ELSEVIER

Accident Analysis and Prevention 38 (2006) 751–766

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

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☆

Dominique Lord*

Zachry Department of Civil Engineering, TAMU 3136, Texas A&M University, College Station, TX 77843-3136, United States

Received 3 November 2005; received in revised form 30 January 2006; accepted 1 February 2006

Abstract

There has been considerable research conducted on the development of statistical models for predicting crashes on highway facilities. Despite numerous advancements made for improving the estimation tools of statistical models, the most common probabilistic structure used for modeling motor vehicle crashes remains the traditional Poisson and Poisson-gamma (or Negative Binomial) distribution; when crash data exhibit over-dispersion, the Poisson-gamma model is usually the model of choice most favored by transportation safety modelers. Crash data collected for safety studies often have the unusual attributes of being characterized by low sample mean values. Studies have shown that the goodness-of-fit of statistical models produced from such datasets can be significantly affected. This issue has been defined as the "low mean problem" (LMP). Despite recent developments on methods to circumvent the LMP and test the goodness-of-fit of models developed using such datasets, no work has so far examined how the LMP affects the fixed dispersion parameter of Poisson-gamma models used for modeling motor vehicle crashes. The dispersion parameter plays an important role in many types of safety studies and should, therefore, be reliably estimated.

The primary objective of this research project was to verify whether the LMP affects the estimation of the dispersion parameter and, if it is, to determine the magnitude of the problem. The secondary objective consisted of determining the effects of an unreliably estimated dispersion parameter on common analyses performed in highway safety studies. To accomplish the objectives of the study, a series of Poisson-gamma distributions were simulated using different values describing the mean, the dispersion parameter, and the sample size. Three estimators commonly used by transportation safety modelers for estimating the dispersion parameter of Poisson-gamma models were evaluated: the method of moments, the weighted regression, and the maximum likelihood method. In an attempt to complement the outcome of the simulation study, Poisson-gamma models were fitted to crash data collected in Toronto, Ont. characterized by a low sample mean and small sample size. The study shows that a low sample mean combined with a small sample size can seriously affect the estimation of the dispersion parameter, no matter which estimator is used within the estimation process. The probability the dispersion parameter becomes unreliably estimated increases significantly as the sample mean and sample size decrease. Consequently, the results show that an unreliably estimated dispersion parameter can significantly undermine empirical Bayes (EB) estimates as well as the estimation of confidence intervals for the gamma mean and predicted response. The paper ends with recommendations about minimizing the likelihood of producing Poisson-gamma models with an unreliable dispersion parameter for modeling motor vehicle crashes.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Statistical models; Poisson-gamma; Low sample mean values; Empirical Bayes; Small sample size

1. Introduction

There has been considerable research conducted on the development of statistical models for predicting crashes on highway facilities (Abbess et al., 1981; Hauer et al., 1988; Persaud and Dzbik, 1993; Kulmala, 1995; Poch and Mannering, 1996; Lord, 2000; Ivan et al., 2000; Lyon et al., 2003; Miaou and Lord, 2003; Oh et al., 2003; Lord et al., 2005a; Miaou and Song, 2005). Despite numerous developments for improving the estimation tools of statistical models, such as random-effects (RE) models (Miaou and Lord, 2003), the Generalized Estimating Equations (GEE) (Lord and Persaud, 2000; Abdel-Aty and Addella, 2004) or the full-Bayesian methods (Qin et al., 2004; Miaou and Lord, 2003; Miaou and Song, 2005), the most common probabilistic

☆ This paper was presented at the 85th Annual Meeting of the Transportation Research Board.

* Tel.: +1 979 458 3949; fax: +1 979 845 6481.
E-mail address: d-lord@tamu.edu.

0001-4575/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2006.02.001

structure of the models used for modeling motor vehicle crashes remains the traditional Poisson and Poisson-gamma (or Negative Binomial) distribution.

Crash data have been shown to exhibit over-dispersion, meaning that the variance is greater than the mean. The over-dispersion can be caused by various factors, such as data clustering, unaccounted temporal correlation, model misspecification, but it has been shown to be mainly attributed to the actual nature of the crash process, namely the fact that crash data are the product of Bernoulli trials with unequal probability of events (this is also known as Poisson trials). Lord et al. (2005b) have reported that as the number of trials increases and becomes very large, the distribution may be approximated by a Poisson process, where the magnitude of the over-dispersion is dependent on the characteristics of the Poisson trials. (Note: the over-dispersion can be minimized using appropriate mean structures of statistical models, as discussed in Miaou and Song, 2005.)

Although different Poisson-based distributions have been developed to accommodate the over-dispersion (e.g., Poisson-lognormal, etc.), the most common distribution used for modeling crash data remains the Poisson-gamma or Negative Binomial (NB) distribution. The Poisson-gamma distribution offers a simple way to accommodate the over-dispersion, especially since the final equation has a closed form and the mathematics to manipulate the relationship between the mean and the variance structures is relatively simple (Hauer, 1997).

Poisson-gamma models in highway safety applications have been shown to have the following probabilistic structure: the number of crashes at the ith entity (road section, intersections, etc.) and tth time period, Y_{it}, when conditional on its mean μ_{it}, is assumed to be Poisson distributed and independent over all entities and time periods as:

$$Y_{it}|\mu_{it} \sim \text{Po}(\mu_{it}) \qquad i = 1, 2, \ldots, I \quad \text{and} \quad t = 1, 2, \ldots, T \tag{1}$$

The mean of the Poisson is structured as:

$$\mu_{it} = f(X; \beta) \exp(e_{it}) \tag{2}$$

where $f(\cdot)$ is a function of the covariates (X), β a vector of unknown coefficients, and e_{it} is the model error independent of all the covariates.

It is usually assumed that $\exp(e_{it})$ is independent and gamma distributed with a mean equal to 1 and a variance $1/\phi$ for all i and t (with $\phi > 0$). With this characteristic, it can be shown that Y_{it}, conditional on $f(\cdot)$ and ϕ, is distributed as a Poisson-gamma random variable with a mean $f(\cdot)$ and a variance $f(\cdot)(1 + f(\cdot)/\phi)$, respectively. (Note: other variance functions exist for Poisson-gamma models, but they are not covered here since they are seldom used in highway safety studies. The reader is referred to Cameron and Trivedi (1998) and Maher and Summersgill (1996) for a description of alternative variance functions.) The probability density function (PDF) of the Poisson-gamma structure described above is given by the following equation:

$$f(y_{it}; \phi, \mu_{it}) = \frac{\Gamma(y_{it} + \phi)}{\Gamma(\phi) y_{it}!} \left(\frac{\phi}{\mu_{it} + \phi}\right)^{\phi} \left(\frac{\mu_{it}}{\mu_{it} + \phi}\right)^{y_{it}} \tag{3}$$

where y_{it} is the response variable for observation i and time period t, μ_{it} the mean response for observation i and time period t, and ϕ is the inverse dispersion parameter of the Poisson-gamma distribution.

The term ϕ is usually defined as the "inverse dispersion parameter" of the Poisson-gamma distribution. (Note: in the statistical and econometric literature, $\alpha = 1/\phi$ is usually defined as the dispersion parameter; in some published documents, the variable α as also been defined as the "over-dispersion parameter".) Usually the dispersion parameter or its inverse is assumed to be fixed, but recent research in highway safety has shown that the variance structure can potentially be dependent on the covariates (Heydecker and Wu, 2001; Miaou and Lord, 2003; Lord et al., 2005a).

As opposed to data collected in other fields of research, crash data have the uncommon attribute to frequently exhibit a distribution with a low sample mean. Similarly, it is not unusual for researchers and practitioners to develop statistical models using a limited number of observations (or sites) where data can be collected (see e.g., Lord, 2000; Oh et al., 2003; Kumara et al., 2003). Small sample sizes are attributed to the prohibitive costs of collecting crash data and other relevant variables (Lord and Bonneson, 2005).

Data characterized by a low sample mean has been sporadically studied in the traffic safety literature. As such, Maycock and Hall (1984) first raised the issue related to the low sample mean. Fridstrøm et al. (1995) further discussed this issue, while Maher and Summersgill (1996) showed how the goodness-of-fit of statistical models could be affected by a low sample mean. They defined this issue as the "low mean problem" (LMP). Subsequent to the identification and its effects on the development of statistical models, Wood (2002) proposed a method to test the fit of statistical models developed using data characterized with low sample mean values.

Despite the important work done on this topic, nobody has so far examined how the LMP affects the dispersion parameter of a Poisson-gamma model. In the traffic safety literature, the dispersion parameter is often relegated to a second-tier term and assumed to be estimated without any uncertainty (i.e., many studies did or still do not provide any uncertainties associated with the estimated dispersion parameter or its inverse). Given the fact that ϕ, the inverse dispersion parameter, is a critical parameter for developing confidence intervals (Myers et al., 2002; Wood, 2005) and for refining the estimates of the predicted mean when the empirical Bayes (EB) method is used (Hauer, 1997), one has to ensure that the dispersion parameter or its inverse has been properly estimated. In addition to the LMP, there is a need to study how a small sample size can affect the estimation of the dispersion parameter of Poisson-gamma models. When large databases are available for developing statistical models, as it is the case in many other fields of research, the outcome of the modeling effort is often assumed to be asymptotically distributed. In fact, the output provided by commercially available statistical software programs is based on the assumption that the outcome of the analysis is also asymptotically distributed (see Morris, 1997, for a discussion on this topic). This assumption indicates that as the number of observations becomes large, the

statistical inferences associated with the estimated coefficients become approximately normally distributed. Unfortunately, statistical models produced for traffic safety applications usually do not have the luxury of being developed using extremely large databases or with high sample means.

