $\lambda_1 + \lambda_3$. The marginal distributions of the diagonal inflated bivariate Poisson regression model can model either under- or over-dispersed count data, depending on the definition of $D(x;$

θ). For example, if $J = 1$, $\lambda_1 + \lambda_3 = 1$ and $p_m = 0.5$, the resulting distribution is under-dispersed. While $J = 0$ (the simplest case of zero-inflated models), the resulting distribution is over-dispersed. This implies that the diagonal inflated bivariate Poisson regression model is more flexible than the bivariate Poisson regression model and hence a clearly better choice for modeling the AVC data in this study.

The parameters in most multivariate Poisson or related models are difficult to estimate because of the computational issues involved in their applications (Karlis and Ntzoufras, 2005; Ma and Kockelman, 2006). However, recent developments in statistical software models and computer hardware have provided several ways to estimate bivariate Poisson models. In this study, an open source statistical analysis package, R (http://www.r-project.org/, 2009), was used to estimate the models. The expectation–maximization (EM) approach (Dempster et al., 1977; Borman, 2009) is used for estimating the parameters in the diagonal inflated bivariate Poisson regression model. Details of the EM algorithm can be found in Karlis (2003) and Karlis and Ntzoufras (2005).

To explore how changes in the variables affect the outcome Poisson distributions with parameters (means) λ_k, $k = 1, 2$, and 3, elasticity values are computed to determine the marginal effects of the independent variables in the diagonal inflated bivariate Poisson model. For the roadway segment i, λ_k is denoted as λ_{ki}, and its elasticity $E_{\omega_{ki,l}}^{\lambda_{ki}}$ is defined as (Washington et al., 2003)

$$E_{\omega_{ki,l}}^{\lambda_{ki}} = \frac{\partial \lambda_{ki}}{\partial \omega_{ki,l}} \frac{\omega_{ki,l}}{\lambda_{ki}} = \beta_{k,l} \omega_{ki,l} \qquad (6)$$

where $\omega_{ki,l}$ is the lth variable in the vector of explanatory variables for roadway segment i; $\beta_{k,l}$ is the corresponding coefficient of the lth variable for Z_k. The elasticity in Eq. (6) is applicable only when the explanatory variable $\omega_{ki,l}$ is continuous. For the indicator variables, a pseudo-elasticity is estimated as an approximate elasticity of the variables. The pseudo-elasticity is defined as (Washington et al., 2003)

$$E_{\omega_{ki,l}}^{\lambda_{ki}} = \frac{Exp(\beta_{k,l}) - 1}{Exp(\beta_{k,l})} \qquad (7)$$

3. Data description

Ten highways (US2, SR8, US12, SR20, I90, US97, US101, US395, SR525 and SR970) in Washington State were selected as study routes following the recommendation of WSDOT experts. Five years (2002–2006) of the reported AVC and the carcass removal datasets were analyzed for this research.

Table 1 shows all the explanatory variables used in the models. Annual average daily traffic (AADT) is converted into thousands of vehicles, and some variables, such as access control type, terrain type and rural or urban, are binary variables. Three types of animal habitats, while-tailed deer, mule deer and elk, are included in the variables since deer and elk are the main animal types most commonly involved in AVCs in Washington State.

The minimum, maximum, mean, and standard deviation (S.D.) are shown in the last four columns. One can find that both the reported AVC data and the carcass removal data are over-dispersed.

Table 2 is a cross-tabulation for the AVC data. Each cell represents the number of roadway segments that have corresponding numbers of AVC records in the reported AVC data and the carcass removal data in the 5-year study period. For example, 63 in the fifth row and the first column (the $(0, 4)$ cell) indicates there are 63 roadway segments with four records in the carcass removal data and zero record in the reported AVC data. From this table, one can find that most roadway segments have zero records in both data sets. That is the $(0, 0)$ cell has the largest number. It is reasonable in that

Table 1
Description of explanatory variables in the models.

