Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model

Peer-reviewed author version

Estimation of safety performance functions for urban intersections using various functional forms of the negative binomial regression model and a generalized Poisson regression model

ABSTRACT

Intersections are established dangerous entities of a highway system due to the challenging and unsafe roadway environment they are characterized for drivers and other road users. In efforts to improve safety, an enormous interest has been shown in developing statistical models for intersection crash prediction and explanation. The selection of an adequate form of the statistical model is of great importance for the accurate estimation of crash frequency and the correct identification of crash contributing factors. Using a six-year crash data, road infrastructure and geometric design data, and traffic flow data of urban intersections, we applied three different functional forms of negative binomial models (i.e., NB-1, NB-2, NB-P) and a generalized Poisson (GP) model to develop safety performance functions (SPF) by crash severity for signalized and unsignalized intersections. This paper presents the relationships found between the explanatory variables and the expected crash frequency. It reports the comparison of different models for total, injury & fatal, and property damage only crashes in order to obtain ones with the maximum estimation accuracy. The comparison of models was based on the goodness of fit and the prediction performance measures.

The fitted models showed that the traffic flow and several variables related to road infrastructure and geometric design significantly influence the intersection crash frequency. Further, the goodness of fit and the prediction performance measures revealed that the NB-P model outperformed other models in most crash severity levels for signalized intersections. For the unsignalized intersections, the GP model was the best performing model. When only the NB models were compared, the functional form NB-P performed better than the traditional NB-1 and, more specifically, the NB-2 models. In conclusion, our findings suggest a potential improvement in the estimation accuracy of the SPFs for urban intersections by applying the NB-P and GP models.

1. Introduction

Drivers encounter multiple interactions with turning and crossing vehicles, pedestrians, and cyclists at intersections. A plethora of information (e.g., the presence of road signs, street signs and name tags, traffic lights, channelization and road markings, conflicting, crossing and adjacent traffic movements, dedicated lanes for left and right turning vehicles, billboards and advert screens, and many others) at intersections produce an unsafe environment, which poses an enormous challenge for drivers to operate safely. The demand for instant decision-making, complex urban design, dense and rigorous land use, congestion, heavy traffic, vulnerable road users, and many on-and-off-vehicle distractions overload driver's attentional resources. That, in turn, leads to poor judgment of the traffic situation, confusion, inadequate decision, and ultimately a crash. It is not surprising to note that intersections constitute the highest proportion of total crashes on the roads. Tay (2015) has provided some statistics from around the world to highlight this safety concern. In the past, the operational aspects of urban intersections, such as optimization of traffic signals or reduction of vehicular and pedestrian traffic delays, travel time, and congestion, have received significant coverage in the literature (Dong et al., 2014; Roshandeh et al., 2014; Nesheli et al., 2009). However, these operational improvements do not account for the overall performance-based benefits (Roshandeh et al., 2016). The roadway network's overall performance requires consideration of additional aspects like safety, comfort, cost, availability, accessibility, etc. In this paper, we have focused on the safety of intersections in urban areas.

The safety of intersections can be improved by understanding the factors that contribute to the occurrence of crashes and thereby proposing appropriate countermeasures. Concerning this, an intersection safety analysis is typically suggested. One of the tools to measure the safety performance of intersections is by developing crash prediction models (CPMs). The CPMs are mathematical equations obtained through the statistical modeling of crash data and a series of explanatory variables. They are used to estimate the expected average crash frequency of roadway facilities over a specified period. The CPMs are also known as safety performance functions (SPFs) or collision prediction models (CPMs). The SPFs are applied to evaluate the safety of intersections and road segments, identify hazardous locations, assess the safety of applied solutions, and compare and prioritize the best

28   alternative designs (AASHTO, 2010). To address safety issues, the SPFs have been developed for many

29   years now across the globe for numerous highway facilities (Elvik et al., 2019; Abdel-Aty et al., 2016;

30   Janstrup, 2016; Cafiso et al., 2012; Persaud et al., 2012; Vieira Gomes et al., 2012; Srinivasan and

31   Carter, 2011; Wong et al., 2007; Greibe, 2003). Leaving aside the applicability of these models, the

32   development of the SPFs is a critical process in which a modeler makes crucial decisions. To emphasize,

33   Hauer and Bamfo (1997) argued, "*In the course of modeling, the modeler will make two major*

34   *decisions: (a) What explanatory variables to include in the model equation; and, (b) What should be its*

35   *functional form*." Factors, such as the SPF's purpose, the availability, quality, and quantity of the data,

36   and required expertise affect those decisions.

37   American Association of State Highways and Transportation Officials (AASHTO) published

38   the Highway Safety Manual (HSM), first in 2010 (AASHTO, 2010), and then in 2014 with a few

39   supplements (AASHTO, 2014). The HSM offers the SPFs to predict intersection and road segment

40   crashes on several highway facility types, such as rural two-lane and multilane highways, urban and

41   suburban arterials, and freeway ramp terminals (AASHTO, 2014; AASHTO, 2010). The predictive

42   models in the HSM were developed using data from a small number of States. Because of the possible

43   differences in travel behavior, traffic conditions, and road characteristics across different geographical

44   regions, it has been highlighted that the crash relationships in these states may not be necessarily

45   representative of those in the other states. Regarding this, the guidelines in the HSM recommend either

46   (i) the calibration of its base models for applications in other jurisdictions or (ii) the estimation of new

47   SPFs for the regions where sufficient good quality local data is available. Therefore, several states in

48   the US and other countries have developed their own SPFs. The SPFs given in the HSM for intersections

49   estimate only total crashes that might not be an ideal approach since crashes vary by type and severity

50   across intersections (Wang et al., 2019; Zhao et al., 2018; Wang et al., 2017). Some intersections might

51   be crowded by fatal crashes only, and others might experience injury or property damage only (PDO)

52   crashes. Similarly, some intersections could have a higher proportion of a different particular type of

53   crash compared to other intersections. Differences in the distribution of crash severity and/or crash type

54   could be attributed to the variation in the geometric design and traffic characteristics between

55   intersections. In order to consider those variations, studies estimate predictive models for intersections

56   by crash type (Wang et al., 2019; Gates et al., 2018; Liu and Sharma, 2018; Wu et al., 2018; Dixon et

57   al., 2015; Geedipally and Lord, 2010), and/or by severity level (Liu and Sharma, 2018; Wang et al.,

58   2017; Wu et al., 2013; Oh et al., 2010).

59   Regarding the statistical methodologies, crash prediction modeling has come a long way. In the

60   beginning, researchers used linear regression models for the estimation of crashes and determining

61   relationships between crash frequency and explanatory variables (Joshua and Garber, 1990; Okamoto

62   and Koshi, 1989). However, with new research, it was soon realized that linear regression models have

63   certain limitations in treating the non-negative and discrete nature crash data (Lord and Mannering,

64   2010; Miaou and Lum, 1993). This led to the adoption of count data models in crash prediction.

65   Naturally, the first choice of researchers was the Poisson regression model, which assumes that the data

66   variance is equal to its mean. On the other hand, the crash data is frequently characterized by over-

67   dispersion, which occurs when the crash data variance is greater than its mean. To overcome the over-

68   dispersion issue, negative binomial (NB) regression models were used (Abdel-Aty & Radwan, 2000;

69   Miaou, 1994). With the progress in statistical methods and improved computing power, more advanced

70   techniques have been applied recently to model the crash data. Lord and Mannering (2010), and

71   Mannering and Bhat (2014) have provided detailed accounts of the existing trends in the crash

72   prediction and future directions. Despite all the intricacy, the traditional NB model still enjoys great

73   popularity due to its inherent simplicity of estimation and relatively better performance.

74   Several parameterizations of the NB model are available in the literature. Nonetheless, the NB-

75   1 and NB-2 (Cameron and Trivedi, 1986) have been commonly used to model the count data (Wang et

76   al., 2019; Giuffrè et al., 2014; Ismail and Zamani, 2013; Hilbe, 2011; Winkelmann, 2008; Chang and

77   Xiang, 2003; Miaou and Lord, 2003). The two models necessarily differentiate based on the relationship

78   between the variance of the data and the mean of the data. The NB-1 assumes a linear relationship

79   between the variance and the mean, while the NB-2 assumes a quadratic relationship. Detailed

80   estimation procedures of the two alternative forms are given in Hardin (2018), Lord and Park (2015),

81   and Hilbe (2011). In traffic safety, the NB-2 has been frequently used to estimate the SPFs, while the

82 NB-1 has been applied in a few studies. For instance, Chang and Xiang (2003) created SPFs using both

83 the NB-1 and NB-2 models to study the relationship between crashes and congestion levels on freeways.

84 The authors found that both models showed consistent results for the relationship between crashes and

85 traffic volume, the number of through lanes, and median. Giuffrè et al. (2014) applied the NB-1 and

86 NB-2 models to develop the SPFs for urban unsignalized intersections. They found that the NB-1 fits

87 the data better than the NB-2. Wang et al. (2019) used the NB-1 model with the standard Poisson, NB-

88 2, and NB-P models to estimate the SPFs and select a better performing model for rural two-lane

89 intersections.

90   The applications of the NB-1 and NB-2 models, however, come with a few compromises. For

91 instance, the NB-1 and NB-2 models restrict the variance structure in estimating the SPFs (Park, 2010).

92 In other words, the mean-variance relationship of the crash data is constrained to either a linear or

93 quadratic link for the NB-1 and NB-2 models, respectively. The restricted variance structure may result

94 in biased estimates of model parameters and, ultimately, incorrect crash forecasts (Wang et al., 2019).

95 Furthermore, both the NB-1 and NB-2 are non-nested models, and an appropriate statistical test to

96 determine a better model of the two cannot be carried out directly (Wang et al., 2019; Greene, 2008).