The purpose of this study is two-fold. The first objective seeks to verify whether the LMP affects the estimation of the dispersion parameter and, if so, to determine the magnitude of the problem. Three estimators commonly used for estimating the dispersion parameter of Poisson-gamma models for modeling motor vehicular crashes are evaluated: the method of moments (MM), the weighted regression (WR) and the maximum likelihood method (ML). The second objective consists of determining the effects of an unreliably estimated dispersion parameter on common analyses performed in highway safety studies. They include the application of the EB method and estimating confidence intervals for gamma mean (m) for a given site and the predicted response (y) for new sites not used as part of the model development.

To accomplish the objectives of this study, a series of Poisson-gamma distributions are simulated using different values describing the mean, the dispersion parameter, and the sample size. Two sets of distributions are estimated: (1) one with a fixed population mean (e.g., each site has same gamma mean) and (2) one with a population mean varying according to a lognormal distribution. The varying mean is employed to better characterize crash data observed in the field: (A) simulate sites with similar characteristics, but with different levels of exposure and (B) adding noise to the data (e.g., missing values, under-reporting, etc.). In an attempt to complement the output of the simulation study, Poisson-gamma models are fitted to crash data collected in Toronto, Ont. characterized by a low sample mean and small sample size. The study will show that a dataset characterized with a low sample mean combined with a small sample size can seriously affect the estimation of the dispersion parameter for extreme conditions, no matter which estimator is used in the estimation process. Consequently, an unreliably estimated dispersion parameter can significantly undermine EB estimates as well as the values calculated for the confidence intervals on the gamma mean and predicted response.

2. Previous work

Although a full paper could be devoted on previous work done on techniques for estimating the fixed dispersion parameter of Poisson-gamma models, this section only addresses the most relevant literature on this topic. The estimation of the dispersion parameter has been evaluated extensively in various fields, including statistics, econometrics, and biology. It is generally agreed that the first estimator for calculating the dispersion parameter was initially proposed by Fisher (1941). Fisher discussed how the ML method could be used for estimating the parameter. Following the publication of Fisher, several researchers expanded on his work by either refining the estimation method (Anscombe, 1950; Shenton and Wallington, 1962; Gourieroux et al., 1984a,b; Gourieroux and Visser, 1986; Cameron and Trivedi, 1986, 1990; Lawless, 1987) or describing potential biases, such as unstable variance and issues related to small sample sizes (Pieters et al., 1977; Willson et al., 1984a,b; Davidian and Carroll, 1987; Clark and Perry, 1989; Piegorsch, 1990; Dean, 1994; Toft et al., 2006). The studies listed above are only a fraction of the published studies done on this topic and, consequently, the reader is referred to Piegorsch (1990), Dean (1994) and Cameron and Trivedi (1998) for additional information on different estimation techniques.

Up until the late 1980s, researchers who have worked on small sample size (n) estimation of ϕ or α have usually focused their effort on determining whether ϕ should be estimated directly or, indirectly through its reciprocal $\alpha = 1/\phi$ (Clark and Perry, 1989). Given the outcome of these studies, it is the generally agreed that the dispersion parameter α should be estimated directly rather than its inverse ϕ (via the PDF of a Poisson-gamma distribution). According to the literature, the ML estimator of ϕ does not have any formal distribution, since there exists a finite probability that ϕ may not be calculable (Piegorsch, 1990); this usually occurs for data characterized with under-dispersion. It has also been shown that confidence intervals built for α are continuous and usually more symmetric than for ϕ (Ross and Preece, 1985).

Most recent studies on small sample size estimation techniques have usually focused on estimating potential the biases small sample sizes exert on different estimators. The studies have shown that different estimators (among them the MM and ML) perform well, except when the sample mean is low and the sample size becomes small. When this occurs, many estimators provide a biased estimate of the dispersion parameter (the distribution becomes highly skewed) and has a high probability of being mis-estimated.

Among the most significant studies on this topic, Clark and Perry (1989) compared two estimators, the MM and Maximum Quasi-Likelihood method, for different sample sizes ($n = 10$, 20, 30, and 50) and sample means ($\lambda = 1$, 3, 5, 10, 15, and 20). Using simulated data, they reported that both estimators become biased when $\lambda \leq 3.0$ and $n < 20$. In addition, under these conditions, the bias becomes more important as $\phi \to \infty$ (if the true value of the inverse of the dispersion parameter is known).

In a follow up study, Piegorsch (1990) examined the ML estimator and compared the results to the ones of Clark and Perry. Again, using a simulation experiment, the author noted that the ML estimator performed as well as the Quasi-likelihood for large sample sizes. However, Piegorsch reported that the ML estimator was slightly less accurate for small sample sizes than the Quasi-likelihood. It should be pointed out that in both studies, the dispersion parameter was biased for $n = 50$ both for the ML and MM estimators.

In a third study, Dean (1994) evaluated the effects of small sample sizes on the estimation of the dispersion parameter for seven different estimators, including the ML and MM. Dean simulated a NB model with two covariates, i.e. $\mu_i = \exp(\beta_0 + \beta_1 x_i)$, using a sample mean varying between 6 and 16. She reported that the ML estimator produced a biased estimate as the sample size decreased and ϕ increased (even for a sample mean equal

to 6). In fact, the biased results influenced the standard errors of the coefficients of the models.

In the last and very recent study, Toft et al. (2006) studied the stability of the parameters of the Poisson-gamma model when it is used for modeling microorganisms that are randomly distributed in a food matrix. These authors also used simulation to test the stability via the maximum likelihood method. However, in this case, they examined an alternative parameterization of the Poisson-gamma model commonly used in biology, where the mean and variance functions are defined as $\mu = \tau\nu$ and $\sigma^2 = \tau\nu^2$, respectively. Toft et al. (2006) reported that the parameter estimation becomes unstable when the parameter $\nu \to 0$ and $n \to 0$. Indeed, even for $\mu = 10$ and $n = 100$, the maximum likelihood method did not provide a reliable estimate of the parameters.

In summary, all previous studies described above have shown that small sample sizes and low sample mean values can significantly and negatively affect different estimators of the dispersion parameter. The dispersion parameter becomes increasingly underestimated as $n \to 0$, $\mu \to 0$ and $\alpha \to 0$ (the theoretical dispersion parameter if one were to know the true value). It should be pointed out that the researchers for all three studies have not examined the effects of a sample mean below 1 nor a dispersion parameter above 1 (e.g., $\alpha = 1/\phi > 1$) either mutually or independently. In all cases, the authors have argued that NB models characterized with a dispersion parameter equal to 1 is considered highly dispersed and, consequently, unlikely to be observed in the field. Finally, the data were simulated using the PDF of a NB distribution rather than using a simulation approach proposed in this study (described below).

3. Dispersion parameter estimators

Although numerous estimators have been proposed in the literature, the three most commons ones used by transportation safety modelers have been selected for this research. In addition, some of these estimators are used by existing statistical software programs, while others have been proposed for re-calibrating Poisson-gamma models developed using crash data (Persaud et al., 2002; FHWA, 2003; Lord and Bonneson, 2005).

3.1. Estimator 1

The first estimator consists of computing the dispersion parameter using the method of moments. For Estimator 1, the analyst is required to use the output of the regression analysis. Once the value of the dispersion parameter is estimated, the analyst puts the new value into the regression model and performs a new regression analysis. This iteration is performed until all values converge (i.e., dispersion parameter, coefficients, etc.). This estimator usually converges after a single iteration. Estimator 1 has been tested and used extensively in various fields of research (Gourieroux et al., 1984a,b; Lawless, 1987; Hauer, 1997; McCullagh and Nelder, 1989; Cameron and Trivedi, 1998). This estimator was proposed by FHWA (2003) for re-calibrating statistical models of motor vehicle crashes. The estimator is given by the following equation:

$$\hat{\alpha} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{\{(y_i - \hat{\mu}_i)^2 - \hat{\mu}_i\}}{\hat{\mu}_i^2} \tag{4}$$

In Eq. (4), the term p refers to the number of parameters included in the model and n is the sample size. The confidence intervals for estimating the uncertainty associated with this estimator are usually not calculated by analysts who use this estimator due to the complexity of mathematics involved for computing the intervals. Although the delta method can be used for building confidence intervals, its derivation can be very cumbersome (Cameron, 2005). Therefore, it is suggested to use a bootstrapping method for estimating the confidence intervals for Estimator 1 (Cameron and Trivedi, 1998; Cameron, 2005).

3.2. Estimator 2

The second estimator has been proposed by Cameron and Trivedi (1986). For Estimator 2, the dispersion parameter is estimated using a weighted regression analysis. The second estimator is given by the following equation:

$$\frac{(y_i - \hat{\mu}_i)^2 - y_i}{\hat{\mu}_i} = \alpha\hat{\mu}_i + \varepsilon \tag{5}$$

In essence, the functional form of this estimator is very similar to Estimator 1, but the actual crash count is subtracted from the square of the difference between the observed and predicted values. This estimator uses the same iterative procedure as Estimator 1. According to Cameron and Trivedi (1986), this estimator provides a more rational way for estimating the dispersion parameter since the variance of the left-hand of the equation is asymptotically distributed. This is apparently essential when the Poisson distribution is tested using the score test proposed by Lee (1986). Lord and Bonneson (2005) proposed this estimator for re-calibrating Poisson-gamma models developed using motor vehicle crashes. The confidence intervals can be computed directly from the output of the regression analysis.