	Variable	Minimum	Maximum	Mean	S.D.
X[a]	Number of reported AVCs per segment[c]	0	22	0.24	0.81
Y[b]	Number of carcasses per segment[c]	0	95	0.94	3.88
$z1$	Annual average daily traffic (in thousands)	0.31	148.8	13.85	19.76
$z2$	Restrictive access control (Yes: 1; No: 0)			0.24	
$z3$	Posted speed limit (mph)	20	70	52.76	10.79
$z4$	Truck percentage (%)	0	52.28	14.05	8.29
$z5$	Median width (feet)	0	60	7.9	15.62
$z6$	Total number of lanes for both directions	1[d]	9	2.79	1.24
$z7$	Roadway length (feet)	0.01	6.99	0.22	0.4
$z8$	Terrain type (rolling: 1; otherwise: 0)			0.720	
$z9$	Terrain type (mountainous: 1; otherwise: 0)			0.096	
$z10$	Lane width (feet)	10	20	12.5	1.88
$z11$	Left shoulder width (feet)	0	18	2.44	2.04
$z12$	Right shoulder width (feet)	0	20	4.03	3.52
$z13$	Rural or Urban (urban: 0; rural: 1)			0.758	
$z14$	White-tailed deer habitat (yes: 1; no: 0)			0.31	
$z15$	Mule deer habitat (yes: 1; no: 0)			0.51	
$z16$	Elk habitat (yes: 1; no: 0)			0.31	

[a] Reported AVC data record.
[b] Carcass removal data record.
[c] Dependent variable.
[d] Six out of 10,475 segments have only one lane.

most segments do not have AVCs observed during the study period. Among segments having at least one record in both the reported AVC and carcass removal data sets, the (1, 1) and (1, 2) cells contain the largest numbers of segments. Similarly, among the records with at least two records in each data set the (2, 2) cell contains the most records. Thus, the diagonal cells, cells (0, 0), (1, 1), and (2, 2) should be expected to play important roles in the data sets.

4. Model estimation

To compare different models, the model details and evaluation criterion for double Poisson, bivariate Poisson, diagonal inflated bivariate Poisson (DIBP) and zero-inflated double Poisson models are listed in Table 3. In order to compare the effect of different J values on the diagonal inflated bivariate Poisson model, three models – DIBP0, DIBP1 and DIBP2 – with different J values are also estimated. Table 3 shows the details about the variables used for λ_k, the value of J as well as the number of parameters in each model. The insignificant variables were removed from the variables lists. For example, four variables ($z5, z10, z14, z15$), insignificant for λ_1, were removed from all the six fitted models (noted as "-$z5$-$z10$-$z14$-$z15$" in Table 3).

One can see that the bivariate Poisson model has a better fit than the traditional double Poisson model. The diagonal inflated bivariate Poisson and zero-inflated double Poisson models gener-

ally have better fits because they take the zero-inflated portion into account. Overall, the DIBP1 model is considered the best-fitted model because it has the highest ρ^2, lowest AIC and lowest BIC. In comparison with DIBP1, DIBP2 does not show any improvement in its log-likelihood even when the J value becomes larger. That is because the DIBP2 model cannot benefit from the additional diagonal cell when the number of records in the (2, 2) cell of Table 2 is relatively low. Therefore, the selection of the J parameter should depend on the diagonal cell values in the AVC-carcass removal cross-tabulation as well as goodness-of-fit measures. The mixing proportions (p_m) in the last column indicate that the data in the diagonal of the AVC cross-tabulation should be over 66%. This result is also consistent with the statistical result in Table 2 where the sum of the diagonal value is about 79% of the total data.

Table 4 shows estimated values of θ and λ in the diagonal inflated bivariate Poisson models. θ values represent the proportion of the corresponding diagonal cells in the mixing proportion data; λ values denote the proportion of the three regions in Fig. 1. All models have $\theta_0 > 0.99$, indicating that more than 99% of mixing proportion data has zero AVC record and less than 1% of mixing proportion data has at least one AVC record for both data sets. This result is consistent with the statistics in Table 2 where both datasets have large numbers (more than 6698) of zero-accident roadway segments. Note that the value of λ_3 represents the average number of overlapped records per road segment. For the DIBP1 results,

Table 2
Cross-tabulation for AVC and CR data.

	Number of reported AVCs								Cumulated record
	0	1	2	3	4	5	6	>6	
Number of carcasses									
0	6698	361	77	21	10	3	2	2	7174
1	301	67	22	10	6	3	0	1	410
2	228	69	28	5	6	2	2	0	340
3	81	35	9	7	1	0	0	1	134
4	63	26	10	5	1	2	1	0	108
5	35	17	8	1	2	1	0	2	66
6	26	17	7	7	0	0	2	0	59
7	15	14	7	4	1	1	2	0	44
8	17	8	7	4	2	0	0	0	38
>8	81	64	43	31	22	13	10	16	280
Cumulated record	7545	678	218	95	51	25	19	22	8653

Table 3
Details for the six fitted models.