97 To account for that, Greene (2008) introduced a new functional form of the NB regression called an

98 NB-P that nests both the NB-1 and NB-2 models. The NB-P is essentially the extension of the traditional

99 NB models to address the restricted variance structure problem. The NB-P reduces to NB-1 when P=1

100 and to NB-2 when P=2. Since the NB-P model parametrically nests both the NB-1 and NB-2 models,

101 it allows analysts to test the two NB functional forms (NB-1, NB-2) against a more general alternative

102 (NB-P) for a better model (Greene, 2008; Ismail and Zamani, 2013; Hilbe, 2011). The NB-P model has

103 been used in a few studies dealing with count data. For example, Greene (2008) applied the NB-P with

104 the NB-1 and NB-2 models to the German health care data and found that the NB-P outperformed the

105 other two models based on the goodness of fit measures. Ismail and Zamani (2013) used the NB-1, NB-

106 2, and NB-P models to study the Malaysian private car own damage claim counts. They also reported

107 that the best performing model was the NB-P model. In traffic safety, Wang et al. (2019) used the NB-

108 P model with the standard Poisson, NB-1, and NB-2 models to study rural two-lane intersections' safety

109   performance by crash type and intersection type. They developed traffic only models. Their findings

110   revealed that the NB-P model performed better than the Poisson model, NB-1, and NB-2 models for

111   most crash types and intersection types. They concluded that the flexible variance structure of the NB-

112   P model significantly improves the estimation accuracy. Recently, Wang et al. (2020) applied the NB-

113   P model to examine various functional forms of intersection safety performance functions in an urban

114   and suburban context.

115   The literature review shows that the NB-P model, despite the apparent improvement in

116   estimation accuracy compared to the traditional NB models, are still not commonly applied in traffic

117   safety and crash prediction. To the authors' knowledge, only Wang et al. (2020) have used the NB-P

118   model to estimate SPFs for urban roads. However, there has been no evidence that the NB-P model is

119   used to estimate multivariate SPFs. Given that the applications of the NB-P model in road safety are

120   rare, its potential to improve the estimation accuracy by offering a flexible variance structure, and that

121   it allows to statistically test the NB-1 and NB-2 models against a general alternative, are motivations

122   behind this work. Besides, the HSM recommendation of developing local SPFs for locations with

123   enough data was another driving force. In this paper, we applied different functional forms of the NB

124   regression model (NB-1, NB-2, and NB-P) and compared the results with the Generalized Poisson (GP)

125   regression model, also a popular count data modeling technique, in the pursuit of obtaining the best

126   model for the estimation of intersection SPFs in the urban areas. The GP model, discussed in detail in

127   section 2.4, is an extension of the Generalized NB models (Ismail and Zamani, 2013). In the past, the

128   GP models have been applied to study road crashes (Famoye et al., 2004), shipping damage incidents

129   (Ismail and Jemain, 2007), vehicle insurance claims (Ismail and Zamani, 2013), etc. The rationale for

130   choosing the GP model to compare with the NB models was that it could also accommodate the over-

131   dispersed data equally well, has relatively fewer applications in the SPF estimation, and the fact that it

132   is regarded as a potential competitor to the NB models for treatment of over-dispersed count data

133   (Melliana et al., 2013). The contribution of the current study to traffic safety literature is that it applies

134   the functional form NB-P of the NB regression, along with the NB-1, NB-2 and a GP model for the

135   estimation of intersection SPFs in the urban areas. A unique combination of the new approach for the

136 SPFs estimation and the use of not only the traffic flow but also other explanatory variables adds to the

137 novelty of this work. To the best of our knowledge, no micro-level SPFs have been developed for the

138 urban intersections in Belgium. Results of this study could potentially serve the local research

139 community involved in traffic safety as well as the industry in planning level safety assessment of new

140 road infrastructure projects.

141 **2. Methodology**

142 The count data models have been widely applied to estimate crashes at road segments and

143 intersections in a non-negative, discrete, and random fashion (Washington et al., 2010). Since the

144 Poisson regression model is usually not fit for modeling crash data due to its inability to accommodate

145 overdispersion, three different functional forms of the NB model and a GP model were applied to

146 estimate the SPFs in this study.

147 *2.1 Negative binomial model-type 2 (NB-2)*

148 The negative binomial regression is the derivative of the standard Poisson regression. It

149 redefines the conditional mean (i.e., the data variance equals its mean) of the standard Poisson model

150 and incorporates a latent heterogeneity term to account for over-dispersion in data. The expected crash

151 frequency "μ_i" at the intersection "i" obtained by applying the NB model as in Washington et al. (2010)

152 is given by:

$$\mu_i = exp(\beta X_i + \varepsilon_i) \tag{1}$$

153 where "X_i" is the vector of explanatory variables, "β" is the vector of estimable coefficients,

154 and "$exp(\varepsilon_i)$" is the latent heterogeneity term, also known as an error term. When "$exp(\varepsilon_i)$" follows a

155 gamma distribution with mean 1 and variance $1/\sigma = k$ where "k" represents an over-dispersion

156 parameter, a traditional NB model, called the NB-2 model, is derived.

157 For the interest of readers, equation 1, according to the standard Poisson regression model,

158 would have been:

$$\mu_i = exp(\beta X_i) \tag{2}$$

159 As can be seen, **equation 2** lacks the term "$exp(\varepsilon_i)$" to account for over-dispersion.

9

160   The probability density function of the NB-2 model for estimation of the SPFs as in Washington

161   et al. (2010):

$$Prob[y_i|\mu_i] = \frac{\Gamma[(\sigma) + y_i]}{\Gamma(\sigma)y_i!} \left[\frac{\sigma}{(\sigma) + \mu_i}\right]^{\sigma} \left[\frac{\mu_i}{(\sigma) + \mu_i}\right]^{y_i} \qquad (3)$$

162   where Γ is a gamma function. The mean and variance of the NB-2 regression model are equal

163   to $E(y_i) = \mu_i$ and $Var(y_i) = \mu_i + k\mu_i^2 = \mu_i(1 + k\mu_i)$, respectively. When $1/\sigma = k$, the marginal

164   distribution function of the NB-2 model can be reproduced:

$$Prob[y_i|\mu_i] = \frac{\Gamma\left[\left(\frac{1}{k}\right) + y_i\right]}{\Gamma\left(\frac{1}{k}\right)y_i!} \left[\frac{\frac{1}{k}}{\left(\frac{1}{k}\right) + \mu_i}\right]^{\frac{1}{k}} \left[\frac{\mu_i}{\left(\frac{1}{k}\right) + \mu_i}\right]^{y_i} \qquad (4)$$

165   *2.2 Negative binomial model-type 1 (NB-1)*

166   A re-parameterization of the variance structure of the NB model by replacing $\frac{1}{k}$ in the NB-2

167   (**equation 4**) with $\frac{1}{k}\mu_i$ allows for another functional form, called the NB-1 (Wang et al., 2019; Hilbe,

168   2011; Greene, 2008; Cameron & Trivedi, 1986). The marginal distribution function of the NB-1 is given

169   by:

$$Prob[y_i|\mu_i] = \frac{\Gamma\left[\left(\frac{1}{k}\mu_i\right) + y_i\right]}{\Gamma\left(\frac{1}{k}\mu_i\right)y_i!} \left[\frac{\frac{1}{k}\mu_i}{\left(\frac{1}{k}\mu_i\right) + \mu_i}\right]^{\frac{1}{k}\mu_i} \left[\frac{\mu_i}{\left(\frac{1}{k}\mu_i\right) + \mu_i}\right]^{y_i} \qquad (5)$$

170   The mean of the NB-1 is $E(y_i) = \mu_i$ and the variance of the NB-1 is $Var(y_i) = \mu_i + k\mu_i$.

171   *2.3 Negative binomial model-type P (NB-P)*

172   Greene (2008) proposed a new form of the NB regression that uses the parameter "P" to

173   represent the mean-variance relationship. It is known as the NB-P model. The NB-P model is obtained

174   by replacing $\frac{1}{k}$ in the NB-2 model (**equation 4**) with $\frac{1}{k}\mu_i^{2-P}$. The marginal distribution function of the

175   NB-P model is given by:

$$Prob[y_i|\mu_i] = \frac{\Gamma\left[\left(\frac{1}{k}\mu_i^{2-P}\right) + y_i\right]}{\Gamma\left(\frac{1}{k}\mu_i^{2-P}\right)y_i!} \left[\frac{\frac{1}{k}\mu_i^{2-P}}{\left(\frac{1}{k}\mu_i^{2-P}\right) + \mu_i}\right]^{\frac{1}{k}\mu_i^{2-P}} \left[\frac{\mu_i}{\left(\frac{1}{k}\mu_i^{2-P}\right) + \mu_i}\right]^{y_i} \qquad (6)$$

10

176   where mean and variance of the NB-P are $E(y_i) = \mu_i$ and $Var(y_i) = \mu_i + k\mu_i^P$, respectively.

177 "P" represents the functional parameter of the NB-P model.

178   All the NB models used the maximum likelihood estimation (MLE) approach to estimate ~~the~~

179 parameter coefficients.

180 *2.4 Generalized Poisson model (GP)*

181   The generalized Poisson (GP) regression is another popular approach to model count data. As

182 an alternative to the NB regression, the GP models have the advantage of modeling both over-dispersed

183 and under-dispersed data. Like the NB regression, the GP model has an extra parameter, called a scale

184 or dispersion parameter. A distinctive feature of the GP dispersion parameter is that it can take both

185 positive and negative values for over-dispersed and under-dispersed data, respectively. The probability

186 mass function (p.m.f.) of the GP distribution given as in Yang et al. (2009):

$$Prob\ (Y_i = y_i) = \frac{\theta(\theta + ky_i)^{y_i - 1}\exp(-\theta - ky_i)}{y_i!}, \qquad y_i = 0,1,2, \ldots\ , \qquad (7)$$

187   where $\theta > 0$, and $0 \leq k < 1$. From Joe and Zhu (2005), the mean of the GP regression is

188 $E(Y_i) = \mu = (1 - k)^{-1}\theta$, and the variance of the GP regression is $Var(Y_i) = (1 - k)^{-3}\theta =$

189 $(1 - k)^{-2}\mu = \emptyset.\mu$. The term $\emptyset = (1 - k)^{-2}$ is a dispersion factor, and it is used in the GP mass

190 function where "k" is a dispersion parameter. It can be seen that when $k = 0$, a standard Poisson model

191 is obtained. For $k < 0$, under-dispersion is assumed while $k > 0$ represents over-dispersion. Since crash

192 data normally exhibits over-dispersion, this study will assume $k > 0$ condition. There are other

193 parametrizations of the GP, but their applications are left for future studies.