3.3. Estimator 3

The third estimator was originally proposed by Fisher (1941) and later improved by Lawless (1987). This estimator consists of estimating the dispersion parameter using the maximum likelihood method. The log-likelihood function of Poisson-gamma model is given by the following equation:

$$\ell(\alpha, \hat{\mu}_i) = \sum_{i=1}^{n}\left(\ln\left\{\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})}\right\} + y_i \ln\{\hat{\mu}_i\} - (y_i + \alpha^{-1})\ln\{1 + \alpha\hat{\mu}_i\}\right) \tag{6}$$

This function can be written without call to the gamma function, such that

$$\ln\left\{\frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(\alpha^{-1})}\right\} = \sum_{j=0}^{y_i - 1}\ln\left\{\frac{1 + \alpha j}{\alpha}\right\} = \sum_{j=0}^{y_i - 1}\ln\{1 + \alpha j\} \tag{7}$$

(Lawless, 1987; Piegorsch, 1990). Incorporating Eqs. (7) into (6) produces the following log-likelihood function

$$\ell(\alpha, \hat{\mu}_i) = \sum_{i=1}^{n} \left(\sum_{j=0}^{y_i-1} \ln(1+\alpha j) + y_i \ln\{\hat{\mu}_i\} - (y_i + \alpha^{-1}) \ln\{1+\alpha\hat{\mu}_i\} \right) \quad (8)$$

The gradient elements of new log-likelihood function are defined as follows:

$$\nabla_\mu \ell = \frac{y_i}{\hat{\mu}_i} - \frac{1+\alpha y_i}{1+\alpha\hat{\mu}_i} \quad (9a)$$

$$\nabla_\alpha \ell = \sum_{i=1}^{n} \left(\sum_{j=0}^{y_i-1} \frac{j}{1+\alpha j} + \alpha^{-2} \ln\{1+\alpha\hat{\mu}_i\} - \frac{\hat{\mu}_i(y_i+\alpha^{-1})}{1+\alpha\hat{\mu}_i} \right) \quad (9b)$$

Using the gradients, the Newton-Raphson (NR) scoring algorithm can be used to find the values of the log-likelihood function through the maximum likelihood method (Fletcher, 1970; Walsh, 1975). According to Lawless (1987), this estimator is valid only if the parameters βs (the coefficients of the model) and α are asymptotically independent and normally distributed. He demonstrated this property using large-sample approximations (high sample mean and large sample size). The dispersion parameter of the Poisson-gamma model in GENSTAT (Payne, 2000), the software used in this study, is estimated using the approach proposed by Lawless (1987). SAS also uses the same approach for estimating the dispersion parameter (SAS Institute Inc., 2002).

The NR algorithm can be used for building the confidence intervals associated with the estimator (Lawless, 1987; Payne, 2000). In fact, many statistical software programs now provide confidence intervals for the estimated dispersion parameter.

In the subsequent sections, the results are shown using the inverse dispersion parameter (ϕ). Given that most safety studies show the inverse (often called the "k-value"), the results are presented in this manner to render the comparison with previous work on this subject easier.

4. Simulation framework

This section presents the characteristics of the simulation study intended to illustrate how the LMP and the small sample size affect the three estimators described above. Rather than simulating the data using the PDF of a NB distribution, the simulation was performed using a mixed distribution where the sample mean and the count data were simulated in a step-wise fashion. This approach offered more flexibility, particularly for the second series of simulation runs. In this exercise, two series of simulation runs were performed. The first series consisted of simulating a Poisson-gamma distribution using a fixed sample population mean. In other words, each simulated observation was taken from a sample population having the same 'gamma' mean.

The second series involved the simulation of a Poisson-gamma distribution, but in this case the mean was taken from a sample population that is assumed to follow a lognormal distribution. The second series was intended to more adequately replicate real data where the mean for each observation varies as a function of exposure. Larger exposure is associated with a higher mean, although usually increasing at a decreasing rate as exposure increases ($y = \beta_0 F^{\beta_1}$). For instance, a lognormal distribution was fitted for a dataset previously used by Lord (2000) and characterized the data adequately (see Fig. 1). In this dataset, the predicted means ($\hat{\mu}_i$) (after the statistical model was fitted) varied according to a lognormal distribution with a logarithmic mean equal to 1.14 and a variance equal to 0.49. Using a lognormal distribution does not necessarily imply that other distributions, such as the gamma or beta distributions could not fit the data better, but this distribution is simple to manipulate and prevents negative values. In the end, using a varying sample mean can reproduce noises one can observe in the data collected in the field (e.g., missing values, under-reporting collisions, etc.). Finally, as discussed in Section 1, the Poisson-gamma model is in fact used as an approximation for modeling crash data (Lord et al., 2005b).

The simulation framework or algorithm is described below.

(A) Fixed sample population mean
  (1) Generate a mean value (ρ_i) for observation i from a fixed sample population mean (λ):
  $\rho_i = \lambda$
  (2) Generate a value (δ_i) from a gamma distribution with the mean equal to 1 and the parameter ϕ:
  $\delta_i \sim \text{gamma}\left(\phi, \frac{1}{\phi}\right)$
  (3) Calculate the mean (μ_i) for observation i:
  $\mu_i = \rho_i \times \delta_i$
  (4) Generate a discrete value (Y_i) for observation i from a Poisson distribution with mean μ_i:
  $Y_i \sim \text{Poisson}(\mu_i)$
  (5) Repeat steps 1 and 4 "n" times for the number of observations under study (defined here as the sample size).

(B) Varying sample population mean
  (1) Generate a mean value (ρ_i) for observation i from a sample population that follows a lognormal distribution:
  $\rho_i \sim \log\text{normal}(\lambda, \sigma)$
  (2) Generate a value (δ_i) from a gamma distribution with the mean equal to 1 and the parameter ϕ:
  $\delta_i \sim \text{gamma}\left(\phi, \frac{1}{\phi}\right)$
  (3) Calculate the mean (μ_i) for observation i:
  $\mu_i = \rho_i \times \delta_i$
  (4) Generate a discrete value (Y_i) for observation i from a Poisson distribution with mean μ_i:
  $Y_i \sim \text{Poisson}(\mu_i)$
  (5) Repeat steps 1 and 4 "n" times for the number of observations under study.

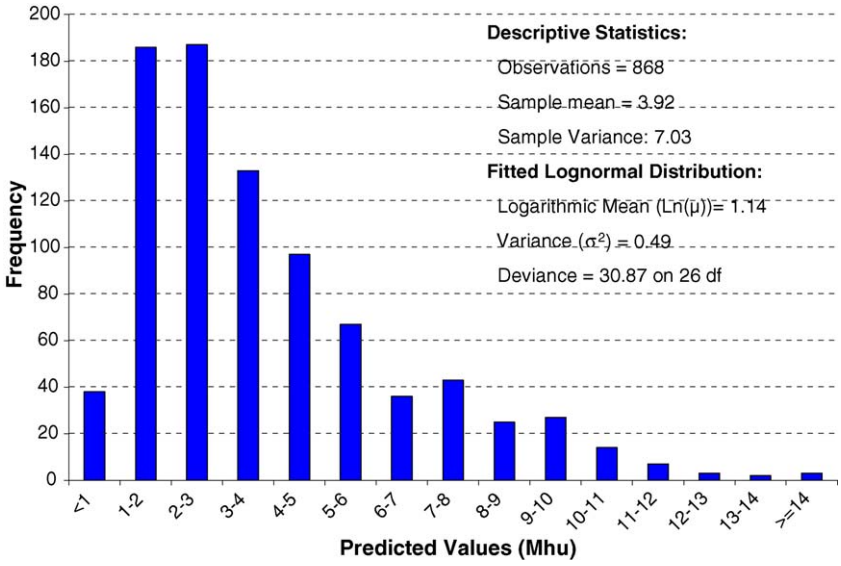

Fig. 1. Distribution of the predicted values (the predicted values were estimated using the following functional form: $\mu_i = 0.0002433 \times F_{1i}^{0.527} \times F_{2i}^{0.568} \times e^{(8.61E-6 \times F_{2i})}$ with $\phi = 6.91$. At the time model was estimated, the software program was unable to provide inferences associated with the estimate of the inverse dispersion parameter) (μ_i) for 868 signalized intersections in Toronto, Ont. for 1995 (Lord, 2000).

The parameterization of the gamma distribution above, $\theta \sim \text{gamma}(a, b)$, needs to be used when its mean and variance are defined as $E(\theta) = ab$ and $\text{Var}(\theta) = ab^2$, respectively. (There also exist an alternative approach for defining $E(\theta)$ and $\text{Var}(\theta)$.) GENSTAT (Payne, 2000) uses this parameterization for generating gamma distributed values. It can be shown that when $E(\theta) = 1$ and $\text{Var}(\theta) = 1/\phi$ (where $a = \phi$ and $b = 1/\phi$), the Poisson-gamma function gives rise to a NB distribution with $\text{Var}(Y) = \mu + \mu^2/\phi$ (Cameron and Trivedi, 1998). The simulation effort was performed using GENSTAT (Payne, 2000) for the following values:

- Sample size or number of observations (n): **50**, **100**, 1000.
- Inverse dispersion parameter (ϕ) = **1/2**, **1**, 2.
- Sample population mean (λ) = **0.5**, **1.0**, 10 (fixed mean only).

The values in bold character characterize data subjected to extreme conditions: low sample mean and/or small sample sizes. (Note: Clark and Perry (1989), Piegorsch (1990), and Dean (1994) reported that data subjected to $\phi = 1$ are considered highly dispersed and unusual.) The other values are used to assess the asymptotic properties of the data for $n \times \lambda \to \infty$ (see Lawless, 1987).

For each combination of sample size, dispersion parameter, and sample mean, the simulation was replicated 30 times. Due to the partial manual manipulations needed to summarize the results of the simulation process and the numerous simulation runs, the number of replications was limited to 30. A few trial runs were performed using a larger number of replications, but the summary statistics as well as the inferences did not change compared to the original values estimated from the original number of replications. At the end of the replications, the standard statistics, such as the mean, standard deviation, maximum and minimum values, and the number of times the estimator did not converge (when it occurred) were computed.

5. Simulation results

This section summarizes the results of the simulation output. The first part describes the simulation results for the fixed sample mean. The second part summarizes the simulation results for the varying sample mean.