	Model details					Evaluation criterion				
	λ_1	λ_2	λ_3	J	Par.	LL	ρ^2	AIC	BIC	p_m
DP	-z5-z10-z14-z15	-z1-z5-z10-z15-z16	–	–	25	−21,802	0.313	43,654	43,852	–
BP	-z5-z10-z14-z15	-z1-z5-z10-z15-z16	-z5-z6-z10-z12-z15	–	37	−21,173	0.333	42,421	42,715	–
DIBP0	-z5-z10-z14-z15	-z1-z5-z10-z15-z16	-z5-z6-z10-z12-z15	0	38	−17,283	0.456	34,642	34,944	0.6612
DIBP1[a]	-z5-z10-z14-z15	-z1-z5-z10-z15-z16	-z5-z6-z10-z12-z15	1	39	−17,275	0.456	34,628	34,938	0.6637
DIBP2	-z5-z10-z14-z15	-z1-z5-z10-z15-z16	-z5-z6-z10-z12-z15	2	40	−17,275	0.456	34,630	34,948	0.6637
ZIDP	-z5-z10-z14-z15	-z1-z5-z10-z15-z16	–	–	27	−17,415	0.451	34,884	35,099	0.6659

[a] Best-fitted model; -z5-z10-z14-z15 indicates variables $z5$, $z10$, $z14$, and $z15$ are removed from the model; (–): the parameter is set zero; LL: log-likelihood; Par.: number of parameters; ρ^2: log-likelihood ratio index (Ben-Akiva and Lerman, 1985); AIC: Akaike's information criterion (Akaike, 1974); BIC: Bayesian information criterion (Schwarz, 1978; Liddle, 2007); p_m: mixing proportion; DP: double Poisson; BP: bivariate Poisson; DIBP: diagonal inflated bivariate Poisson; ZIDP: zero-inflated double Poisson.

the overlapping percentage in the reported AVC data is about 13% (0.0664/(0.0664 + 0.4605)).

Table 5 shows the coefficient, standard deviation, t-value, and average elasticity values for each explanatory variable for λ_1, λ_2, and λ_3. All the listed variables are statistically significant at a 5% significance level.

5. Model interpretation

Table 5 shows the DIBP1 model results, in which factors contributing to AVCs are identified. The positive values of the coefficients indicate that the increase of each of these explanatory variables increases the probability of AVC occurrences. Conversely, negative values of the corresponding coefficients show that the increases of these explanatory variables lower the probabilities of AVCs. In contrast to regular Poisson accident models, the diagonal inflated bivariate Poisson model contains three dependent variables, λ_1, λ_2 and λ_3. λ_1 and λ_2 quantify the effects on the reported AVC and the carcass removal portions, respectively, whereas λ_3 accounts for the combined effects on the overlapping carcass removal and the reported AVC data sets. The significance and interpretation of the explanatory variables for each dependent variable are discussed below.

Among the traffic elements, three variables are found to significantly contribute to the occurrence of AVCs. The estimated coefficients show that the variable of speed limit is the most significant variable affecting the occurrence of AVCs (λ_1: coef. = 0.043, t = 30.302, E_1 = 2.327; λ_2: coef. = 0.06, t = 33.129, E_2 = 3.247; λ_3: coef. = 0.068, t = 10.298, E_3 = 3.680). The elasticity values here show that a 1% increase in posted speed limit increases the λ_k by 2.327% for λ_1, 3.247% for λ_2, and 3.680% for λ_3. Higher speed limits tend to increase the likelihood of AVCs. This may be because drivers travel at higher speeds under a higher speed limit, and high-speed vehicles require longer stopping distances. Therefore, drivers may not be able to stop quickly enough to avoid colliding with an animal on the road. This finding is consistent with most AVC-related research that has concluded that speed limits have an increasing relationship with the AVC rates (Allen and McCullough, 1976; Rolley and Lehman, 1992).

AADT is found to have increasing effects on λ_1 (coef. = 0.013, t = 13.556, E_1 = 0.202) and λ_3 (coef. = 0.069, t = 14.827, E_3 = 1.072), but have no significant effect on λ_2. This may be because that AVCs

are more likely to be reported on a highway segment with heavier traffic since more travelers can observe and therefore call and report the AVCs. Meanwhile, once an AVC happens, the carcass could be removed by other agencies or persons other than WSDOT. This explains the reason why the AADT variable also contributes to AVC occurrences in the overlapping portion of the reported AVC and carcass removal data. Overall, a higher AADT increases the chance of AVCs because a higher volume elevates the level of accident exposure and shortens vehicle headways needed for animals to cross the road. This result is consistent with the accident research conducted by Chin and Quddus (2003).