194 *2.5 Model structure*

195   The literature offers several ways to model the relationships between intersection crash

196 frequency and explanatory variables (Barbosa et al., 2014; Park and Lord, 2009; Nambuusi et al., 2008;

197 Miaou and Lord, 2003). They are differentiated based on the type of variables, the number of variables,

198 the form that the variables take during the modeling process and the transformation applied to the

199 variables (Oh et al., 2003). In this study, the following model structure was used to estimate the expected

200 crash frequency "μ_i" of the intersection "i":

$$\mu_i = exp(\beta_0 + \beta_1 \ln(AADT_{major}) + \beta_2 \ln(AADT_{minor}) + \Sigma_{m=3}^{n}\beta_m X_m) \qquad (8)$$

where β_0 represents the intercept, $AADT_{major}$ is the major approach average annual daily traffic (AADT), β_1 represents the coefficient estimate of the major approach AADT, $AADT_{minor}$ represents the minor approach AADT, β_2 represents the coefficient estimate of the minor approach AADT, β_m is the vector of the coefficient estimates of explanatory variables and "X_m" denotes the vector of explanatory variables. For the NB models (NB-1, NB-2, and NB-P) and GP model, the coefficients denoted by β_m and a dispersion parameter denoted by "k" were estimated, but for the NB-P model, an additional parameter "P", known as the functional parameter, was also estimated.

2.6 Model comparison

For model comparison, both the likelihood-based and the predictive ability-based measures were used. The likelihood-based measures consisted of the likelihood ratio test (LRT), the Akaike Information Criteria (AIC), and the Bayesian Information Criteria (BIC). The LRT was used when comparing the hierarchically nested models only (Greene, 2008; Wang et al., 2019). The AIC and the BIC were used for comparing the non-nested models (Ismail and Jemain, 2007).

The predictive ability-based measures compared all developed models for predictive performance using the validation data. Those included in the study were; mean prediction bias (MPB), mean absolute deviation (MAD), and mean squared prediction error (MSPE) as in Oh et al. (2003), and % CURE deviation and a validation factor (Hauer, 2015; Wang et al., 2019). To measure the degree to which our models were well-calibrated and to assess the certainty level of crash predictions by our models, we also estimated a range of prediction intervals (50%, 75%, 90%, and 95%) for each intersection type using the method proposed by Wood (2005), and then compared our models using those intervals.

3. Data

The data used for modelling was obtained for urban intersections of Antwerp, Belgium. A dataset consisting of crash data of six years (2010-2015), road geometry data, and traffic flow data was created to estimate the SPFs. An online database of the regional government called the Flanders road

12

226 register was consulted for the intersection data. A total of 760 intersections were used for analysis, of

227 which 198 were signalized, and 562 were unsignalized. Around 470 were three-legged intersections,

228 and the remaining 290 were four-legged intersections.

229 Table 1 Variables description for urban intersections of Antwerp

Variable Description	Variable levels
AADT on the major approach	-
AADT on the minor approach	-
Skewness	1: Intersection angle is less than/equal to 75-degrees 2: Intersection angle is greater than 75-degrees
Legs/approaches of the intersection	1: For 4 legged intersections 0: For 3 legged intersections
Existence of stop sign on the minor approach	1: Stop sign is present on at least one minor approach 0: No stop sign on the minor approaches
Existence of stop line on the minor approach	1: Stop line is present on at least one minor approach 0: No stop line on the minor approaches
Number of left turn lane on the major approach	2: At least one left turn lane exists on each direction of the major approach 1: At least one left turn lane exists on only one direction of the major approach 0: No left turn lane exists
Number of right turn lane on the major approach	2: At least one right turn lane exists on each direction of the major approach 1: At least one right turn lane exists on only one direction of the major approach 0: No right turn lane exists
Number of through lanes of the minor approach	4 or 4+: Four and more through lanes of the minor approach 1-3: One to three through lanes of the minor approach 0: No through lane of the minor approach
Left turn (LT) movements on the minor approach	2: LT movement on each minor approach 1: LT movement on only one minor approach 0: No LT movement on the minor approach
Existence of crosswalk on minor approach	2: Crosswalk on each minor approach 1: Crosswalk on only one minor approach 0: No crosswalk
Existence of crosswalk on major approach	2: Crosswalk on each major approach 1: Crosswalk on only one major approach 0: No crosswalk
Size of the intersection [a]	4: for 5*4, 5*8, 6*4, 6*6, 6*8, 8*4, 8*6, 8*8, 8*10, 10*8, 10*10 3: for 3*2, 3*4, 3*6, 4*2, 4*4, 4*6 2: for 2*2, 2*3, 2*4, 2*6 1: for 1*2, 1*3, 1*4

[a] The first number is the total number of approach lanes for a minor approach, and the second number is the total number of through lanes for a major approach (as per, Abdel-Aty and Haleem 2011)

230   Because the skewness of intersection has been reported to impact its safety (Kumfer et al., 2019;

231 Nightingale et al., 2017; Haleem and Abdel-Aty, 2010), we decided to include skewness as a potential

232 explanatory variable. The smallest angle between the two adjacent approaches of the intersection,

233 known as an intersection angle (Nightingale et al., 2017), was used as a surrogate to define the level of

234 skewness. A 75 degrees intersection angle used by Haleem and Abdel-Aty (2010) was chosen as a

235 threshold to define the levels of skewness. An intersection angle less than or equal to 75 degrees

236 represented skewness level 1, while an intersection angle greater than 75 degrees represented skewness

237 level 2. Two hundred and seventeen intersections had a skewness level 1, and five hundred and forty-

238 three intersections had a skewness level 2. **Table 1** describes the variables employed in this study for

239 urban signalized and unsignalized intersections.

240   The police of Antwerp provided the crash data. The crash records featured the severity level of

241 a crash, coordinates of a crash location, time and date of a crash, number of the vehicles involved and

242 their type, maneuver of the involved vehicles at the time of the crash, data about the involved drivers,

243 and road and pavement conditions. Only intersection and intersection-related crashes were used in the

244 analysis. Because of the inconstancy in the influence area's definition to classify a crash as intersection-

245 related (Wang et al., 2008), we chose to use the HSM guidelines to differentiate the intersection and

246 intersection-related crashes from segment crashes. According to the HSM (AASHTO, 2014, 2010);

247   -   An intersection crash is one that has occurred within the physical boundaries of an

248   intersection area

249   -   An intersection related crash is one that has occurred on the road segment but the presence

250   of the intersection was the cause of that crash, and it falls within its influence area

251   Using the above definition, 5128 intersection and intersection-related crashes were identified

252 for analysis (**Table 2**). To account for the potential variation in the SPFs by crash severity, those crashes

253 were divided into total crashes, injury & fatal crashes and property damage only (PDO) crashes. In the

254 beginning, we had plans to disaggregate crash data further by type and developed separate models for

255 various crash types, but the partial availability of the information in our dataset and the resulting smaller

256 sample size did not allow for the SPF estimation by crash type.

257   Table 2 Descriptive statistics of crash data (by severity) and traffic flow data for signalized and unsignalized intersections

Variables	Signalized Intersections				Unsignalized Intersections			
	Min.	Max.	Aver.	Std. Dev.	Min.	Max.	Aver.	Std. Dev.
Total Crashes	0	87	13.899	13.848	0	51	4.347	5.223
PDO Crashes	0	50	6.979	7.760	0	49	2.540	3.671
Injury & Fatal Crashes	0	39	6.919	7.224	0	25	1.806	2.557
Ln (AADT)$_{major}$	183	41915	14559	9424.8	13	30648	3511	2884.1
Ln (AADT)$_{minor}$	31	26837	5225	4905.8	9	7595	1001	815.2

258

259   The traffic data was acquired from Lantis, a mobility management company based in Antwerp.

260 Lantis also provides its services to the Mobiliteit en Parkeren Antwerpen Ag, an office for Antwerp

261 city's parking and mobility services. The data was received in two sets, actual counts and traffic model

262 estimates. The actual counts were collected using either manual counting techniques or loop detectors

263 installed at random locations on the roads in the studied network. The traffic model estimates were

264 generated using a microsimulation traffic model called Dynamisch Model Kernstad Antwerpen

265 (DMKA). It is important to note that the model was calibrated for 2010-2015, when the crash data was

266 recorded. Results from several runs of the simulation model were obtained and averaged to get a better

267 convergence towards the actual counts. Actual counts and model generated counts were compared at

268 locations where both were available to check for the residuals. An absolute difference of not greater

269 than 5% between the simulation counts and actual counts was reported for most locations. The outliers

270 were discarded. The authors agreed to use a combination of actual counts and traffic model estimates

271 to ensure as many intersections included in the SPFs estimation as possible with a maximum degree of

272 accuracy.

273 **4. Results**

274   **Table 3** and **Table 4** present the parameter estimates (β) of the NB-1, NB-2, NB-P, and GP

275 models developed by crash severity (total crashes, PDO crashes, and injury & fatal crashes) for

276 signalized and unsignalized intersections, respectively. The numbers enclosed within the parenthesis

277 correspond to their p-values. The SPFs show that the signs of estimated parameters are similar across

278 different models developed for the same severity level. This indicates that given the same severity level,

279 the potential impact of explanatory variables on the expected crash frequency obtained from different

280 models is similar. However, the estimated parameters vary slightly across different severity levels,

281 which could be one reason that implies the need to develop separate models for each crash severity.

282 Using a 90% confidence level as in Vieira Gomes et al. (2012) for similar data, we found that five

283 variables were significant in the models for signalized intersections and four variables in the models for

284 unsignalized intersections. The significant variables included the traffic flow, intersection skewness,

285 existence of crosswalk on a minor approach, the number of through lanes on a minor approach, and the

286 number of approaches. To our surprise, the presence of exclusive left and right turn lanes were not

287 significant in any model. The intersection size and crosswalk on the major approaches were other

288 insignificant explanatory variables.

289 *4.1 SPFs of signalized intersections*

290 **Table 3** provides the SPFs for the signalized intersections. It shows a statistically significant

291 increase in the crash frequency with an increase in the natural logarithm of AADTs (which necessarily

292 indicates an increase in traffic flow) of the major and the minor approaches of the intersection. The

293 crosswalk on a minor approach was significant only when it existed on both approaches of a signalized

294 intersection across all developed models and all severity levels. However, there was an exception in the

295 case of the NB-2 and NB-P models of total crashes, in which, in addition to a crosswalk on each minor

296 approach, a crosswalk variable was also significant when present on only one of the minor approaches

297 of an intersection. The estimated coefficients in the former case were approximately double that of the

298 latter. This was not true for other crash severity levels (i.e., PDO, and injury & fatal crashes) and model

299 types. The intersection skewness was significant for total crashes (all the NB models only), and injury

300 & fatal crashes (all models) but not for the PDO crashes. The coefficient estimates were negative in the

301 developed models. Since the high skewness level was a base case, the negative sign indicates that no

302 skewness or low skewness level (i.e., intersection angle greater than 75 degrees, please see section 3

303 for details) results in a reduced crash frequency. In other words, intersections with no or low skewness

304 were safer than intersections with high skewness. An important observation from the results was that

16

305   the absence of skewness causes a more significant decrease in injury & fatal crashes than total crash

306   frequency.