5.1. Fixed sample mean

In order to test the asymptotic properties of the estimators, a series of simulation runs was performed using a sample mean equal to 10. The results of the simulation are summarized in Table 1.

Table 1 shows that for a sample size of 1000, all three estimators accurately predicted the theoretical values of the inverse dispersion parameter for $\phi = 1/2$, $\phi = 1$, and $\phi = 2$. For a sample size of 100, the three estimators also predicted the theoretical values accurately, but the standard deviation is a little larger and the extreme simulated values (i.e., max and min) are becoming more noticeable. With a sample size of 50, which is considered a small sample size according to Clark and Perry (1989), the estimators tend to overestimate the theoretical mean, particularly for $\phi = 1$ and $\phi = 2$, respectively (i.e., meaning the predicted values are higher than the theoretical value for the inverse dispersion parameter). Furthermore, the values describing the standard deviations are very large for the same two inverse dispersion parameters. For $\phi = 2$, the difference between the maximum and minimum values varies by a factor of two.

The overestimation of the inverse dispersion parameter is explained by the fact the parameter is no longer normally distributed. A closer look at the data shows that the distribution for $\phi = 1$ and $\phi = 2$ was highly skewed. As described above, this outcome was also noted by Clark and Perry (1989) and Piegorsch (1990). In short, for a sample size equal to 50 and a sample mean

Table 1
Simulation results for $\lambda = 10$ (fixed mean)

Characteristics	$\phi = 1/2$				$\phi = 1$				$\phi = 2$			
	$\hat{\lambda}$	MM[a]	WR[b]	ML[c]	$\hat{\lambda}$	MM	WR	ML	$\hat{\lambda}$	MM	WR	ML
$n = 50$[d]												
Mean	10.47 (2.36)[e]	0.55 (0.14)	0.56 (0.14)	0.51 (0.07)	9.60 (1.37)	1.10 (0.31)	1.12 (0.31)	1.08 (0.22)	10.30 (0.96)	2.23 (0.82)	2.27 (0.84)	2.13 (0.62)
Max	13.90	0.83	0.85	0.70	11.76	1.84	1.88	1.72	12.14	5.18	5.28	4.47
Min	5.86	0.33	0.34	0.38	7.02	0.52	0.53	0.83	8.38	1.40	1.42	1.59
$n = 100$												
Mean	9.73 (1.05)	0.53 (0.14)	0.54 (0.14)	0.51 (0.09)	9.96 (0.90)	1.03 (0.15)	1.04 (0.15)	1.01 (0.13)	10.05 (0.74)	2.07 (0.35)	2.09 (0.36)	2.01 (0.35)
Max	11.82	0.84	0.85	0.69	11.70	1.30	1.32	1.28	11.75	2.87	2.90	3.12
Min	7.94	0.32	0.33	0.35	8.24	0.63	0.64	0.78	8.68	1.58	1.60	1.42
$n = 1000$												
Mean	10.04 (0.49)	0.50 (0.04)	0.50 (0.04)	0.50 (0.02)	9.92 (0.28)	0.97 (0.08)	0.97 (0.08)	0.98 (0.06)	9.99 (0.25)	2.00 (0.14)	2.00 (0.14)	2.01 (0.10)
Max	11.43	0.57	0.58	0.53	10.50	1.15	1.15	1.09	10.40	2.24	2.24	2.18
Min	9.11	0.41	0.41	0.46	9.32	0.82	0.82	0.85	9.50	1.81	1.81	1.79

[a] Method of moments (Estimator 1).
[b] Weighted regression (Estimator 2).
[c] Maximum likelihood (Estimator 3).
[d] Sample size.
[e] Standard deviation.

equal to 10, the estimated dispersion parameter has a slight probability of being mis-estimated, although less than 5%.

Between the three estimators, the third estimator (ML) offers a slightly better prediction both in terms of mean and standard deviations. However, the comparison between the three estimators was not statistically significant.

The results of the simulation for $\lambda = 1.0$ and $\lambda = 0.5$ are presented in Tables 2 and 3, respectively. In Table 2, it can be observed that for a sample size equal to 1000, all three estimators predict the theoretical value adequately. However, the estimators start to overestimate the theoretical value with a sample size equal to 100. For $\phi = 2$, the mean of the estimated values is in fact 1.5 times larger than theoretical value. All three estimators for $\phi = 2$ and $n = 50$ are completely unreliable. In fact, the data produced from many simulation runs erroneously exhibited a pure Poisson distribution. It should be pointed out that for $\lambda = 1.0$, no estimator outperformed the others.

Table 3 exhibits similar characteristics as the ones presented in Table 2. However, the values now become unreliable or unstable for a sample size equal to 100 for $\phi = 2$. To some degree, the values are also unstable for $\phi = 1$ and a sample size equal to 100. For a sample size equal to 50, the data exhibited pure Poisson properties since many estimated values describing the inverse dispersion parameter were above 10. However, for the same sample size, some simulation runs exhibited significant over-dispersion with values $\phi < 1$. The simulation output shows that the inverse dispersion parameter estimated from data characterized by low sample mean values and small sample size is highly unreliable as the theoretical value of ϕ increases.

In summary, the simulation results for the fixed sample mean have shown the following characteristics:

1. For large sample size and high mean, all the three estimators predicted values very close to the theoretical value;
2. As the sample size decreases, the distribution of the estimated values becomes more skewed, which significantly increases the mean of the estimated values of the inverse dispersion parameter;
3. As a result, the inverse dispersion parameter is more likely to be unreliable estimated, no matter which estimator is used;
4. The standard error of the predicted values increases as the sample size becomes smaller; this characteristic is more important as the theoretical inverse dispersion parameter increases;
5. The results are consistent with previous work on this topic.

The next section describes the simulation results for the varying sample mean.

5.2. Varying sample mean

In this exercise, the sample population mean is assumed to follow a lognormal distribution. It should be pointed out that the generated sample mean will be slightly higher since $E(Y) = \exp(\mu + \sigma^2/2)$. Nonetheless, this outcome did not affect

Table 2
Simulation results for $\lambda = 1$ (fixed mean)

Characteristics	$\phi = 1/2$				$\phi = 1$				$\phi = 2$			
	$\hat{\lambda}$	MM[a]	WR[b]	ML[c]	$\hat{\lambda}$	MM	WR	ML	$\hat{\lambda}$	MM	WR	ML
$n = 50$[d]												
Mean	1.02 (0.21)[e]	0.81 (0.38)	0.82 (0.39)	0.69 (0.31)	1.03 (0.15)	1.32 (0.63)	1.35 (0.64)	1.30 (0.75)	1.09 (0.15)	5.62 (11.06)	5.73 (11.29)	5.73 (12.61)
Max	1.46	2.01	2.05	1.66	1.40	3.33	3.40	4.27	1.44	62.43	63.71	71.66
Min	0.66	0.41	0.42	0.34	0.76	0.58	0.59	0.58	0.80	0.92	0.94	1.09
$n = 100$												
Mean	1.01 (0.15)	0.55 (0.15)	0.56 (0.15)	0.52 (0.12)	1.02 (0.16)	1.06 (0.47)	1.07 (0.47)	1.03 (0.43)	1.00 (0.11)	3.27 (2.72)	3.31 (2.75)	3.38 (3.02)
Max	1.31	0.97	0.98	0.91	1.43	2.60	2.63	2.31	1.21	12.37	12.50	14.00
Min	0.64	0.23	0.23	0.32	0.81	0.36	0.36	0.49	0.77	1.06	1.07	1.23
$n = 1000$												
Mean	0.99 (0.06)	0.51 (0.06)	0.51 (0.06)	0.50 (0.04)	0.99 (0.04)	1.02 (0.13)	1.02 (0.13)	1.01 (0.12)	0.99 (0.03)	2.01 (0.28)	2.01 (0.28)	2.01 (0.30)
Max	1.08	0.63	0.63	0.57	1.08	1.40	1.40	1.33	1.05	2.73	2.73	2.84
Min	0.87	0.39	0.39	0.42	0.87	0.75	0.75	0.81	0.92	1.61	1.61	1.56

[a] Method of moments (Estimator 1).
[b] Weighted regression (Estimator 2).
[c] Maximum likelihood (Estimator 3).
[d] Sample size.
[e] Standard deviation.

Table 3
Simulation results for $\lambda = 0.5$ (fixed mean)

Characteristics	$\phi = 1/2$				$\phi = 1$				$\phi = 2$			
	$\hat{\lambda}$	MM[a]	WR[b]	ML[c]	$\hat{\lambda}$	MM	WR	ML	$\hat{\lambda}$	MM	WR	ML
$n = 50$[d]												
Mean	0.54 (0.15)[e]	0.70 (0.35)	0.71 (0.36)	0.67 (0.36)	0.51 (0.09)	3.18 (4.67)	3.25 (4.77)	2.74 (3.68)	0.52 (0.10)	6.04 (8.18)	6.17 (8.35)	5.72 (8.08)
Max	0.96	1.67	1.70	1.80	0.66	25.75	26.27	20.10	0.74	29.65	30.25	32.98
Min	0.32	0.32	0.21	0.27	0.34	0.52	0.53	0.47	0.30	0.91	0.92	1.21
$n = 100$												
Mean	0.47 (0.11)	0.74 (0.38)	0.75 (0.39)	0.67 (0.29)	0.52 (0.09)	1.45 (0.98)	1.47 (0.99)	1.28 (0.83)	0.53 (0.10)	3.90 (3.83)	3.94 (3.87)	3.67 (3.42)
Max	0.67	1.89	1.91	1.45	0.68	4.30	4.34	3.48	0.70	19.72	19.92	17.82
Min	0.26	0.23	0.23	0.29	0.32	0.36	0.36	0.3	0.37	1.35	1.36	1.25
$n = 1000$												
Mean	0.51 (0.03)	0.52 (0.07)	0.52 (0.07)	0.49 (0.06)	0.49 (0.03)	1.01 (0.21)	1.01 (0.21)	1.00 (0.20)	0.51 (0.02)	2.08 (0.47)	2.08 (0.47)	2.09 (0.47)
Max	0.56	0.78	0.78	0.75	0.54	1.53	1.53	1.51	0.56	3.00	3.00	3.10
Min	0.46	0.38	0.38	0.41	0.44	0.66	0.66	0.69	0.47	1.29	1.29	1.29

[a] Method of moments (Estimator 1).
[b] Weighted regression (Estimator 2).
[c] Maximum likelihood (Estimator 3).
[d] Sample size.
[e] Standard deviation.

the simulated data. Simulation runs performed for a sample mean equal to 5 or above has shown that, for a sample size of 1000, all three estimators predicted values similar to the values predicted in Table 1. The variance (σ^2) of the sample mean was set to 0.50 (similar to the dataset shown in Fig. 1), which is intended to reproduce a large variation (in exposure) within the lognormal distribution.