A higher truck percentage is found to decrease the likelihood of reported AVCs and carcass removal for all λ's. One reason may be that drivers are more cautious when more trucks are on the road. Another reason may be that trucks are usually associated with louder noise which may scare animals away. Trucks also tend to have better driver visibility forward, which could provide more time for drivers to react. This result is similar to the motor vehicle accident research by Milton and Mannering (1998), which identified a decreasing relationship between truck percentage and accident probability.

Among the geometric design elements, five variables are significantly associated with the occurrence of AVCs. Roadway segments with restrictive access control tend to have lower accident risk with fairly significant t-ratios for λ_1 (t = −35.753), λ_2 (t = −15.973), and λ_3 (t = −14.988). Usually, highways with higher restrictive access control (e.g., interstates) also have more physical obstructions along the highway that may limit the crossing of animals. In this case, animals find it more difficult to access highways protected by physical obstructions, and consequently, the number of AVCs is smaller for the highways with more restrictive access control.

The variable, total number of lanes, is found to be significant at a 5% significance level for λ_1 (t = −9.761) and λ_2 (t = −22.882) but not λ_3. With an increase in the total number of lanes, the roadway becomes wider, increasing the crossing difficulty for animals. Thus, wider road segments may be less attractive for animals to cross and hence reducing the likelihood of AVCs.

As expected, longer roadway segment length appears to increase the occurrence of AVCs (λ_1: coef. = 0.499, t = 58.069, E_1 = 0.105; λ_2: coef. = 0.471, t = 17.042, E_2 = 0.099; λ_3: coef. = 0.912, t = 30.785, E_3 = 0.192). This may be because the longer the roadway segment is, the more likely it is to segment animal habitats, between which

Table 4
Estimated values of θ and λ in DIBP models.

Models	θ estimation			Mean of parameter λ		
	θ_0	θ_1	θ_2	λ_1	λ_2	λ_3
DIBP0	1	–	–	0.4608	1.9205	0.0659
DIBP1[a]	0.9976	0.0023	–	0.4605	1.9359	0.0664
DIBP2	0.9976	0.0023	0.0000	0.4606	1.9359	0.0664

[a] Best-fitted model; DIBP: diagonal inflated bivariate Poisson.

Table 5
The DIBP1 model for AVC.

Explanatory variables	λ_1				λ_2				λ_3			
	Coef.	St. err	t-Value	E_1	Coef.	St. err	t-Value	E_2	Coef.	St. err	t-Value	E_3
Constant	−2.904	0.093	−31.067	–	−3.164	0.101	−31.369	–	−26.763	2.769	−9.665	–
Annual average daily traffic (in thousands)	0.013	0.001	13.556	0.202	–	–	–	–	0.069	0.005	14.827	1.072
Restrictive access control (yes: 1; no: 0)	−1.141	0.032	−35.753	−2.130	−0.986	0.062	−15.973	−1.680	−2.036	0.136	−14.988	−6.660
Posted speed limit (mph)	0.043	0.001	30.302	2.327	0.060	0.002	33.129	3.247	0.068	0.007	10.298	3.680
Truck percentage (%)	−0.049	0.001	−39.589	−0.634	−0.011	0.003	−4.055	−0.142	−0.069	0.004	−16.417	−0.892
Total number of lanes	−0.198	0.020	−9.761	−0.592	−0.395	0.017	−22.882	−1.180	–	–	–	–
Roadway segment length (feet)	0.499	0.009	58.069	0.105	0.471	0.028	17.042	0.099	0.912	0.030	30.785	0.192
Terrain type (rolling: 1; otherwise: 0)	−0.302	0.029	−10.543	−0.353	0.105	0.044	2.417	0.100	−1.925	0.096	−20.152	−5.855
Terrain type (mountainous: 1; otherwise: 0)	−0.958	0.037	−25.646	−1.606	−0.182	0.066	−2.755	−0.200	−2.027	0.182	−11.159	−6.591
Left shoulder width (feet)	0.036	0.004	9.718	0.189	0.038	0.004	8.836	0.199	0.092	0.012	7.416	0.482
Right shoulder width (feet)	0.034	0.003	12.466	0.310	0.032	0.003	11.340	0.291	–	–	–	–
Rural or Urban (urban: 0; rural: 1)	0.560	0.046	12.114	0.424	0.780	0.049	15.790	0.591	19.984	0.232	86.172	15.140
White-tailed deer habitat (yes: 1; no: 0)	–	–	–	–	0.973	0.088	11.005	0.622	1.607	2.743	0.586	0.800
Elk habitat (yes: 1; no: 0)	0.203	0.018	11.162	0.184	–	–	–	–	1.417	0.078	18.102	0.758