307   Table 3 Estimated models for urban signalized intersections

	NB-1	NB-2	NB-P	GP
Variables	β (p-value)	β (p-value)	β (p-value)	β (p-value)
TOTAL CRASHES				
Intercept	-3.9067 (0.0000)	-4.2775 (0.0000)	-4.2761 (0.0000)	-3.7402 (0.0000)
AADT $_{Major}$	0.4058 (0.0000)	0.3623 (0.0000)	0.3621 (0.0000)	0.3934 (0.0000)
AADT $_{Minor}$	0.2450 (0.0000)	0.3002 (0.0000)	0.3003 (0.0000)	0.2425 (0.0000)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	N/S [a]	0.5547 (0.0690)	0.5551 (0.0690)	N/S
Crosswalk on each of the minor approach: 2	0.8867 (0.0050)	1.2151 (0.0000)	1.2155 (0.0000)	0.8549 (0.0040)
Skewness: 1 (Base)				
Skewness: 2	-0.1572 (0.0970)	-0.2180 (0.0360)	-0.2181 (0.0360)	N/S
Over-dispersion	4.1062	0.2977	0.2953	0.5778
P	1.000 (0.0000)	2.000 (0.0000)	2.0031 (0.0000)	
Log L [ba]	-653.03	-640.65	-640.65	-651.79
AIC	1320.06	1295.31	1297.31	1317.58
BIC	1343.07	1318.33	1323.62	1340.60
PDO CRASHES				
Intercept	-4.4085 (0.0000)	-4.8088 (0.0000)	-4.8899 (0.0000)	-4.2727 (0.0000)
AADT $_{Major}$	0.3396 (0.0100)	0.2992 (0.0010)	0.3153 (0.0010)	0.3269 (0.0010)
AADT $_{Minor}$	0.2954 (0.0000)	0.3367 (0.0000)	0.3357 (0.0000)	0.2942 (0.0000)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	N/S	N/S	N/S	N/S
Crosswalk on each of the minor approaches: 2	0.9062 (0.0150)	1.3397 (0.0010)	1.2820 (0.0020)	0.9008 (0.0130)
Skewness: 1 (Base)				
Skewness: 2	N/S	N/S	N/S	N/S
Over-dispersion	2.7650	0.3840	0.6319	0.5022
P	1.000 (0.0000)	2.000 (0.0000)	1.7530 (0.0000)	
Log L	-538.89	-533.32	-532.91	-538.16
AIC	1091.79	1080.65	1081.81	1090.33
BIC	1114.80	1103.66	1108.12	1113.35
INJURY & FATAL CRASHES				
Intercept	-4.9921 (0.0000)	-5.6066 (0.0000)	-5.6210 (0.0000)	-4.9458 (0.0000)
AADT $_{Major}$	0.4963 (0.0000)	0.4797 (0.0000)	0.4835 (0.0000)	0.4952 (0.0000)
AADT $_{Minor}$	0.1917 (0.0030)	0.2586 (0.0000)	0.2563 (0.0000)	0.1879 (0.0040)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	N/S	N/S	N/S	N/S
Crosswalk on each of the minor approaches: 2	0.8633 (0.0150)	1.0876 (0.0030)	1.0856 (0.0040)	0.8576 (0.0140)

Skewness: 1 (Base)				
Skewness: 2	-0.2056 (0.0620)	-0.3258 (0.0070)	-0.3213 (0.0090)	-0.1984 (0.0740)
Over-dispersion	2.3591	0.3324	0.3679	0.4727
P	1.000 (0.0000)	2.000 (0.0000)	1.9500 (0.0000)	
Log L	-531.95	524.40	-524.39	-530.87
AIC	1077.91	1062.81	1064.77	1075.74
BIC	1100.93	1085.83	1091.08	1098.76

308   *Notes: [a] Not Significant, [b] Log-Likelihood*

309   *4.2 SPFs of unsignalized intersections*

310   **Table 4** presents the coefficient estimates of the SPFs for unsignalized intersections. The traffic

311 flows of major and minor approaches were significantly associated with crash frequency except for

312 injury & fatal crashes, where the AADT of the minor approach was found insignificant. The presence

313 of a crosswalk on the minor approach was only significant for total crashes, and injury & fatal crashes

314 across all developed models. Unlike signalized intersections, the crosswalk was significant when it was

315 present on only one of the minor approaches of unsignalized intersections. However, crosswalk

316 presence on one or both approaches was significant only in injury & fatal crashes, as can be seen in the

317 NB-2 and NB-P models. The number of approaches/legs of an intersection was a significant predictor

318 of the total, and PDO crashes at unsignalized intersections at a 90% confidence level. Intersections with

319 three approaches/legs as a base level, the positive signs of the estimated coefficients show higher

320 expected crash frequency on intersections with four approaches compared to intersections with three

321 approaches. Another statistically significant variable was the number of through lanes of the minor

322 approaches of unsignalized intersection. A positive association was found between crash frequency and

323 the number of through lanes of its minor approach for the total crashes, and injury & fatal crashes.

324 While the first level of this variable was not significant, the second level, which represents four or more

325 through lanes of minor approaches, was significant for total crashes. For injury & fatal crashes, all levels

326 of the number of through lanes were significant. This means that a significant increase can be expected

327 in total crashes, and injury & fatal crashes with an increase in the number of through lanes of the minor

328 approach of an unsignalized intersection. It is noteworthy that this result can be generalized only to

329 four-legged unsignalized intersections because through lanes were reported only for such facility type

330 in this study.

Table 4 Estimated models for urban unsignalized intersections

Variables	NB-1 β (p-value)	NB-2 β (p-value)	NB-P β (p-value)	GP β (p-value)
TOTAL CRASHES				
Intercept	-1.2095 (0.0000)	-1.4860 (0.0000)	-1.4082 (0.0000)	-1.1683 (0.0000)
AADT $_{Major}$	0.1948 (0.0000)	0.2155 (0.0000)	0.2113 (0.0000)	0.1883 (0.0000)
AADT $_{Minor}$	0.1262 (0.0010)	0.1539 (0.0010)	0.1379 (0.0010)	0.1266 (0.0010)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.2668 (0.0010)	0.1709 (0.0500)	0.2485 (0.0040)	0.2728 (0.0010)
Crosswalk on each of the minor approaches: 2	N/S [a]	N/S	N/S	N/S
No. of approaches: 3 (Base)				
No. of approaches: 4	0.3878 (0.0010)	0.2369 (0.0950)	0.3547 (0.0070)	0.3994 (0.0010)
No. of through lanes on the minor approaches: 0 (Base)				
No. of through lanes on the minor approach: 1-3	N/S	N/S	N/S	N/S
No. of through lanes on the minor approach: 4 & 4+	0.8029 (0.0000)	0.8797 (0.0120)	0.8176 (0.0010)	0.7760 (0.0000)
Over-dispersion	2.3705	0.5680	1.4209	0.4708
P	1.0000 (0.0000)	2.0000 (0.0000)	1.3598 (0.0000)	
Log L [b]	-1369.80	-1372.45	-1368.53	-1362.96
AIC	2757.61	2762.89	2757.06	2743.93
BIC	2796.61	2801.89	2800.39	2782.93
PDO CRASHES				
Intercept	-1.039 (0.0020)	-1.2663 (0.0000)	-1.2223 (0.0010)	-1.0004 (0.0030)
AADT $_{Major}$	0.1011 (0.0370)	0.0989 (0.0600)	0.1067 (0.038)	0.0987 (0.0420)
AADT $_{Minor}$	0.1619 (0.0010)	0.2040 (0.0000)	0.1842 (0.0010)	0.1583 (0.0010)
No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	N/S	N/S	N/S	N/S
Crosswalk on each of the minor approaches: 2	N/S	N/S	N/S	N/S
No. of approaches: 3 (Base)				
No. of approaches: 4	0.3189 (0.0000)	0.2122 (0.0280)	0.2912 (0.0030)	0.3291 (0.0000)
No. of through lanes on the minor approaches: 0 (Base)				
No. of through lanes on the minor approach: 1-3	N/S	N/S	N/S	N/S
No. of through lanes on the minor approach: 4 & 4+	N/S	N/S	N/S	N/S
Over-dispersion	1.7930	0.7114	1.1867	0.4167
P	1.0000 (0.0000)	2.0000 (0.0000)	1.4507 (0.0000)	
Log L	-1149.44	-1149.70	-1148.76	-1140.98
AIC	2308.87	2309.41	2309.51	2291.96
BIC	2330.54	2331.07	2335.51	2313.63
INJURY & FATAL CRASHES				
Intercept	-3.6679 (0.0000)	-4.0729 (0.0000)	-4.0075 (0.0000)	-3.6730 (0.0000)
AADT $_{Major}$	0.4225 (0.0000)	0.4507 (0.0000)	0.4489 (0.0000)	0.4229 (0.0000)
AADT $_{Minor}$	N/S	N/S	N/S	N/S

No crosswalk: 0 (Base)				
Crosswalk on one of the minor approaches: 1	0.3155 (0.0030)	0.3582 (0.0020)	0.3439 (0.0020)	0.3136 (0.0030)
Crosswalk on each of the minor approaches: 2	N/S	0.3740 (0.0290)	0.3091 (0.0570)	N/S
No. of through lanes on the minor approaches: 0 (Base)				
No. of approaches: 3 (Base)				
No. of approaches: 4	N/S	N/S	N/S	N/S
No. of through lanes on the minor approaches: 1-3	0.4208 (0.0100)	0.5060 (0.0080)	0.4678 (0.0090)	0.4167 (0.0110)
No. of through lanes on the minor approaches: 4 & 4+	1.2610 (0.0000)	1.2376 (0.0020)	1.2515 (0.0000)	1.2484 (0.0000)
Over-dispersion	1.2707	0.5680	0.9755	0.3489
P	1.0000 (0.0000)	2.0000 (0.0000)	1.4601 (0.0000)	
Log L	-931.85	-933.04	-929.64	-931.01
AIC	1881.70	1884.07	1879.28	1880.02
BIC	1920.70	1923.07	1922.61	1919.02

Notes: [a] Not Significant, [b] Log-Likelihood

4.3 Comparison and performance evaluation of the developed SPFs

The likelihood ratio test (LRT) was used for comparing either the NB-1 with the NB-P model or the NB-2 with the NB-P model since both (NB-1 and NB-2) are parametrically nested by the NB-P (Greene, 2008). However, the LRT was not applied to compare the non-nested models, i.e., the NB-1 model against the NB-2 model, or the NB models against the GP model. Instead, the AIC and BIC were used as in Ismail and Jemain (2007).