Table 4 summarizes the simulation results for a sample mean equal to 1. This table shows interesting characteristics. First, the estimators no longer provide similar values. In fact, for Estimator 3 (ML), the estimated value is usually lower than the first two estimators, at least for $\phi = 1$ and $\phi = 2$. It is unclear at this point why the ML estimates estimate lower values than the theoretical values. Second, for a sample size equal to 1000, the estimated values are very close to the theoretical value, although the standard errors are larger than for the fixed sample mean. Third, for a sample size of 100, the method of moments does not provide accurate estimates. The problem is more important for $\phi = 1/2$ and $\phi = 2$. Fourth, none of the estimators provide good estimates for $\phi = 2$, particularly for a sample size equal to 50 and 100. It should be pointed out that the method of moments failed to converge (or provided non-positive values or under-dispersion) for many simulation runs for a sample size equal to 50.

Table 5 summarizes the simulation runs for a sample mean equal to 0.5. This table shows more drastic results than the ones shown in Table 4. For instance, all three estimators become unstable for $\phi = 1$ and $\phi = 2$, and a sample size equal to 100. As described in Table 4, the first estimator performs poorly for a sample size below 100 and for $\phi = 1/2$. For a sample size equal to 50 and $\phi = 2$, none of the estimators converged or provided reasonable results.

In summary, the simulation results for the varying sample mean have shown the following characteristics:

1. All three estimators performed relatively well for a sample size equal to 1000, with the exception of Estimator 1 for $\phi = 2$;
2. Estimator 1 becomes highly unreliable for a sample size below 100;
3. In most circumstances, Estimator 3 provides better results and is usually more stable than the other two estimators;
4. Overall, Estimator 2 performed relatively better than Estimator 1;
5. Although the confidence intervals produced by Estimators 2 and 3 indicate that the inverse dispersion parameter was adequately estimated for extreme conditions, the theoretical value of the inverse dispersion parameter lied outside the 95% confidence interval of the estimated value;
6. Data characterized by a low sample mean and a small sample size are most likely highly dispersed, although the estimators may show otherwise (the dispersion parameter is underestimated).

The next section describes the performance of the three estimators when observed data are used for developing Poisson-gamma models.

6. Observed data

A sample dataset was utilized to determine if the effects of low sample mean and small sample size on the inverse dispersion parameter described via simulation could be replicated using actual data collected in the field. The dataset was initially collected for a project related to the development of statistical models for predicting the safety performance of unsignalized intersections in Toronto, Ont. (Lord, 2000). The data were collected for the years 1990–1995 at 59 unsignalized intersections. Fatal and non-fatal injury collisions were used in this example, since the sample population mean was about one crash per year. The characteristics of the data are summarized in Table 6.

A Poisson-gamma model was initially fitted for the entire dataset. For this dataset, each year was treated as an independent observation. Consequently, the model for the complete dataset contained 354 observations. For the sake of simplicity, the temporal effect was not included in the development of the model. Since the dataset does not include missing values, only the standard errors of the coefficients will be affected no matter the type of correlation structure used in the model development (see Lord and Persaud, 2000). The dispersion parameter was estimated for the three estimators described above.

Two subsets of 50 observations were then extracted. The observations were taken from the same year (out of the 59 sites). Then, for each subset, a Poisson-gamma model was fitted and the three estimators calculated. It should be pointed out that three additional subsets were tested, but the statistical models produced counterintuitive values for the coefficients (i.e., negative values). The results are therefore not shown herein.

The functional form used for the example was the following:

$$\mu_i = \beta_0 F_{i1}^{\beta_1} F_{i2}^{\beta_2} \tag{10}$$

where μ_i is the predicted number of crashes per year for site i; F_{i1}, F_{i2} the entering AADT flows for the major and minor approaches for site i; β_0, β_1, β_2 are the coefficients to be estimated.

Although the functional form is not the most adequate for describing the relationship between crashes and exposure, this form is still the most favored by transportation safety modelers for modeling crash data at intersections. As reported by Miaou and Lord (2003), the functional form above does not appropriately fit the data near the boundary conditions.

The modeling results are presented in Table 7. This table shows that even for the full dataset, the dispersion parameter varies greatly among the three estimators. In addition, the uncertainty associated with the last two estimators is relatively large. For the two subsets, one can see that the dispersion parameter becomes highly unstable for all three estimators. On the one hand, the estimators for subset 1 are all positive, but the standard errors are extremely large (at least for Estimators 2 and 3). On the other hand, the statistical model for subset 2 exhibits, depending on the estimator, either a pure Poisson distribution or under-dispersion. The results of the modeling process using actual data show that the three estimators do not provide consistent values for the same dataset and appear to correspond

Table 4
Simulation results for original $\lambda = 1$ (varying mean)

Characteristics	$\phi = 1/2$				$\phi = 1$				$\phi = 2$			
	$\hat{\lambda}$[a]	MM[b]	WR[c]	ML[d]	$\hat{\lambda}$[a]	MM	WR	ML	$\hat{\lambda}$[a]	MM	WR	ML
$n = 50$[e]												
Mean	1.26 (0.37)[f]	1.19 (1.01)	1.08 (0.95)	0.59 (0.26)	1.21 (0.19)	3.32 (4.39)	1.72 (0.84)	1.21 (0.39)	1.25 (0.21)	6.55 (13.75)	5.03 (5.42)	7.82 (18.26)
Max	1.84	4.25	3.91	1.36	1.70	18.84	3.39	1.75	1.66	77.36	30.11	75.82
Min	0.54	0.19	0.14	0.28	0.80	0.47	0.52	0.53	0.86	0.59	0.95	0.98
Not converged[g]		2				5				11		
$n = 100$												
Mean	1.34 (0.27)	0.73 (0.61)	0.63 (0.29)	0.47 (0.12)	1.30 (0.21)	2.41 (3.69)	1.24 (0.52)	1.06 (0.43)	1.37 (0.20)	2.66 (2.31)	2.74 (2.43)	2.36 (1.63)
Max	1.93	2.89	1.26	0.71	1.74	16.56	2.75	2.58	2.00	8.77	14.11	9.62
Min	0.85	0.21	0.14	0.28	1.00	0.37	0.20	0.50	0.99	0.31	0.71	1.11
Not converged										4		
$n = 1000$												
Mean	1.28 (0.08)	0.51 (0.11)	0.57 (0.15)	0.48 (0.05)	1.28 (0.08)	1.10 (0.32)	1.07 (0.22)	0.92 (0.09)	1.29 (0.05)	2.12 (0.75)	2.21 (0.43)	1.67 (0.22)
Max	1.46	0.77	0.78	0.62	1.43	1.84	1.49	1.09	1.36	4.08	2.91	2.21
Min	1.16	0.32	0.23	0.38	1.14	0.67	0.65	0.79	1.14	0.80	1.60	1.24

[a] Theoretical value based on the lognormal simulation: 1.28.
[b] Method of moments (Estimator 1).
[c] Weighted regression (Estimator 2).
[d] Maximum likelihood (Estimator 3).
[e] Sample size.
[f] Standard deviation.
[g] The number of times the estimator did not converge.

Table 5
Simulation results for original $\lambda = 0.5$ (varying mean)

Characteristics	$\phi = 1/2$				$\phi = 1$				$\phi = 2$			
	$\hat{\lambda}$[a]	MM[b]	WR[c]	ML[d]	$\hat{\lambda}$[a]	MM	WR	ML	$\hat{\lambda}$[a]	MM	WR	ML
$n = 50$[e]												
Mean	0.61 (0.29)[f]	0.99 (1.64)	1.01 (0.68)	0.73 (0.44)	0.62 (0.13)	1.77 (2.95)	2.07 (2.36)	1.85 (1.69)	N/A	N/A	N/A	N/A
Max	1.32	8.80	2.96	2.49	0.96	13.77	11.86	7.84	N/A	N/A	N/A	N/A
Min	0.14	0.04	0.17	0.29	0.34	0.08	0.21	0.52	N/A	N/A	N/A	N/A
Not converged[g]		9		2		13		2		30	30	30
$n = 100$												
Mean	0.66 (0.13)	1.40 (3.04)	0.73 (0.52)	0.59 (0.44)	0.63 (0.09)	2.29 (6.74)	1.93 (2.62)	2.12 (4.45)	0.69 (0.11)	2.27 (2.84)	3.15 (2.55)	3.13 (3.69)
Max	0.91	12.99	2.99	2.83	0.83	38.33	15.41	25.68	0.93	12.73	13.14	19.26
Min	0.30	0.11	0.13	0.23	0.47	0.31	0.60	0.53	0.47	0.28	0.72	0.86
Not converged		3								7		6
$n = 1000$												
Mean	0.65 (0.04)	0.55 (0.15)	0.53 (0.14)	0.47 (0.06)	0.63 (0.04)	1.35 (0.63)	1.14 (0.33)	0.92 (0.15)	0.63 (0.03)	7.93 (27.39)	2.13 (0.88)	1.56 (0.23)
Max	0.75	0.93	0.77	0.59	0.70	2.82	2.17	1.37	0.69	154.70	4.58	2.19
Min	0.55	0.30	0.19	0.36	0.52	0.43	0.56	0.71	0.56	0.95	0.98	1.18
Not converged						1						

[a] Theoretical value based on the lognormal simulation: 0.64.
[b] Method of moments (Estimator 1).
[c] Weighted regression (Estimator 2).
[d] Maximum likelihood (Estimator 3).
[e] Sample size.
[f] Standard deviation.
[g] The number of times the estimator did not converge.