E_1 average elasticity value for λ_1; E_2 average elasticity value for λ_2; E_3 average elasticity value for λ_3.

animals will move. Similarly, more vehicle-miles are traveled on longer segments for the same traffic, number of lanes, etc. For the same per vehicle-mile traveled or per segment mile of length risk of collision a longer segment increases the total AVCs.

Both left and right shoulders are found to have an increasing effect on λ_1 ($t = 9.718$ for the left, $t = 12.466$ for the right) and λ_2 ($t = 8.836$ for the left, $t = 11.340$ for the right). Generally, drivers may have a broader view on roadways with shoulders. However, the results indicate that shoulders do not give drivers enough time to react to the appearance of animals because drivers tend to drive faster on segments with shoulders.

In terms of area types, three variables have significant impacts on the occurrences of AVCs: rolling area, mountainous area, and rural areas. In comparison with level terrain, rolling areas are associated with low numbers of reported AVCs (λ_1: coef. = −0.302, $t = −10.543$, $E_1 = −0.353$) and AVCs in the overlapping portion of the reported AVC and carcass removal data sets (λ_3: coef. = −1.925, $t = −20.152$, $E_3 = −5.855$). However, rolling areas are found to be associated with a higher number of carcasses (λ_2: coef. = 0.105, $t = 2.417$, $E_2 = 0.100$) than level terrain areas. The contradiction in the estimated coefficient values may imply that AVCs occurred in rolling areas are under reported compared with those occurred in level terrain areas.

In comparison with level terrain, mountainous areas tend to have a low likelihood of AVCs (λ_1: coef. = −0.958, $t = −25.646$, $E_1 = −1.606$; λ_2: coef. = −0.182, $t = −2.755$, $E_2 = −0.200$; λ_3: coef. = −2.027 $t = −11.159$). This may be because in mountainous areas, people drive more carefully, and vehicles are also slower. Another possible reason is that carcasses may not be easily found or require removal when they come to rest in areas off of roadways. Similarly, roadways in mountainous terrain tend to be in valleys and tunnels which may limit the coverage of cell phone for reporting. Some might argue that animals should be also active in the mountainous areas, resulting in more AVCs. However, the valleys and tunnels may also impede animals' movements because these geometric characteristics may physically separate different habitats. Therefore, animal crossing activities could be reduced.

Compared to highways in urban areas, those in rural areas are found to have more reported AVC and carcass removal records in both data sets (λ_1: coef. = 0.560, $t = 12.114$, $E_1 = 0.424$; λ_2: coef. = 0.780, $t = 15.790$, $E_2 = 0.591$; λ_3: coef. = 19.984, $t = 86.172$, $E_3 = 15.140$). This is under our expectation because animals are more active and populated in rural areas. However, looking at the overlapping portion of two data sets, this "rural effect" is more obvious (λ_3: coef. = 19.984, $t = 86.172$, $E_3 = 15.140$). This result highlights rural AVCs as a potential focus for future AVC research.

In terms of high density animal distribution areas, white-tailed deer habitat is associated with a higher λ_1 ($t = 11.005$), as expected. Elk habitat is also found to have an increasing impact on λ_1 ($t = 11.162$) and λ_3 ($t = 18.102$). It makes sense that the areas with higher density animal distribution tend to have a higher AVC rate. However, mule deer habitat is not found to significantly affect the likelihood of AVCs. One main reason may be that the mule deer population distribution in Washington State is relatively uniformly and widely distributed and covers a large portion of the study routes.

In summary, speed limit, restrictive access control, and roadway segment length are the most significant explanatory variables affecting all λ's (the absolute values of their t-ratios are over 10). According to the average elasticity values, rural area, restrictive access control, and terrain type (rolling or mountainous) have the most significant marginal effects on λ_3 (the absolute values of their average elasticity values are all over 5%). It should be noted that the posted speed limit is the only variable with the absolute values of all the average elasticity values being over 2% for all λ's. Hence, reducing the posted speed limit could result in a reduction of AVCs effectively.