The LRT indicated that the NB-P model performed better than the NB-1 model for total crashes, PDO crashes, and injury & fatal crashes in case of signalized intersections (**Table 5**). However, the result of the LRT test was inconclusive when the NB-P and NB-2 were compared and, hence, it cannot be said with certainty, which of the two was a better model. Based on other measures, e.g., the log-likelihood, AIC, and BIC (**Table 3**), it can be seen that the NB-P and NB-2 performed relatively closely, but both performed better than the NB-1 models and GP models for all crash severities. The functional parameter "P" of the estimated NB-P models was statistically significant across all severity levels. The estimated value of the functional parameter "P" of the NB-P models for total crashes, and injury & fatal crashes was close to 2, while for the PDO crashes, it was significantly different from either 1 or 2 (**Table 3**). Although this does not entirely verify the assumption that the NB-1 or NB-2 models' restricted variance structure may lead to biased estimates of model parameters, it does not entirely reject the

20

possibility either, as indicated by the PDO crashes on signalized intersections and the result for the NB-

351 1 models.

352 Table 5 Likelihood ratio (NB-1 vs NB-P and NB-2 vs NB-P) for signalized and unsignalized intersections

	Signalized Intersections		Unsignalized Intersections	
TOTAL CRASHES				
Test/Criteria	**NB-1**	**NB-P**	**NB-1**	**NB-P**
Log L[a]	-653.028	-640.655	-1369.804	-1368.529
Likelihood ratio ($\chi 2$)		**24.75** (0.0000) [b]		2.55 (0.1104)
Test/Criteria	**NB-2**	**NB-P**	**NB-2**	**NB-P**
Log L	-640.655	-640.6552	-1372.4456	-1368.529
Likelihood ratio ($\chi 2$)		0.0002 (0.9893)		**7.83 (0.0051**)
PDO CRASHES				
Test/Criteria	**NB-1**	**NB-P**	**NB-1**	**NB-P**
Log L	-538.894	-532.906	-1149.436	-1148.756
Likelihood ratio ($\chi 2$)		**11.98 (0.0005)**		1.36 (0.2436)
Test/Criteria	**NB-2**	**NB-P**	**NB-2**	**NB-P**
Log L	-533.324	-532.906	-1149.705	-1148.756
Likelihood ratio ($\chi 2$)		0.84 (0.3606)		1.90 (0.1682)
INJURY & FATAL CRASHES				
Test/Criteria	**NB-1**	**NB-P**	**NB-1**	**NB-P**
Log L	-531.954	-524.388	-931.851	-929.6407
Likelihood ratio ($\chi 2$)		**15.13 (0.0001)**		**4.42 (0.0355)**
Test/Criteria	**NB-2**	**NB-P**	**NB-2**	**NB-P**
Log L	524.404	-524.388	-933.036	-929.6407
Likelihood ratio ($\chi 2$)		0.03 (0.8569)		**6.79 (0.0092)**

353 *Notes: Bold values indicate statistically significant results of the LRT*
354 *[a] Log-Likelihood*
355 *[b] Values in parenthesis indicate the p-value when the likelihood ratio ($\chi 2$) was computed.*

356 The LRT for unsignalized intersections showed that the NB-P and NB-1 models performed

357 equally closely for total crashes and PDO crashes. We cannot say that the difference in the NB-P and

358 NB-1 estimates was significant. However, for injury & fatal crashes, the results were in favor of the

359 NB-P models. When compared with the NB-2 model, the NB-P model was a better performing model

360 for total crashes and injury & fatal crashes, but there was no significant difference in the NB-2 and NB-

361 P models for PDO crashes. Based on the AIC and BIC values, the NB-1 models performed better than

362 the NB-2 models (non-nested models comparison, **Table 4**) for total crashes and injury & fatal crashes,

363 while results for the PDO crashes were reasonably close for the two traditional NB models. However,

364 the AIC and BIC showed better model fit for the GP models in all crash severity levels except injury &

365 fatal crashes. So, it will be safe to say that the GP model outperformed the NB models in the case of

un-signalized intersections for most crash severities in this study. The functional parameter "P" of variance structure was significant for the NB-P models across all severity levels, and it was not close to either 1 or 2. This verifies the assumption that the restricted variance structure of the NB-1 and NB-2 models might lead to biased estimates of model parameters for unsignalized intersection, and, hence the NB-P that takes into account the flexible variance structure would be more reliable in the accurate estimation of model parameters when there is no GP model considered.

Table 6 Predictive performance evaluation and validation of estimated models of signalized and unsignalized intersections

Crash Severity	Criteria	Signalized Intersections (198)				Unsignalized Intersections (562)			
		NB-1	NB-2	NB-P	GP	NB-1	NB-2	NB-P	GP
TOTAL CRASHES	MPB	-0.233	-0.268	-0.268	-0.237	-0.035	-0.034	-0.035	-0.034
	MAD	1.083	1.082	1.082	1.088	0.509	0.510	0.507	0.509
	MSPE	2.998	2.932	2.932	3.042	0.472	0.473	0.470	0.471
	CURE Deviation (%)	26	4	4	36	0	1	0	0
	Validation Factor (VF)	1.094	1.110	1.110	1.096	0.954	0.952	0.953	0.955
PDO CRASHES	MPB	0.043	0.031	0.040	0.041	-0.025	-0.027	-0.026	-0.027
	MAD	0.688	0.693	0.693	0.691	0.340	0.337	0.337	0.337
	MSPE	1.100	1.106	1.097	1.111	0.419	0.420	0.419	0.418
	CURE Deviation (%)	19	5	6	21	0	0	0	1
	Validation Factor (VF)	1.033	1.024	1.031	1.032	0.946	0.943	0.944	0.944
INJURY & FATAL CRASHES	MPB	0.063	0.039	0.040	0.064	-0.002	0.002	-0.001	-0.002
	MAD	0.629	0.623	0.624	0.629	0.248	0.246	0.247	0.246
	MSPE	0.930	0.911	0.913	0.930	0.119	0.121	0.121	0.119
	CURE Deviation (%)	7	2	2	6	1	1	1	1
	Validation Factor (VF)	1.058	1.036	1.037	1.059	0.996	1.006	0.998	0.997

Notes: MPB: Mean Prediction Bias, MAD: Mean Absolute Deviation, MSPE: Mean Squared Prediction Error

Besides the likelihood-based criteria, predictive ability-based measures were also computed to validate the developed models and examine their predictive performance. It is important to note that randomly selected 80% of data were used to estimate models, while the remaining 20% were used to validate the developed models. We compute the MPB, MAD, MSPE, % CURE deviation, and a validation factor. According to Oh et al. (2003), the smaller the absolute values of the MPB, MAD, and MSPE, the better the developed models' performance. The % CURE deviation, which denotes the

379   percentage of the data points falling outside the two standard deviation limits of the Cumulative

380   Residual (CURE) (Hauer, 2015), shows a good fit when its values are small (Wang et al., 2019). Finally,

381   a factor that we called a validation factor, was calculated as the total predicted crashes' ratio to the total

382   observed crashes using the validation sample. A value close to one indicated a better model (Wang et

383   al., 2019). Wang et al., (2019) called it a calibration factor.

384   In the case of signalized intersections, the NB-2 and NB-P models' predictive performances

385   were better than that of the NB-1 model for all crash severity levels based on the measures in **Table 6**.

386   Similarly, the NB-2 and NB-P models also outperformed the GP model for total crashes, the PDO

387   crashes, and injury & fatal crashes. Of particular interest was the percentage CURE Deviation, the

388   values of which were very high for all crash severity levels on the signalized intersections in the case

389   of the NB-1 and GP models that indicate a poor prediction performance. For unsignalized intersections,

390   the difference between the predictive performance measures across the developed SPFs was minimal

391   and somewhat inconsequential. However, close observation of the results showed that the NB-1, NB-

392   P, and GP models' performances were almost similar and slightly better than the NB-2 model. This

393   finding is in line with the results from **Table 4** and **Table 5**. To put things into perspective, the GP

394   regression was the better performing model for total crashes and PDO crashes, and the NB-P regression

395   was the better performing model for injury & fatal crashes on the unsignalized intersections. Among

396   the NB models, the NB-P and NB-1 performed closely and relatively better than the NB-2 model.

397   In the end, we also estimated prediction intervals for the validation sample using the approach

398   developed by Wood (2005). A prediction interval (PI) is an interval that allows a future or new

399   observation to fall within it with $100(1-\alpha)$ % confidence when the observation is recorded. The greater

400   the observed data fall within the PI, the more the future observations are correctly predicted by the

401   model. For instance, if the model is well-fitted to the data, it should capture about 50% of the actual

402   counts of the new similar sites for 50% PI, 75% of the actual counts of the new similar sites for 75%

403   PI, etc.

404   **Table 7** provides a range of PIs calculated for each intersection type. For signalized

405   intersections, the 95% PIs captured 100% data of the validation sample for all models and all severity

levels except the NB-1 model for total crashes. The 90% PI captured 100% of the actual PDO crash

counts of validation sites for all models. However, the observations captured by the 90% PIs for total

and injury & fatal crashes were less than 100%. The 90% PI developed from the GP model caught

around 97.7% of the validation sample, the highest in the total crash category. The 90% PIs of NB

models captured an almost similar percentage of actual total crashes (i.e., 95.5%). In the case of injury

& fatal crashes, the 90% PIs for four models caught a similar proportion of data. For tighter intervals

(i.e., 75% and 50%), the NB-P model's PIs were winners for total crashes and injury & fatal crashes.

When PDO crashes were considered, the NB-2 was a winner for 50% PI, but it tied with the NB-P for

75% PI. -In general, the NB-2 and NB-P models performed reasonably close to each other and better

than the NB-1 and GP models. This finding confirms the results obtained from the likelihood-based

measures and predictive ability-based measures in Table 5 and Table 6, respectively.