Table 6
Data characteristics of unsignalized 4-legged intersections (Lord, 2000)

Statistic	Injury crashes per year	Flow major (AADT)	Flow minor (AADT)
Mean	1.00	26005	1878
Standard deviation	1.12	10832	1616
Max	5	53531	8836
Min	0	5669	378

Table 7
Modeling output for full dataset and subsets 1 and 2

Statistic	Full dataset	Subset 1	Subset 2
Intercept (β_0)	−7.65 (1.83)	−9.58 (4.92)	−9.55 (4.96)
ln(F1) (β_1)	0.607 (0.142)	0.711 (0.364)	0.687 (0.367)
ln(F2) (β_2)	0.209 (0.103)	0.333 (0.279)	0.356 (0.281)
ϕ (MM)	2.91	4.52	2.30
ϕ (WR)	5.78 (3.00)	12.49 (21.17)	−22.9
ϕ (ML)	4.83 (2.37)	10.57 (24.9)	N/A[a]
Observations	354	50	50
Mean	1.00	1.06	0.98
Standard deviation	1.12	1.15	1.03

[a] Did not converge; data may be characterized by under-dispersion.

with the outcome of the simulation output (particularly with the varying sample mean), in which the inverse dispersion parameter becomes unreliably estimated as the sample size decreases.

7. Discussion

The results of the analysis presented above raise a few important issues that merit further discussion. First, as described above, the statistical inferences associated with the estimation of the dispersion parameter or its inverse is usually not a concern for many transportation safety modelers. For instance, before the wide availability of commercial statistical software programs, very few researchers, if none at all, provided information on the confidence intervals associated with the dispersion parameter. The author could not find any paper that provided such information in the 1980s and mid-1990s. However, to be fair with the researchers from this period, a large part of the problem was related to the complexity of building confidence intervals and the lack of tools to do so (see Lawless, 1987; McCullagh and Nelder, 1989). Nonetheless, there are currently still published documents that do not provide such information (e.g., see Persaud et al., 2004; Tarko and Kanodia, 2004; Hauer et al., 2004). Interestingly, some researchers do not even bother providing information about the dispersion parameter for Poisson-gamma models produced from crash data (e.g., Noland and Quddus, 2004; Kumara and Chin, 2004).

Even with increasing use of statistical software programs that can now provide the confidence intervals for the dispersion parameter (note: the values are often estimated using approximation techniques, such as the delta method), the author has yet to see a paper in which the characteristics of the dispersion parameter are discussed to the same degree as to the coefficients of the statistical model (if we make abstraction of recent papers on varying dispersion parameters). In datasets characterized with a low sample mean and a small sample size, it is more than likely that the standard errors associated with the coefficients of the covariates of the statistical models may be erroneous. Many methods used for estimating the standard errors of Poisson-gamma models are based on the dispersion parameter (Cameron and Trivedi, 1998; Myers et al., 2002; Wood, 2005).

What is more troublesome is the fact that the safety modeler may not be aware the dispersion parameter was actually mis-estimated. As detailed in the simulation results, for many simulation runs, Estimators 2 and 3 seemed to provide very good statistical inferences. Yet, the theoretical value used in the simulation was actually located beyond the 95% confidence intervals provided by GENSTAT (Payne, 2000) for the estimated value. For large sample sizes ($n = 1000$) and high sample mean ($\lambda = 10$), all the theoretical values were located inside the 95% confidence bound of the estimated values.

The second issue is related to the effects of an unreliably estimated inverse dispersion parameter on two types of analysis commonly used in highway safety. The first type of analysis is the widely applied EB method. This method has become increasingly popular since it corrects for the regression-to-the-mean (RTM) bias, refines the predicted mean of an entity, and is relative simple to manipulate compared to the full-Bayes approach (Hauer and Persaud, 1984; Hauer, 1997). The EB method combines information obtained from a reference group having similar characteristics with the information specific to the site under study with characteristics similar to the ones found in the reference group. A weight factor is assigned to both the reference population and the site under study. The equation can be defined the following way:

$$\hat{\hat{\mu}}_i = \gamma_i \hat{\mu}_i + (1 - \gamma_i) y_i \tag{11}$$

where $\hat{\hat{\mu}}_i$ is the EB estimate of the expected number of crashes per year for site i; $\hat{\mu}_i$ the ML estimate produced from a Poisson-gamma model fitted using the reference population for site i (crashes per year); $\gamma_i = 1/(1 + \hat{\mu}_i/\phi)$, the weight factor estimated as a function of the ML estimate and the inverse dispersion parameter; y_i is the observed number of crashes per year at site i.

As it can be seen above, the inverse dispersion parameter plays an important role for estimating the weight factor.

To determine how an unreliably estimated inverse dispersion parameter affects the weight factor and the EB estimate, the output for one of the simulation runs was used to this effect. The original run consisted of simulating data for $n = 100$, $\lambda = 0.5$, and $\phi = 1$. As detailed in Tables 2 and 4, a mis-estimated inverse dispersion parameter can be off by a factor of two to three, i.e. $\phi = 2$, $\phi = 3$. The results are shown in Table 8.

Table 8 shows that, even with a small error in the mis-specification of the inverse dispersion parameter, the EB estimate can be greatly affected. In this example, the magnitude of the relative difference could be as high as 43%. When the dispersion parameter is grossly mis-estimated, i.e. $\phi \to \infty$, the EB estimate could be off by 100% (from 2.0 to 1.0 for the most extreme values). In this case, the transportation safety modeler may erroneously believe that RTM may not exist with the

Table 8
Effects of an unreliably estimated dispersion parameter ($\lambda = 0.5$)

y	Freq[a]	$\phi = 1$[b]		$\phi = 2$			$\phi = 3$		
		γ	$\hat{\hat{\mu}}$	γ	$\hat{\hat{\mu}}$	Diff (%)[c]	γ	$\hat{\hat{\mu}}$	Diff (%)[c]
0	67	0.67	0.33	0.80	0.40	20.0	0.86	0.46	28.6
1	26	0.67	0.67	0.80	0.670	10.0	0.86	0.57	14.3
2	5	0.67	1.00	0.80	0.860	20.0	0.86	0.71	28.6
3	1	0.67	1.33	0.80	5.0	25.0	0.86	0.86	35.7
4	0	–	–	–	–	–	–	–	–
5	1	0.67	2.00	0.80	1.40	30.0	0.86	1.14	42.9
Total	100								
Ave	0.44								

[a] Frequency or number of observations.
[b] Theoretical value used for the simulation.
[c] Relative difference.

crash data under study. It should be pointed out that the absolute difference may appear to be small (i.e., below 0.5 crash per year for most counts). However, as the inverse dispersion parameter becomes increasingly miss-specified, the MLE estimate becomes as good an estimator as the EB estimate even for small absolute differences. In fact, the difference between both estimates is also small (as shown for $\phi = 3$ where the EB estimate for $y = 0$ is 0.46 and the MLE estimate is 0.50). The blurry line delimiting both estimates is particularly true when one considers the uncertainty associated with each estimate.

The second type of analysis is related to the estimation of confidence intervals for the mean (μ), gamma mean (m) and predicted response (y). There are numerous applications where the confidence intervals on the predicted mean (y) can play an important role. For instance, when predictive models are used for estimating the safety performance of different highway design alternatives, the confidence intervals (around μ and m) can play an vital role in the selection process competitive projects. Examples where the predicted crashes are used for identifying competitive highway design alternatives can be found here: FHWA (2003), Lord and Persaud (2000), and Kononov and Allery (2004). In another application, confidence intervals can also be useful for screening hazardous sites, "black spots" or the so-called "sites with promise" (Hauer, 1996). Erroneously selecting sites for treatment (i.e., false positive) can lead to a significant waste of financial resources, not withstanding miss opportunities to save lives.

There are difference methods for estimating the confidence intervals of the predicted values generated from generalized linear models (Cameron and Trivedi, 1998; Myers et al., 2002). The most recent and relevant method has been proposed by Wood (2005) who specifically developed a procedure for computing confidence intervals for statistical models developed from crash data. For the sake of simplicity, the equations for building the 95% confidence intervals on the mean, gamma mean, and predicted response are reproduced in Table 9. This table shows that confidence intervals used to estimate the uncertainty of the gamma mean and the predicted response both incorporate the inverse dispersion parameter.

As described above, confidence intervals on the mean and predicted response were built for the predictive model shown in Table 9. In this case, the sample mean equal to 0.5 was used. A similar exercise as described above was performed for $\phi = 1$ (the theoretical value), $\phi = 2$ and $\phi = 3$, both assumed to be mis-estimated ($\mathrm{Var}(\eta)$ is assumed to be constant). The 95% percentile confidence intervals on the gamma mean (m) (for the given site) and the predicted response (y) were calculated using the equations illustrated in Table 9. The results are shown in Figs. 2 and 3, respectively.