Based on the analysis above, the authors suggest that in areas where the highway segments the habitats of non-domestic animals, especially deer, transportation agencies should further examine speed limit and access control options to develop suitable countermeasures. Constructing fences and crossing infrastructure (e.g., tunnels and over bridges) along and within the hot spots could be helpful for connecting segmented animal habitats and preventing animals from interacting with vehicles in the areas with frequent AVCs (Donaldson, 2007).

6. Conclusion

Animal-vehicle collision (AVC) is an important roadway-safety concern in many areas around the world. In order to investigate the contributing factors of AVCs, reported AVC and carcass removal data sets are commonly used in previous studies but these two significantly different data sets are usually analyzed separately. Although the two data sets complement each other, they have not been analyzed jointly. This paper applies a diagonal inflated bivariate Poisson (DIBP) regression model to fit these data sets concurrently. As an inflated version of the bivariate Poisson regression model, the diagonal inflated bivariate Poisson model outperformed other models because among the models (double Poisson, bivariate Poisson, diagonal inflated bivariate Poisson, and zero-inflated double Poisson) studied in this paper, the diagonal inflated bivariate Poisson models are the best-fitted models with the lowest AIC and BIC values. Functionally, the diagonal inflated bivariate Poisson model not only can handle under- or over-dispersed count data but also can model paired data sets with correlation.

The contributing factors of AVCs are identified by the DIBP1 model. Three dependent variables (λ_1, λ_2 and λ_3) are each linked with a group of explanatory variables including traffic elements, geometric design factors, and geographic characteristics associated with AVCs. Two traffic elements, speed limit and AADT, and two geometric design factors, shoulder width and roadway segment length, are found to have an increasing effect on the likelihood of AVCs. In terms of the variables of geographic characteristics, rural area segments tend to have higher numbers of reported AVCs and carcass removals. The areas with dense animal distributions, such as white-tailed deer and elk habitats, are also found to increase the occurrence probability of AVC.

In this study, the diagonal inflated bivariate Poisson model has been found effective in modeling AVCs. The methodology developed in this study may be applied to model other types of accident with two datasets of similar characteristics. Since the datasets used in this study happen to be over-dispersed, diagonal inflated bivariate Poisson's capability for handling under-dispersed data may be demonstrated in future studies. Moreover, comparisons between the diagonal inflated bivariate Poisson model and other multivariate models, such as bivariate negative binomial and bivariate Poisson-lognormal models, will be desired extensions of this study.

Although the diagonal inflated bivariate Poisson model is effective in predicting and assessing contributing factors of AVCs using concurrently the reported AVC and carcass removal data sets collected from the ten study routes in Washington State, more data are needed to further investigate the approach and accident causation. Transferability testing is also needed when applying this model to different animal types or locations. Moreover, it will be more desirable to investigate the potential contribution of time factors (e.g., day vs. night) in the future research since AVCs associated with a specific animal type may be more frequent in certain periods of a day.

References

Aguero-Valverde, J., Jovanis, P.P., 2008. Analysis of road crash frequency with spatial models. Transp. Res. Rec. 2061, 55–63.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. 19 (6), 716–723.

Allen, R.E., McCullough, D.R., 1976. Deer-car accidents in Southern Michigan. J. Wildl. Manage. 40, 317–325.

Ben-Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis: theory and application to travel demand. MIT Press Series in Transportation Studies, vol. 9. MIT Press, Cambridge, Mass, pp. 167–168.

Borman, S., 2009. The Expectation Maximization Algorithm – A Short Tutorial, http://www.seanborman.com/publications/EM_algorithm.pdf.

Chin, H.C., Quddus, M.A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. Accid. Anal. Prev. 35, 253–259.

Daniels, S., Brijs, T., Nuyts, E., Wets, G., 2010. Explaining variation in safety performance of roundabouts. Accid. Anal. Prev. 42 (2), 292–402.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B: Stat. Methodol. 39, 1–38.

Donaldson, B.M., 2007. Use of highway underpasses by large mammals and other wildlife in Virginia: factors influencing their effectiveness. Transp. Res. Rec. 2011, 157–164.

Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. Accid. Anal. Prev. 38 (3), 590–599.

El-Basyouny, K., Sayed, T., 2006. Comparison of two negative binomial regression techniques in developing accident prediction models. Transp. Res. Rec. 1950, 9–16.

Garber, N.J., Wu, L., 2001. Stochastic Models Relating Crash Probabilities with Geometric and Corresponding Traffic Characteristics Data (UVACTS-5-15-74). Center for Transportation Studies, University of Virginia, Charlottesville, VA.