Table 7 The proportion of validation data captured by 50%, 75%, 90% and 95% prediction intervals of NB-1, NB-2, NB-P and GP models- Signalized and unsignalized intersections

Crash Severity	Prediction Interval (%)	Signalized Intersections (198)				Unsignalized Intersections (562)			
		NB-1	NB-2	NB-P	GP	NB-1	NB-2	NB-P	GP
TOTAL CRASHES	50	0.778	0.800	0.822	0.800	0.805	0.814	0.867	0.823
	75	0.889	0.911	0.933	0.933	0.901	0.920	0.947	0.903
	90	0.956	0.956	0.955	0.977	0.973	0.982	0.973	0.973
	95	0.978	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PDO CRASHES	50	0.844	0.867	0.844	0.844	0.832	0.850	0.858	0.805
	75	0.933	0.956	0.956	0.933	0.920	0.929	0.920	0.894
	90	1.000	1.000	1.000	1.000	0.973	0.982	0.982	0.965
	95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
INJURY & FATAL CRASHES	50	0.800	0.800	0.800	0.800	0.823	0.832	0.841	0.814
	75	0.889	0.911	0.933	0.911	0.964	0.965	0.973	0.920
	90	0.978	0.978	0.978	0.978	0.982	0.982	0.982	0.982
	95	1.000	1.000	1.000	1.000	0.991	0.991	0.991	0.982

In the case of unsignalized intersections, the 95% PIs for all models caught 100% of total

crashes and PDO crashes in the validation sample. However, the 95% PIs in the injury & fatal crash

category caught 99.1% and 98.2% of the validation data for all NB models and the GP model, respectively. The 90% PI performed almost similar across all four models for total crashes and exactly similar for injury & fatal crashes. Only the NB-2 model was an exception to this generalization about total crashes for which the 90% PI caught 98.2% of the data, while for other models, the proportion of caught data was 97.3%. For PDO crashes when 90% PI was considered, the NB-2 and NB-P performance was identical and only slightly better than the remaining two models. For tighter intervals (i.e., 50% and 75%), the results were mixed. For instance, the data captured by the 75% PIs for the NB-P models was highest in the case of total crashes and injury & fatal crashes. When PDO crashes were considered, the result for the 75% PI for the NB-2 model was better, but it was very close to the NB-P model. For 50% PI, the winner was the NB-P model in all crash severity levels, which captured the highest data percentage. In general, the results were mixed and inconsequential for unsignalized intersections, similar to the criteria in Table 5 and Table 6. All models performed reasonably close to one another for wider PIs (i.e., 95% and 90%). However, for tighter PIs, the NB-P models were favored. The GP models' PIs caught fewer data than that of the NB models, but the difference was negligible.

We noticed that the calculated PIs captured relatively higher percentages of the validation data than expected. This could be the result of many factors acting simultaneously. In this study, we followed the approach proposed by Wood (2005) to estimate the PIs. According to Wood (2005), the lower limit of the interval should be set to zero to avoid negative boundary measures. Setting the lower limit of the PI to zero resulted in comparatively wider intervals that potentially captured more actual counts than expected. Another reason could be the dependability of the approach applied to estimate PIs on the model's dispersion parameter. We estimated our models with the fixed dispersion parameters. Geedipally and Lord (2008) reported that fixed dispersion parameter models tend to have larger PIs than those with varying dispersion parameters, and, in extreme cases, the difference could reach up to 40%. We possibly got the considerably wider PIs for our models because of assuming a fixed dispersion parameter for the whole network. Other studies (Ramírez et al., 2009; Geedipally and Lord, 2008; Lord, 2006; Wood, 2005) estimated the PIs for exposure-only models. However, we considered additional

448 variables (**Table 1**). Maybe the presence of these additional variables improved the predictive

449 performance of the SPFs developed and, hence, resulted in more inclusive PIs.

5. Discussion

451 The current study investigated the application of NB-1, NB-2, NB-P, and a GP model to

452 develop SPFs and, among them, find a statistical model capable of improved estimation accuracy. For

453 this purpose, several goodness of fit (likelihood-based) and predictive performance measures were

454 calculated and compared. Before that, several variables (**Table 1 and Table 2**) were used to develop

455 the SPFs (**Table 3 and Table 4**). Statistical modeling revealed that only a few explanatory variables

456 have a significant relationship with the crash occurrence. The following section presents the discussion

457 of the results obtained in section 4.

5.1 Predictor variables of crashes on urban intersections

459 A positive association between the crash frequency and traffic volume of major and minor

460 intersections approaches was found for almost every severity level and every intersection type

461 considered. This result was in accordance with our expectations. When the number of vehicles entering

462 and/or leaving the intersection increases, it induces new turning and crossing maneuvers, that result in

463 an increased risk of new conflicts. Those other conflicts, in some cases, are translated into actual

464 crashes. Many studies have reported similar results (Wang et al., 2019; Barbosa et al., 2014; Ferreira

465 and Couto, 2013; Vieira Gomes et al., 2012; Miaou and Lord, 2003; Greibe, 2003). It is important to

466 note that, of all the models developed across all severity levels for both intersection types, only the

467 traffic flow of a minor approach of unsignalized intersections in the injury & fatal crash models was

468 not significant. Since the majority of unsignalized intersections were located on local streets where the

469 traffic volume of minor approaches and the corresponding speed limits were relatively very low, it is

470 possible that those factors might have contributed to a reduced number of fatal and injury crashes and

471 hence resulted in the insignificance of traffic volume of the minor approaches in models for fatal and

472 injury crashes.

473   The presence of crosswalks on the minor approaches, although significant, gave somewhat mixed

474   results for signalized and unsignalized intersections. In the case of signalized intersections, the

475   crosswalk on the minor approaches had a significant positive association with crash frequency only

476   when it was present on both approaches. The crosswalk, however, was not significant when it existed

477   on one of the minor approaches. Moreover, the estimated coefficients were often more than double for

478   intersections with crosswalks on both minor approaches than intersections with crosswalks on only one

479   approach, although not significant in the latter case. A possible explanation could be that, at signalized

480   intersections with crosswalks on both minor approaches, an existing and/or entering or turning traffic

481   will have two possible vehicle-pedestrian interactions and, thus, the chances of involvement in crashes

482   will be greater. In contrast, intersections with a crosswalk on only one minor approach will have one

483   possible vehicle-pedestrian interaction, so lower risk of a crash. In the case of unsignalized intersection,

484   a crosswalk on a minor approach was significant across all but PDO models when it was present on

485   only one of the minor approaches. As we know, two crosswalks on the minor approaches were only

486   present on four-legged intersections. In the unsignalized category, the majority of intersections were

487   three-legged, which could accommodate only one crosswalk on its minor approach at a time. Thus,

488   most three-legged intersections, and the consequent presence of only one crosswalk on a minor

489   approach possibly contributed significantly to crashes on unsignalized intersections.

490   The intersection skewness was statistically significant in the models for total crashes, and injury

491   & fatal crashes in the case of signalized intersections. The association found indicates that more crashes

492   were expected on the intersections with a high skewness level than those with none or low skewness

493   level. For recollection of readers, we classified an intersection angle of less than or equal to 75 degrees

494   as a high skewness level and an intersection angle of greater than 75 degrees as low skewness or no

495   skewness level. Nightingale et al. (2017) and Harkey (2013) reported similar results when studying the

496   skewness angle's influence on intersection crash frequency. However, Kumfer et al. (2019) found

497   somewhat mixed results and concluded that the most dangerous intersection arrangements might not be

498   those with the smallest intersection angles between the adjacent major and minor approaches. Instead,

499   the crashes tend to be their highest when the intersection angle is between 50 and 70 degrees. Further,

500 they found most intersections with skew angles of 50 degrees or below relatively safe. This indicates

501 that analysts must exercise caution when interpreting the impact of intersection skewness on its safety.

502 Reasons for the significance of skewness in signalized intersection SPFs might be related to the fact

503 drivers tend to have greater safety perception at these intersection types compared to unsignalized

504 intersections. This reduces the scale of decision-making on the driver's behalf, necessary for safe

505 driving. When they encounter a skewed intersection, this potentially leads to confusion, which is

506 reinforced by other undesirable characteristics (obstructed views, distorted sight distances, large

507 intersection area, large turns, etc.) of skewed intersection may result in a crash.

508 The number of approaches of the intersection was found to influence the expected crash

509 frequency at unsignalized intersections only. This was particularly true in the case of PDO crashes and

510 total crashes. The intersections with four or more legs were expected to experience more crashes than

511 the intersections with three legs. This was an expected outcome. An increase in the number of

512 legs/approaches increases intersection complexity. It invites additional traffic, which could be related

513 to an increased risk of involvement in a crash.

514 Another significant predictor of crashes at un-signalized intersections was the number of

515 through lanes of the minor approach. The association between the number of through lanes and the

516 expected crash frequency was positive, which means more crashes with an increased number of through

517 lanes. Abdel-Aty and Nawathe (2006) found similar results for urban intersections, but their study was

518 focused on signalized intersections. Zhao et al. (2018) and Kamrani et al. (2017) also reported a

519 significant positive association between crash frequency and the number of through lanes for

520 intersections. The number of through lanes on a minor approach also indirectly informs about the size

521 of an intersection and correspondingly higher traffic volumes. An intersection with many through lanes

522 on a minor approach could have a higher expected crash frequency because of its large size that carries

523 more traffic. This result can be generalized only to four-legged unsignalized intersections since through

524 lanes were only reported for such facility type.

525 We also found some unexpected results, especially the insignificance of the exclusive left and

526 right turn lanes in the developed models. It was rather opposite to some studies' results (Al-Kaisy and

28

527   Roefaro, 2012; Abdel-Aty and Haleem, 2011; Zhou et al., 2010). The reason may be that the number

528   of intersections with exclusive left and right turn lanes in the study data was not that much to be

529   significant in the final models. The influence of the intersection size on crash frequency was also

530   insignificant as a predictor variable. This might be because other variables, like, the number of through

531   lanes on a minor approach and the number of legs/approaches of the intersection, could have acted as

532   proxies for intersection size in the modelling process.

533   *5.2   The appropriate model(s) for crash estimation on urban intersections*

534   In the case of signalized intersections, both the likelihood-based and predictive ability-based

535   measures revealed that the NB-P and NB-2 models performed better than the NB-1 and GP models.