Figs. 2 and 3 show that an unreliably estimated dispersion parameter can significantly affect the computation of confidence intervals. In Fig. 2, an inverse dispersion parameter that is mis-estimated by a factor of 100% (from 1 to 2) reduces the confidence interval by 19% for the upper bound. When the factor

Table 9
Ninety-five percent confidence and prediction intervals for Poisson-gamma models (Wood, 2005)

Parameter	Intervals
μ	$\left[\frac{\hat{\mu}}{e^{1.96\sqrt{\mathrm{Var}(\hat{\eta})}}}, \hat{\mu}\, e^{1.96\sqrt{\mathrm{Var}(\hat{\eta})}}\right]$
m	$\left[\max\left\{0, \hat{\mu} - 1.96\sqrt{\hat{\mu}^2\,\mathrm{Var}(\hat{\eta}) + \frac{\hat{\mu}^2\,\mathrm{Var}(\hat{\eta})+\hat{\mu}^2}{\phi}}\right\}, \hat{\mu} + 1.96\sqrt{\hat{\mu}^2\,\mathrm{Var}(\hat{\eta}) + \frac{\hat{\mu}^2\,\mathrm{Var}(\hat{\eta})+\hat{\mu}^2}{\phi}}\right]$
y	$\left[0, \left\lfloor \hat{\mu} + \sqrt{19}\sqrt{\hat{\mu}^2\,\mathrm{Var}(\hat{\eta}) + \frac{\hat{\mu}^2\,\mathrm{Var}(\hat{\eta})+\hat{\mu}^2}{\phi}} + \hat{\mu} \right\rfloor\right]$

Note: $\mathrm{Var}(\hat{\eta}) = \mathbf{x}_0(\mathbf{X}\mathbf{V}\mathbf{X}')^{-1}\mathbf{x}_0$, $\lfloor x \rfloor$ denotes the largest integer less or equal than x.

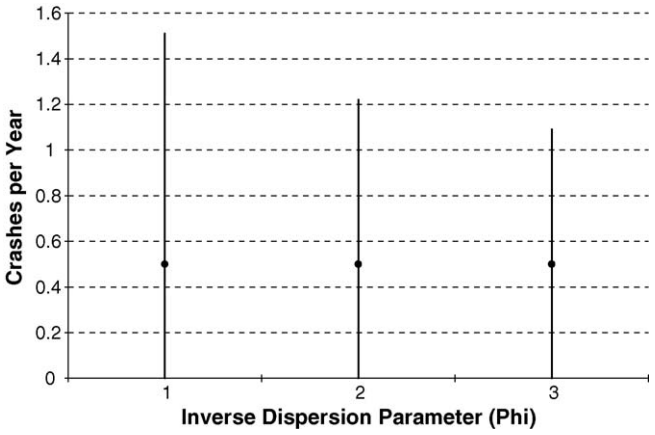

Fig. 2. Ninety-five percent confidence interval for the gamma mean (m) for a given site (theoretical value used for the simulation: $\phi = 1$).

is mis-estimated by a factor 200%, the confidence interval for the upper bound decreases by 27%. In Fig. 3, the confidence interval for the predicted response diminishes by 1 crash or 15% for a change in 100%. If the statistical model exhibits a pure erroneous Poisson characteristic (not shown in this graph), the confidence intervals of the predicted response could be off by as much as 60%. At that level, an unreliably estimated inverse dispersion parameter can have drastic consequences when Poisson-gamma models, in this case a pure Poisson model, are used for decision-making processes (e.g., selection of countermeasures or competitive highway design alternatives).

The last issue is related to the recalibration of predictive models. Up until now, the procedure for re-calibrating models did not include the recalibration of the inverse dispersion parameter. Given the fact that the mean structure of model will be different (note: usually only β_0 is re-calibrated; see Lord and Bonneson, 2005 for additional description on this assumption), it is expected that variance structure will also be different. Thus, the dispersion parameter should be re-calibrated using the new dataset. As detailed in Tables 5, 6 and 8, the estimators produced inconsistent values, particularly for small sample sizes. Consequently, the selection of the proper estimator is very critical and should be selected carefully by the safety modeler (more details on this topic below). In recent studies, FHWA (2003) recommends using Estimator 1 for applications related to the *SafetyAnalyst*, whereas Lord and Bonneson (2005) recommend Estimator 2 for re-calibrating the inverse dispersion parameter when models are transferred from one jurisdiction to another.

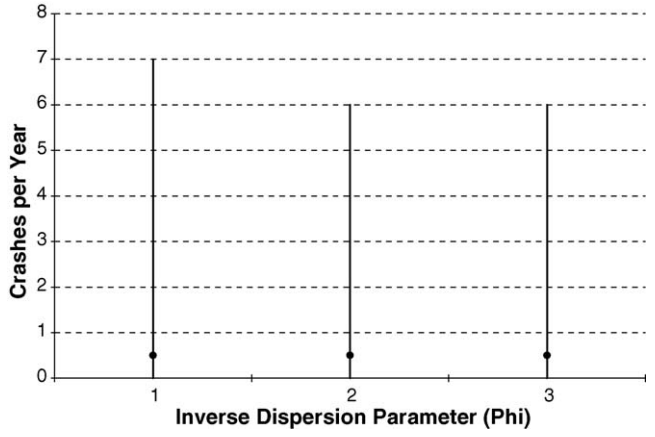

Fig. 3. Ninety-five percent confidence interval for the predicted response (y) (theoretical value used for the simulation: $\phi = 1$).

The discussion presented above leads to two important questions: (1) which estimator should be used for data characterized by a low sample mean and a small sample size and (2) what should be the minimum sample size a safety modeler should use to avoid or minimize an unreliably estimated dispersion parameter.

To answer the first question, one can examine the results of Tables 4 and 5. These tables show that on average Estimator 3 estimated values closer to the "theoretical" value used for the simulation runs more frequently. In addition, the standard deviation was usually smaller than the other two estimators. In the event Estimator 3 cannot be used, Estimator 2 should be used over Estimator 1. The number of instances where Estimator 2 was mis-estimated was much less frequent than Estimator 1. Nonetheless, for a mean below 0.5 and a sample size below 50, no estimator outperformed the others.

The second question can be answered using the simulation results shown in Tables 2 and 4. For $n = 1000$, $\lambda = 1.0$ and $\phi = 2$, all three estimators performed equally well. Assuming that the data exhibit proper asymptotic properties with the values noted above ($n \times \lambda$ or $1000 \times 1.0 = 1000$) (see Lawless, 1987), a matrix can be created to determine the recommended number of observations for different values describing the sample mean. Keeping the multiplied values fixed at 1000, the minimum sample size suggested for different values of λ is summarized in Table 10. (Note: the sample size refers to the number of observations or sites, e.g. intersections or segments, in the data. It does not reflect the number of collisions collected at all the sites.) Given the prohibitive costs to collect crash data and other related variables, the sample size recommended for very small sample means (0.5 and below) may be difficult to collect in practice. With the preponderance of evidence detailed in other fields of research on this topic, the author recommends that no Poisson-gamma models be estimated for a sample size below 100, even when the sample mean is equal to 5.

Table 10
Recommended minimum sample size[a] to minimize an unreliably estimated dispersion parameter

Population sample mean (λ)	Minimum sample size
5.00	200
4.00	250
3.00	335
2.00	500
1.00	1000
0.75	1335
0.50	2000
0.25	4000

[a] The sample size refers to the number of observations, e.g. intersections or segments, in the data. It does not reflect the number of collisions collected at the sites that are part of the sample.

There are a few avenues for further work. First, there is a need to determine if a small sample size and a low sample mean affects the estimation of the inverse dispersion parameter (posterior value) for Bayesian models that make use of Markov Chain Monte Carlo (MCMC) simulation techniques. The author did not find any work on this topic in the statistical literature. Given the recent interest in the development and application of hierarchical Bayes models in safety research, there is a need to determine whether they are as affected as for MLE models. Preliminary results seem to show that Poisson-gamma models developed using a Bayesian framework where the coefficients are estimated with WinBUGS (Spiegelhalter et al., 2003) suffer from the same limitations, but the mis-estimation of the inverse dispersion parameter (the posterior value) starts occurring at lower sample mean and smaller sample size values than the MLE (Lord and Miranda-Moreno, 2006). Second, further work should be performed on finding approaches to "correct" or "adjust" mis-estimated dispersion parameters. It may be possible to adjust the dispersion parameter given the characteristics of the data at hand. Finally, with the extensive work done on small sample sizes in the statistical literature, innovative estimation techniques specifically tailored for crash data should be evaluated (see Dean, 1994).

8. Summary and conclusions

The objectives of this study were to determine how the LMP combined with a small sample size can affect the estimation of the dispersion parameter and, consequently, how an unreliably estimated dispersion parameter can influence safety analyses that make direct use of the dispersion parameter or its inverse. This paper was motivated by the fact that, based on information obtained from the literature, crash databases used for developing statistical models are often characterized by low sample mean values and/or a small sample size. Previous research has shown that a low sample mean can significantly affect the quality-of-fit of statistical models. Given the importance the dispersion parameter plays in various types of safety studies, including the EB method, there is a need to determine the conditions in which the LMP affects the estimation of the dispersion parameter of Poisson-gamma models.

A series of Poisson-gamma distributions were simulated using different values describing the mean, the dispersion parameter, and the sample size. The simulation runs were estimated using a fixed and varying sample population means. Three estimators commonly used for estimating the dispersion parameter of Poisson-gamma models were evaluated: the method of moments, the weighted regression, and the maximum likelihood method. To complement the simulation study, the estimators were tested with crash data collected at unsignalized intersections in Toronto, Ontario.

Two main conclusions are drawn from this research. First, crash data characterized by a low sample mean combined with a small sample size can seriously affect the estimation of the dispersion parameter. All three estimators are affected equally for extreme conditions. The likelihood that the estimated dispersion parameter becomes mis-estimated, usually exhibiting erroneous pure Poisson characteristics, increases significantly as the sample size diminishes. An important problem with this characteristic is that the transportation safety modeler may not even be aware that the dispersion parameter is unreliably estimated.