HSIS, 2009. http://www.hsisinfo.org/ (accessed 15.06.09).

http://www.r-project.org/, 2009 (accessed 05.09.09).

Huijser, M.P., Bergers, P.J.M., 2000. The effect of roads and traffic on Hedgehog (Erinaceus europaeus) populations. Biol. Conserv. 95, 111–116.

Huijser, M.P., Fuller J., Wagner, M.E., Hardy, A., Clevenger, A.P., 2007. Animal-vehicle Collision Data Collection: A Synthesis of Highway Practice. National Cooperative Highway Research Board Program: Synthesis 370. Transportation Research Board, Washington, DC.

Hubbard, M.W.B.J., Danielson, Schmitz, R.A., 2000. Factors influencing the location of deer-vehicle accidents in Iowa. J. Wildl. Manage. 64, 707–712.

Jovanis, P.P., Chang, H., 1986. Modeling the relationship of accidents to miles traveled. Transp. Res. Rec. 1068, 42–51.

Kadane, J.B., Shmueli, G., Minka, T.P., Borle, S., Boatwright, P., 2006. Conjugate analysis of the Conway–Maxwell–Poisson distribution. Bayesian Anal. 1, 363–374.

Karim, E., Tarek, S., 2009. Collision prediction models using multivariate Poisson-lognormal regression. Accid. Anal. Prev. 41 (4), 820–828.

Karlis, D., 2003. An EM algorithm for multivariate Poisson distribution and related models. J. Appl. Stat. 30 (1), 63–77.

Karlis, D., Ntzoufras, I., 2005. Bivariate Poisson and diagonal inflated Poisson regression models in R. J. Stat. Softw. 14 (10).

Kim, J., Wang, Y., Ulfarsson, G., 2007. Modeling the probability of freeway rear-end crash occurrence. ASCE J. Transp. Eng. 133 (1), 11–19.

Knapp, K.K., Lyon, C., Witte, A., Kienert, C., 2007. Crash or carcass data: a critical definition and evaluation choice. Transp. Res. Rec. 2019, 189–196.

Knapp, K.K., Yi, X., 2004. Deer-Vehicle Crash Patterns and Proposed Warning Sign Installation Guidelines. In: Proceedings of the TRB 2004 Annual Meeting, TRB 04-000571.

Kumara, S.S.P., Chin, H.C., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. Traffic Inj. Prev. 3 (4), 53–57.

Lao, Y., Wu, Y., Wang, Y., McAllister, K., 2010. Fuzzy logic-based mapping algorithm for improving animal-vehicle collision data. Transportation Research Record, Submitted.

Lee, J., Mannering, F., 2002. Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accid. Anal. Prev. 34 (2), 149–161.

Li, C.C., Lu, J.C., Park, J., Kim, K., Brinkley, P.A., Peterson, J.P., 1999. Multivariate zeroin-flated Poisson models and their applications. Technometrics 41 (1), 29–38.

Liddle, A.R., 2007. Information criteria for astrophysical model selection. Mon. Not. R. Astron. Soc.: Lett. 377 (1), L74–L78.

Lord, D., Washington, S.P., Ivan, J.N., 2005a. Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory. Accid. Anal. Prev. 37 (1), 35–46.

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prev. 38 (4), 751–766.

Lord, D., Guikema, S.D., Geedipally, S.R., 2008. Application of the Conway–Maxwell–Poisson generalized linear model for analyzing motor vehicle crashes. Accid. Anal. Prev. 40 (3), 1123–1134.

Lord, D., Manar, A., Vizioli, A., 2005b. Modeling crash-flow-density and crashflow – V/C ratio for rural and urban freeway segments. Accid. Anal. Prev. 37 (1), 185–199.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transp. Res. A 44 (5), 291–305.

Lord, D., Miranda-Moreno, L.F., 2008. Effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter of Poisson-gamma models for modeling motor vehicle crashes: a Bayesian perspective. Safety Sci. 46 (5), 751–770.

Ma, J., Kockelman, K., 2006. Bayesian multivariate Poisson regression for models of injury count, by severity. Transp. Res. Rec. 1950, 24–34.

Maher, M.J., Summersgill, I.A., 1996. Comprehensive methodology for the fitting of predictive accident models. Accid. Anal. Prev. 28 (3), 281–296.

Malo, J.E., Suarez, F., Diez, A., 2004. Can we mitigate animal-vehicle accidents using predictive models? J. Appl. Ecol. 41, 701–710.