536   The estimated PIs for models of signalized intersections also confirmed these findings. For un-

537   signalized intersections, the GP model was a winner for total and PDO crashes. However, for injury &

538   fatal crashes, the NB-P outperformed the competing models based on the goodness of fit (likelihood-

539   based) measures. Though the predictive ability-based measures' results were inconsequential, a close

540   observation favored the GP models only marginally compared to the NB-1 and NB-P models. When

541   the comparison was made across the NB models for unsignalized intersections, the NB-P and NB-1

542   models performed better than the NB-2. When the PIs were estimated for unsignalized intersections,

543   similar results were obtained for the developed models. Generally, in situations where the NB models

544   were considered, the flexible variance structure allowed the NB-P model to outperform the traditional

545   NB-1 or NB-2 model. Another observation was that for one type of facility (un-signalized intersections),

546   the better performing model was the NB-1, while for the other type of facility (signalized intersection),

547   the better performing model was the NB-2 when only the traditional NB models were compared. This

548   finding suggests that it is necessary to check for an appropriate model form in advance.

549   The use of several functional forms of the NB regression and the equally powerful but relatively

550   less used GP model in this study revealed that the accurate estimation of crash frequency is subjected

551   to selecting the appropriate functional form and model type. The flexible variance structure of the NB

552   model can improve estimation accuracy. Further, the study results showed that it is possible that a model

29

553 functional form, appropriate for one sub-type of the same infrastructure, might not be appropriate for

554 another sub-type of that infrastructure.

6. Conclusions

556   This study developed multiple SPFs by crash severity for urban intersections using the NB-1,

557 NB-2, NB-P, and GP regression to obtain models with higher estimation accuracy. The data was

558 obtained for the intersections of Antwerp, Belgium. Only those intersections were included in modeling

559 for which sufficient good quality data was available. The major and minor approaches' AADT and

560 several other variables related to road infrastructure and geometry were used as the explanatory

561 variables. The traffic volume was a significant predictor of crash frequency for almost all developed

562 models and all crash severity levels. Other significant variables include the presence of a crosswalk on

563 the minor approach and the intersection skewness in signalized intersections. For unsignalized

564 intersection, the presence of a crosswalk on the minor approach, the number of through lanes of the

565 minor approach, and the number of legs were significant.

566   Several measures were computed for model comparison. The likelihood-based measures,

567 including the LRT, AIC, and BIC, were used for checking the goodness of fit of the models, while the

568 predictive ability-based measures were used for the predictive performance and validation of the

569 models. The likelihood-based measures showed that the NB-P and NB-2 models performed better than

570 the NB-1 and GP models for signalized intersections and for all crash severity levels. However, for

571 unsignalized intersections, the GP model was relatively better than the NB models for most crash

572 severity levels. A comparison among the NB models showed that the NB-P and NB-1 outperformed the

573 NB-2 model. The predictive ability-based measures confirmed the above results by indicating an

574 improvement in prediction accuracy in the NB-P model and GP model for signalized and unsignalized

575 intersections, respectively. Similarly, the PIs calculated for the validation data confirmed those findings

576 though the data caught by the PIs were comparatively higher than expected.

577   All models (NB-1, NB-2, NB-P, and GP) developed in this study were promising in estimating

578 the SPFs for intersections. Irrespective of the functional form or type, they showed similar results for

579 explanatory variables' influence on crash frequency. Further, it was shown that the use of the flexible

580 variance structure of the NB-P model and/or an entirely different GP model could bring an improvement

581 in the estimation accuracy as indicated by the comparison of the goodness of fit and later verification

582 by the predictive performance measures and prediction intervals for -the validation sample.

583   Finally, we hope that this study's outcome adds to the SPF estimation knowledge concerning

584 the selection of the appropriate parametrization or model type and improvement in the accuracy and

585 reliability of the crash estimates. Nonetheless, future research efforts can focus on investigating the

586 applications of the NB-P model to several other facility types or/and crash types or using the NB-P

587 model in conjunction with other techniques, for instance, exploring the functional forms of the GP

588 model of which a traditional form called GP-1 has already been applied in this study. Since this study

589 developed separate models for different crash severity levels, a comparison with multivariate models

590 with multiple outcomes (i.e., count of total crashes, count of fatal and injury crashes, and count of

591 property damage only [PDO] crashes) can be made in the future.

References

596 AASHTO, 2014. Highway Safety Manual(HSM), First Edition, 2014 Supplement American
597   Association of State and Highway Transportation Officials, Washington DC.
598 AASHTO, 2010. Highway Safety Manual(HSM), First Edition, American Association of State and
599   Highway Transportation Officials, Washington DC.
600 Abdel-Aty, M., Haleem, K., 2011. Analyzing angle crashes at unsignalized intersections using machine
601   learning techniques. Accid. Anal. Prev. 43 1 , 461–470. doi:10.1016/j.aap.2010.10.002
602 Abdel-Aty, M., Nawathe, P., 2006. A Novel Approach for Signalized Intersection Crash Classification
603   and Prediction. Adv. Transp. Stud. 9.
604 Abdel-Aty, M., Radwan, A.E., 2000. Modeling traffic accident occurrence and involvement. Accid.
605   Anal. Prev. 32 5 , 633–642. doi:10.1016/S0001-4575(99)00094-9
606 Abdel-Aty, M.A., Lee, J., Eluru, N., Cai, Q., Al Amili, S., Alarifi, S., 2016. Enhancing and Generalizing
607   the Two-Level Screening Approach Incorporating the Highway Safety Manual (HSM)
608   Methods, Phase 2 (No. BDV-24-977-06). Florida Department of Transportation.
609 Al-Kaisy, A., Roefaro, S., 2012. Channelized right-turn lanes at signalized intersections: the U.S.
610   experience 13.

611   Barbosa, H., Cunto, F., Bezerra, B., Nodari, C., Jacques, M.A., 2014. Safety performance models for
612   urban intersections in Brazil. Accid. Anal. Prev. 70, 258–266. doi:10.1016/j.aap.2014.04.008
613   Cafiso, S., Di Silvestro, G., Di Guardo, G., 2012. Application of Highway Safety Manual to Italian
614   Divided Multilane Highways. Procedia - Soc. Behav. Sci., SIIV-5th International Congress -
615   Sustainability of Road Infrastructures 2012 53, 910–919. doi:10.1016/j.sbspro.2012.09.940
616   Cameron, A.C., Trivedi, P.K., 1986. Econometric Models Based on Count Data: Comparisons and
617   Applications of Some Estimators and Tests. J. Appl. Econom. 1 1 , 29–53.
618   Chang, G.-L., Xiang, H., 2003. The Relationship between Congestion Levels and Accidents. Maryland
619   State Highway Administration.
620   Dixon, K., Monsere, C., Avelar, R., Barnett, J., Escobar, P., Kothuri, S., Wang, Y., 2015. Improved
621   safety performance functions for signalized intersections (No. FHWA-OR-RD-16-03). Oregon
622   Department of Transportation.
623   Dong, C., Richards, S.H., Clarke, D.B., Zhou, X., Ma, Z., 2014. Examining signalized intersection crash
624   frequency using multivariate zero-inflated Poisson regression. Saf. Sci. 70, 63–69.
625   doi:10.1016/j.ssci.2014.05.006
626   Elvik, R., Sagberg, F., Langeland, P.A., 2019. An analysis of factors influencing accidents on road
627   bridges in Norway. Accid. Anal. Prev. 129, 1–6. doi:10.1016/j.aap.2019.05.002
628   Famoye, F., Singh, K.P., Wulu, J.T., 2004. On the Generalized Poisson Regression Model with an
629   application to Accident Data. J. Data Sci. 2, 287–295.
630   Ferreira, S., Couto, A., 2013. Traffic flow-accidents relationship for urban intersections on the basis of
631   the translog function. Saf. Sci. 60, 115–122. doi:10.1016/j.ssci.2013.07.007
632   Gates, T., Savolainen, P., Avelar, R., Geedipally, S., Lord, D., Anthony, I., Stapleton, S., 2018. Safety
633   Performance Functions for Rural Road Segments and Rural Intersections in Michigan | Blurbs
634   New | Blurbs | Main (No. SPR-1645). Michigan Department of Transportation.
635   Geedipally, S.R., Lord, D., 2010. Investigating the effect of modeling single-vehicle and multi-vehicle
636   crashes separately on confidence intervals of Poisson–gamma models. Accid. Anal. Prev. 42 4
637   , 1273–1282. doi:10.1016/j.aap.2010.02.004
638   Geedipally, S.R., Lord, D., 2008. Effects of Varying Dispersion Parameter of Poisson–Gamma Models
639   on Estimation of Confidence Intervals of Crash Prediction Models. Transp. Res. Rec. 2061 1 ,
640   46–54. doi:10.3141/2061-06
641   Giuffrè, O., Granà, A., Giuffrè, T., Marino, R., Marino, S., 2014. Estimating the Safety Performance
642   Function for Urban Unsignalized Four-Legged One-Way Intersections in Palermo, Italy. Arch.
643   Civ. Eng. 60 1 , 41–54. doi:10.2478/ace-2014-0002
644   Greene, W., 2008. Functional forms for the negative binomial model for count data. Econ. Lett. 99 3 ,
645   585–590. doi:10.1016/j.econlet.2007.10.015
646   Greibe, P., 2003. Accident prediction models for urban roads. Accid. Anal. Prev. 35 2 , 273–285.
647   doi:10.1016/S0001-4575(02)00005-2
648   Haleem, K., Abdel-Aty, M., 2010. Examining traffic crash injury severity at unsignalized intersections.
649   J. Safety Res. 41 4 , 347–357. doi:10.1016/j.jsr.2010.04.006
650   Hardin, J.W., Hilbe, J.M., 2018. Generalized Linear Models and Extensions, Second Edition. Stata
651   Press.
652   Harkey, D.L., 2013. Effect of Intersection Angle on the Safety of Intersections. North Carolina State
653   University.
654   Hauer, E., 2015. The Art of Regression Modeling in Road Safety. Springer International Publishing.
655   doi:10.1007/978-3-319-12529-9
656   Hauer, E., Bamfo, J., 1997. Two tools for finding what function links the dependent variable to the
657   explanatory variables., in: ICTCT 1997Conference. Lund, p. 18.
658   Hilbe, J.M., 2011. Negative Binomial Regression, Second. ed. Cambridge university press, New York.
659   doi:10.1017/CBO9780511973420
660   Ismail, N., Jemain, A.A., 2007. Handling Overdispersion with Negative Binomial and Generalized
661   Poisson Regression Models. Casualty Actuar. Soc. Forum Citeseer.
662   Ismail, N., Zamani, H., 2013. Estimation of Claim Count Data using Negative Binomial, Generalized
663   Poisson, Zero-Inflated Negative Binomial and Zero-Inflated Generalized Poisson Regression
664   Models. Casualty Actuar. Soc. E-Forum 41 20 , 1–28.