Second, in the event the dispersion parameter is mis-estimated, common analyses performed in highway safety could be seriously undermined. For instance, the EB estimates as well as the estimation of confidence intervals for the gamma mean and predicted response could potentially be erroneous. Thus, the safety of road users could potentially be affected, e.g. selecting the wrong design alternative, if decisions are made using erroneous modeling output. In conclusion, there is a need for transportation safety modelers to carefully assess whether all the components of the statistical models, including the dispersion parameter or its inverse, are properly estimated.

Acknowledgements

The author would like to thank Prof. Graham Wood for providing useful comments on an earlier draft of this paper. The paper benefited from the input of TRB reviewers and two anonymous reviewers.

References

Abbess, C., Jarett, D., Wright, C.C., 1981. Accidents at blackspots: estimating the effectiveness of remedial treatment. With special reference to the "regression-to-mean" effect. Traffic Eng. Control 22 (10), 535–542.

Abdel-Aty, M., Addella, M.F., 2004. Linking roadway geometrics and real-time traffic characteristics to model daytime freeway crashes: generalized estimating equations for correlated data. Transport. Res. Rec. 1897, 106–115.

Anscombe, F.J., 1950. Sampling theory of the negative binomial and logarithmic series distributions. Biometrika 36, 358–382.

Cameron, A.C., 2005. Personal communication. College Station, TX.

Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data: comparisons and applications of some estimators and tests. J. Appl. Econom. 1, 29–53.

Cameron, A.C., Trivedi, P.K., 1990. Regression-based tests for overdispersion in the Poisson model. J. Econom. 46, 347–364.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, UK.

Clark, S.J., Perry, J.N., 1989. Estimation of the negative binomial parameter κ: by Maximum Quasi-Likelihood. Biometrics 45, 309–316.

Davidian, M., Carroll, R.J., 1987. A note on extended Quasi-likelihood. J. Roy. Stat. Soc.: Ser. B 82, 1079–1091.

Dean, C.B., 1994. Modified pseudo-likelihood estimator of the overdispersion parameter in Poisson mixture models. J. Appl. Stat. 21 (6), 523–532.

Federal Highway Administration, 2003. Module 1 – Network Screening. SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites. U.S. Department of Transportation, Washington, DC (web page: http://www.safetyanalyst.org/whitepapers/module1.pdf, accessed on June 10, 2005).

Fisher, R.A., 1941. The negative binomial distribution. Ann. Eugenics 11, 182–187.

Fletcher, R., 1970. A new approach to variable metric algorithms. Comput. J. 13, 317–322.

Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., Thomsen, L.K., 1995. Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. Accident Anal. Prev. 27 (1), 1–20.

Gourieroux, C., Monfort, A., Trognon, A., 1984a. Pseudo maximum likelihood methods: theory. Econometrika 52, 681–700.

Gourieroux, C., Monfort, A., Trognon, A., 1984b. Pseudo maximum likelihood methods: application. Econometrika 52, 701–720.

Gourieroux, C., Visser, M., 1986. A count data model with unobserved heterogeneity. J. Econom. 79, 247–268.

Hauer, E., 1996. Identification of sites with promise. Transport. Res. Rec. 1542, 54–60.

Hauer, E., 1997. Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety. Elsevier Science Ltd., Oxford.

Hauer, E., Persaud, B.P., 1984. Problem of identifying hazardous locations using accident data. Transport. Res. Rec. 975, 36–43.

Hauer, E., Council, F.M., Mohammedshah, Y., 2004. Safety models for urban four-lane undivided road segments. Transport. Res. Rec. 1897, 96–105.

Hauer, E., Ng, J.C.N., Lovell, J., 1988. Estimation of safety at signalized intersections. Transport. Res. Rec. 1185, 48–61.

Heydecker, B.G., Wu, J., 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference. Adv. Eng. Softw. 32, 859–869.

Ivan, J.N., Wang, C., Bernardo, N.R., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. Accident Anal. Prev. 32 (6), 787–795.

Kononov, J., Allery, B.K., 2004. Explicit consideration of safety in transportation planning and project scoping. Transport. Res. Rec. 1897, 116–125.

Kulmala, R., 1995. Safety at Rural Three- and Four-Arm Junctions: Development and Applications of Accident Prediction Models. VTT Publications 233, Technical Research Centre of Finland, Espoo.

Kumara, S.S.P., Chin, H.C., 1897. Study of fatal traffic accidents in Asia Pacific countries. Transport. Res. Rec., 43–47.

Kumara, S.S.P., Chin, H.C., Weerakoon, W.M.S.B., 2003. Identification of accident causal factors and prediction of hazardousness of intersection approaches. Transport. Res. Rec. 1840, 116–122.

Lawless, J.F., 1987. Negative binomial and mixed Poisson regression. Can. J. Stat. 15 (3), 209–225.

Lee, F., 1986. Specification test for Poisson regression models. Int. Econ. Rev. 27, 689–706.

Lord, D., 2000. The prediction of accidents on digital networks: characteristics and issues related to the application of accident prediction models. Ph.D. Dissertation. Department of Civil Engineering, University of Toronto, Toronto, Ontario.

Lord, D., Bonneson, J.A., 2005. Calibration of predictive models for estimating the safety of ramp design configurations. Transport. Res. Rec. 1908, 88–95.

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

Lord, D., Persaud, B.N., 2000. Accident prediction models with and without trend: application of the generalized estimating equations procedure. Transport. Res. Rec. 1717, 102–108.

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

Lord, D., Miranda-Moreno, L.F., 2006. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Dispersion Parameter of Poisson-gamma Models: A Bayesian Perspective. Working Paper. Zachry Department of Civil Engineering, Texas A&M University, College Station, TX.

Lyon, C., Oh, J., Persaud, B.N., Washington, S.P., Bared, J., 2003. Empirical investigation of the IHSDM accident prediction algorithm for rural intersections. Transport. Res. Rec. 1840, 78–86.

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

Maycock, G., Hall, R.D., 1984. Accidents at 4-arm roundabouts. TRRL Laboratory Report 1120. Transportation and Road Research Laboratory, Crowthorne, Bershire.

McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd ed. Chapman and Hall, London, UK.

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

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

Morris, C.N., 1997. Fitting Hierarchical Models. In: Proceedings of the Workshop on Statistics and Epidemiology: Environment and Health, Minneapolis, MN (web page: http://www.ima.umn.edu/summerstat/week6.html#wk6tue, accessed on January 3, 2006).

Myers, R.H., Montgomery, D.C., Vining, G.G., 2002. Generalized Linear Models: With Applications in Engineering and the Sciences. Wiley Publishing Co., New York, NY.

Noland, R.B., Quddus, M.A., 2004. Analysis of pedestrian and bicycle casualties with regional panel data. Transport. Res. Rec. 1897, 43–47.

Oh, J., Lyon, C., Washington, S.P., Persaud, B.N., Bared, J., 2003. Validation of the FHWA crash models for rural intersections: lessons learned. Transport. Res. Rec. 1840, 41–49.

Payne, R.W. (Ed.), 2000. The Guide to Genstat. Lawes Agricultural Trust, Rothamsted Experimental Station, Oxford, UK.

Persaud, B.N., Dzbik, L., 1993. Accident prediction models for freeways. Transport. Res. Rec. 1401, 55–60.

Persaud, B.N., Bahar, G., Mollett, C.J., Lyon, C., 2004. Safety evaluation of permanent raised snow-plowable pavement markers. Transport. Res. Rec. 1897, 148–163.

Persaud, B.N., Lord, D., Palminaso, J., 2002. Issues of calibration and transferability in developing accident prediction models for urban intersections. Transport. Res. Rec. 1784, 57–64.

Piegorsch, W.W., 1990. Maximum likelihood estimation for the negative binomial dispersion parameter. Biometrics 46, 863–867.

Pieters, E.P., Gates, C.E., Martin, J.H., Sterling, W.L., 1977. Small-sample comparison of different estimators of negative binomial parameters. Biometrics 33, 718–723.

Poch, M., Mannering, F.L., 1996. Negative binomial analysis of intersection-accident frequency. J. Transport. Eng. 122 (2), 105–113.

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

Ross, G.J.S., Preece, D.A., 1985. The negative binomial distribution. The Statistician 34, 323–661.

SAS Institute Inc., 2002. Version 9 of the SAS System for Windows. SAS Institute Inc., Cary, NC.

Shenton, L.R., Wallington, P.A., 1962. The bias of moment estimators with an application to the negative binomial. Biometrika, 193–204.

Spiegelhalter, D.J., Thomas, A., Best, N.G., Lun, D., 2003. WinBUGS Version 1.4.1 User Manual. MRC Biostatistics Unit, Cambridge. Available from http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.

Tarko, A.P., Kanodia, M., 2004. Effective and fair identification of hazardous locations. Transport. Res. Rec. 1897, 64–70.

Toft, N., Innocent, G.T., Mellor, D.J., Reid, S.W.J., 2006. The Gamma-Poisson model as a statistical method to determine if micro-organisms are randomly distributed in a food matrix. Food Microbiol. 23 (1), 90–94.

Walsh, G.R., 1975. Methods of Optimization. Wiley Publishing Co., London, UK.

Willson, L.J., Folks, J.L., Young, J.H., 1984a. Multistage estimation compared with sample-size estimation of the negative binomial κ. Biometrics 40, 109–117.

Willson, L.J., Folks, J.L., Young, J.H., 1984b. Complete sufficiency and maximum likelihood estimation for the two-parameter negative binomial distribution. Metrika 33, 349–362.

Wood, G.R., 2002. Generalised linear accident models and goodness of fit testing. Accident Anal. Prev. 34 (4), 417–427.

Wood, G.R., 2005. Confidence and prediction intervals for generalized linear accident models. Accident Anal. Prev. 37 (2), 267–273.