Malyshkina, N., Mannering, F., 2010. Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accid. Anal. Prev. 42 (1), 131–139.

Maycock, G., Hall, R.D., 1984. Accidents at 4-arm Roundabouts. Laboratory Report LR1120. Transport Research Laboratory, Crowthorne, Berks, U.K.

Miaou, S.P., 1994. The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 (4), 471–482.

Miaou, S., Hu, P.S., Wright, T., Rathi, A.K., Davis, S.C., 1992. Relationship between truck accidents and highway geometric design: a Poisson regression approach. Transp. Res. Rec. 1376, 10–18.

Miaou, S.P., Lord, D., 2003. Modeling traffic crash-flow relationships for intersections: dispersion parameter, functional form, and Bayes versus empirical Bayes. Transp. Res. Rec. 1840, 31–40.

Miaou, S.P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 (6), 689–709.

Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence. Accid. Anal. Prev. 37 (4), 699–720.

Miaou, S.-P., Bligh, R.P., Lord, D., 2005. Developing median barrier installation guidelines: a benefit/cost analysis using Texas data. Transp. Res. Rec. 1904, 3–19.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. Transportation 25 (4), 395–413.

Noland, R.B., Quddus, M.A., 2004. A spatially disaggregate analysis of road casualties in England. Accid. Anal. Prev. 36 (6), 973–984.

Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway highway interfaces. Accid. Anal. Prev. 38 (2), 346–356.

Park, E.S., Lord, D., 2007. Multivariate Poisson-lognormal models for jointly modeling crash frequency by severity. Transp. Res. Rec. 2019, 1–6.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection accident frequencies. J. Transp. Eng. 122 (2), 105–113.

Proctor, M.F., 2003. Genetic Analysis of Movement, Dispersal, and Population Fragmentation of Grizzly Bears in Southwestern Canada, Ph.D. dissertation, The University of Calgary, Calgary, AB, Canada.

Qin, X., Ivan, J.N., Ravishankar, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid. Anal. Prev. 36 (2), 183–191.

Reilley, R.E., Green, H.E., 1974. Deer MORTALITY on a Michigan interstate highway. J. Wildl. Manage. 38, 16–19.

Rodriguez, D.A., Rocha, M., Khattak, A.J., Belzer, M.H., 2003. Effects of truck driver wages and working conditions on highway safety: case study. Transp. Res. Rec. 1833, 95–102.

Rolley, R.E., Lehman, L.E., 1992. Relationships among raccoon road – kill surveys, harvests, and traffic. Wildl. Soc. Bull. 20, 313–318.

Romin, L.A., Bissonette, J.A., 1996. Deer-vehicle collisions: status of state monitoring activities and mitigation efforts. Wildl. Soc. Bull. 24 (2), 276–283.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.

Seiler, A., 2005. Predicting locations of moose-vehicle collisions in Sweden. J. Appl. Ecol. 42, 371–382.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. Accid. Anal. Prev. 27 (3), 371–389.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accid. Anal. Prev. 29 (6), 829–837.

Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergall, M.B., 2003. Modeling crashes involving pedestrians and motorized traffic. Safety Sci. 41 (7), 557–640.

Shmueli, G., Minka, T.P., Kadane, J.B., Borle, S., Boatwright, P., 2005. A useful distribution for fitting discrete data: revival of the Conway–Maxwell–Poisson distribution. J. R. Stat. Soc. Ser. C: Appl. Stat. 54, 127–142.

Wang, Y., Nihan, N.L., 2004. Estimating the risk of collisions between bicycles and automobiles at signalized intersections. Accid. Anal. Prev. 36 (3), 313–321.

Wang, Y., Ieda, H., Mannering, F.L., 2003. Estimating rear-end accident probabilities at signalized intersections: an occurrence-mechanism approach. J. Transp. Eng. 129 (4), 1–8.

Warton, D.I., 2005. Many zeros does not mean zero inflation: comparing the goodness-of-fit of parametric models to multivariate abundance data. Environmetrics 16 (2), 275–289.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical and Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC, Boca Raton, Florida, pp. 242–243.

Winkelmann, R., Zimmermann, K., 1995. Recent developments in count data modeling: theory and applications. J. Econ. Surveys 9, 1–24.

Van der Zee, F.F., Wiertz, J., ter Braak, C.J.F., van Apeldoorn, R.C., Vink, J., 1992. Landscape change as a possible cause of the Badger Meles meles L. Decline in The Netherlands. Biol. Conserv. 61, 17–22.