665 Janstrup, K.H., 2016. Statistical modelling of the frequency and severity of road accidents. DTU
666   Transport.
667 Joe, H., Zhu, R., 2005. Generalized Poisson distribution: the property of mixture of Poisson and
668   comparison with negative binomial distribution. Biom. J. Biom. Z. 47 2 , 219–229.
669   doi:10.1002/bimj.200410102
670 Joshua, S.C., Garber, N.J., 1990. Estimating truck accident rate and involvements using linear and
671   Poisson regression models. Transp. Plan. Technol. 15 1 , 41–58.
672   doi:10.1080/03081069008717439
673 Kamrani, M., Wali, B., Khattak, A.J., 2017. Can Data Generated by Connected Vehicles Enhance
674   Safety?: Proactive Approach to Intersection Safety Management. Transp. Res. Rec. 2659 1 ,
675   80–90. doi:10.3141/2659-09
676 Kumfer, W., Harkey, D., Lan, B., Srinivasan, R., Carter, D., Patel Nujjetty, A., Eigen, A.M., Tan, C.,
677   2019. Identification of Critical Intersection Angle through Crash Modification Functions.
678   Transp. Res. Rec. J. Transp. Res. Board 2673 2 , 531–543. doi:10.1177/0361198119828682
679 Liu, C., Sharma, A., 2018. Using the multivariate spatio-temporal Bayesian model to analyze traffic
680   crashes by severity. Anal. Methods Accid. Res. 17, 14–31. doi:10.1016/j.amar.2018.02.001
681 Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects
682   of low sample mean values and small sample size on the estimation of the fixed dispersion
683   parameter. Accid. Anal. Prev. 38 4 , 751–766. doi:10.1016/j.aap.2006.02.001
684 Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and
685   assessment of methodological alternatives. Transp. Res. Part Policy Pract. 44 5 , 291–305.
686   doi:10.1016/j.tra.2010.02.001
687 Lord, D., Park, B.-J., 2015. Appendix D: Negative Binomial Regression Models and Estimation
688   Methods.
689 Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and
690   future directions. Anal. Methods Accid. Res. 1 0 .
691 Melliana, A., Setyorini, Y., Eko, H., Rosi, S., Purhadi, 2013. The Comparison Of Generalized Poisson
692   Regression And Negative Binomial Reression Methods In Overcoming Overdispersion [WWW
693   Document]. undefined. URL /paper/The-Comparison-Of-Generalized-Poisson-Regression-In-
694   Melliana-Setyorini/9e679999bb7585e9965b271dc5afc462b8572114 (accessed 8.21.20).
695 Miaou, S.-P., 1994. The relationship between truck accidents and geometric design of road sections:
696   Poisson versus negative binomial regressions. Accid. Anal. Prev. 26 4 , 471–482.
697   doi:10.1016/0001-4575(94)90038-8
698 Miaou, S.-P., Lord, D., 2003. Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion
699   Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods: Transp. Res. Rec.
700   doi:10.3141/1840-04
701 Miaou, S.-P., Lum, H., 1993. Modeling vehicle accidents and highway geometric design relationships.
702   Accid. Anal. Prev. 25 6 , 689–709. doi:10.1016/0001-4575(93)90034-T
703 Nambuusi, B.B., Brijs, T., Hermans, E., 2008. A review of accident prediction models for road
704   intersections. UHasselt.
705 Nesheli, M.M., Che Puan, O., Roshandeh, A.M., 2009. Evaluation of Effect of Traffic Signal
706   Coordination System on Congestion. WSEAS Trans. Adv. Eng. Educ. 6 7 , 203–212.
707 Nightingale, E., Parvin, N., Seiberlich, C., Savolainen, P.T., Pawlovich, M., 2017. Investigation of
708   Skew Angle and Other Factors Influencing Crash Frequency at High-Speed Rural Intersections.
709   Transp. Res. Rec. 2636 1 , 9–14. doi:10.3141/2636-02
710 Oh, J., Kim, E., Kim, M., Choo, S., 2010. Development of conflict techniques for left-turn and cross-
711   traffic at protected left-turn signalized intersections. Saf. Sci. 48 4 , 460–468.
712   doi:10.1016/j.ssci.2009.12.011
713 Oh, J., Lyon, C., Washington, S., Persaud, B., Bared, J., 2003. Validation of FHWA Crash Models for
714   Rural Intersections: Lessons Learned. Transp. Res. Rec. 1840 1 , 41–49. doi:10.3141/1840-05
715 Okamoto, H., Koshi, M., 1989. A method to cope with the random errors of observed accident rates in
716   regression analysis. Accid. Anal. Prev. 21 4 , 317–332. doi:10.1016/0001-4575(89)90023-7
717 Park, B.-J., 2010. Application of finite mixture models for vehicle crash data analysis. Texas A&M
718   University.

Park, B.-J., Lord, D., 2009. Application of finite mixture models for vehicle crash data analysis. Accid. Anal. Prev. 41 4 , 683–691. doi:10.1016/j.aap.2009.03.007

Persaud, B., Saleem, T., Faisal, S., Lyon, C., Chen, Y., Sabbaghi, A., 2012. Adoption of Highway Safety Manual Predictive Technologies for Canadian Highways, in: 2012 Conference of the Transportation Association of Canada. Fredericton, New Brunswick.

Ramírez, B.A., Izquierdo, F.A., Fernández, C.G., Méndez, A.G., 2009. The influence of heavy goods vehicle traffic on accidents on different types of Spanish interurban roads. Accid. Anal. Prev. 41 1 , 15–24. doi:10.1016/j.aap.2008.07.016

Roshandeh, A.M., Agbelie, B.R.D.K., Lee, Y., 2016. Statistical modeling of total crash frequency at highway intersections. J. Traffic Transp. Eng. Engl. Ed. 3 2 , 166–171. doi:10.1016/j.jtte.2016.03.003

Roshandeh, A.M., Levinson, H.S., Li, Z., Patel, H., Zhou, B., 2014. New Methodology for Intersection Signal Timing Optimization to Simultaneously Minimize Vehicle and Pedestrian Delays. J. Transp. Eng. 140 5 , 04014009. doi:10.1061/(ASCE)TE.1943-5436.0000658

Srinivasan, R., Carter, D., 2011. Development of Safety Performance Functions for North Carolina (No. FHWA/NC/2010-09). North Carolina Department of Transportation.

Tay, R., 2015. A random parameters probit model of urban and rural intersection crashes. Accid. Anal. Prev. 84, 38–40. doi:10.1016/j.aap.2015.07.013

Vieira Gomes, S., Geedipally, S.R., Lord, D., 2012. Estimating the safety performance of urban intersections in Lisbon, Portugal. Saf. Sci. 50 9 , 1732–1739. doi:10.1016/j.ssci.2012.03.022

Wang, K., Ivan, J.N., Ravishanker, N., Jackson, E., 2017. Multivariate poisson lognormal modeling of crashes by type and severity on rural two lane highways. Accid. Anal. Prev. 99, 6–19. doi:10.1016/j.aap.2016.11.006

Wang, K., Zhao, S., Jackson, E., 2020. Investigating exposure measures and functional forms in urban and suburban intersection safety performance functions using generalized negative binomial - P model. Accid. Anal. Prev. 148, 105838. doi:10.1016/j.aap.2020.105838

Wang, K., Zhao, S., Jackson, E., 2019. Functional forms of the negative binomial models in safety performance functions for rural two-lane intersections. Accid. Anal. Prev. 124, 193–201. doi:10.1016/j.aap.2019.01.015

Wang, X., Abdel-Aty, M., Nevarez, A., Santos, J.B., 2008. Investigation of Safety Influence Area for Four-Legged Signalized Intersections: Nationwide Survey and Empirical Inquiry. Transp. Res. Rec. doi:10.3141/2083-10

Washington, S., Karlaftis, M.G., Mannering, F., Anastasopoulos, P.Ch., 2010. Statistical and Econometric Methods for Transportation Data Analysis, Second. ed. Chapman and Hall/CRC.

Winkelmann, R., 2008. Econometric Analysis of Count Data, 5th ed. Springer-Verlag, Berlin Heidelberg. doi:10.1007/978-3-540-78389-3

Wong, S.C., Sze, N.N., Li, Y.C., 2007. Contributory factors to traffic crashes at signalized intersections in Hong Kong. Accid. Anal. Prev. 39 6 , 1107–1113. doi:10.1016/j.aap.2007.02.009

Wood, G.R., 2005. Confidence and prediction intervals for generalised linear accident models. Accid. Anal. Prev. 37 2 , 267–273. doi:10.1016/j.aap.2004.10.005

Wu, Y., Abdel-Aty, M., Cai, Q., Lee, J., Park, J., 2018. Developing an algorithm to assess the rear-end collision risk under fog conditions using real-time data. Transp. Res. Part C Emerg. Technol. 87, 11–25. doi:10.1016/j.trc.2017.12.012

Wu, Z., Sharma, A., Mannering, F.L., Wang, S., 2013. Safety impacts of signal-warning flashers and speed control at high-speed signalized intersections. Accid. Anal. Prev. 54, 90–98. doi:10.1016/j.aap.2013.01.016

Yang, Z., Hardin, J.W., Addy, C.L., 2009. A score test for overdispersion in Poisson regression based on the generalized Poisson-2 model. J. Stat. Plan. Inference 139 4 , 1514–1521. doi:10.1016/j.jspi.2008.08.018

Zhao, M., Liu, C., Li, W., Sharma, A., 2018. Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach. J. Transp. Saf. Secur. 10 3 , 251–265. doi:10.1080/19439962.2017.1323059

Zhou, H., Ivan, J.N., Sadek, A.W., Ravishanker, N., 2010. Safety Effects of Exclusive Left-Turn Lanes at Unsignalized Intersections and Driveways. J. Transp. Saf. Secur. 2 3 , 221–238. doi:10.1080/19439962.2010.502613