Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes

Amir Pooyan Afghari, Md Mazharul Haque, Simon Washington

Both crash count and severity are thought to quantify crash risk at defined transport network locations (e.g. intersections, a particulate section of highway, etc.). Crash count is a measure of the likelihood of occurring a potential harmful event, whereas crash severity is a measure of the societal impact and harm to the society. As the majority of safety improvement programs are focused on preventing fatal and serious injury crashes, identification of high-risk sites—so-called blackspots—should ideally account for both severity and frequency of crashes. Past research efforts to incorporate crash severity into the identification of high-risk sites include multivariate crash count models, equivalent property damage only models and two-stage mixed models. These models, however, often require suitable distributional assumptions for computational efficiency, neglect the ordinal nature of crash severity, and are inadequate for capturing unobserved heterogeneity arising from possible correlations between crash counts of different severity levels. These limitations can ultimately lead to inefficient allocation of resources and misidentification of sites with high risk of fatal and serious injury crashes. Moreover, the implication of these models in blackspot identification is an important, unanswered question.

While a joint econometric model of crash count and crash severity has the flexibility to account for the limitations mentioned previously, its ability to identify high-risk sites also needs to be examined. This study aims to fill this research gap by employing the joint model for blackspot identification. Using data from state-controlled roads in Queensland, Australia, a new risk score is developed based on predicted crash counts by severity, weighted by the cost ratio of severity levels. This weighted risk score is then used for identifying road segments with high risk of fatal and injury crashes. Results show that the joint model of crash count and crash severity has substantially improved prediction accuracy compared to the traditional count models. The correlation between crash counts of different severity levels captures the unobserved heterogeneity caused by the extra-variation in total crash counts and moderates the parameters in the joint model. In comparison with the traditional approaches, the proposed weighted risk score approach with the joint model of crash count and crash severity leads to the identification of a higher number of fatal and serious injury crashes in the top ranked sites flagged for safety improvements.

1. Introduction

Crash count and crash severity have been widely used in transport safety as two indicators of crash risk along road segments (Washington et al., 2018). Crash count is a measure of the likelihood of crash occurrence, whereas crash severity is a measure of the societal impact of crashes and their harm to the society. Although federal road agencies and departments of transport aim to reduce crash counts across their road networks, safety improvement programs are primarily focused on preventing fatal and serious injury crashes, as the cost per person of fatal and serious injury crashes are substantially higher than minor injury and non-injury crashes (Harmon et al., 2018). In addition, the collective social cost of crashes in the society is substantially higher than the individual social costs (Tay, 2002). As a result, considering crash severity in conjunction with crash count is paramount for identification of high-risk sites (also referred to as blackspots or hotspots).

https://doi.org/10.1016/j.aap.2020.105615

Received 15 December 2019; Received in revised form 26 February 2020; Accepted 26 May 2020

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).
The two-stage mixed modelling of crash count and crash severity again such, additional complexity in crash severity data, for example unparametric models (such as quantile regression) because EPDO data are crashes of different severity levels because the resulting count model has a hierarchical representation in the Bayesian platform (Shao et al., 2019). Other distributions such as Poisson-Gamma (or Negative Binomial) make the multivariate crash count models computationally burdensome. More importantly, multivariate crash count models neglect the ordered nature of crash severity. The EPDO approach requires non-parametric models (such as quantile regression) because EPDO data are not distributed according to well-known statistical distributions. As such, additional complexity in crash severity data, for example unobserved heterogeneity, is not easily accounted for in this approach. The two-stage mixed modelling of crash count and crash severity again fails to account for the possible correlation between crash counts of different severity levels. These shortcomings render the existing approaches suitable for certain, ‘well behaved’ conditions only, and perhaps leave researchers to develop a rigorous methodology that can capture additional complexity combining crash frequency and injury severities.

Recently, an alternative approach was proposed by Yasin and Eluru (2018) to explicitly incorporate the severity of crashes into crash count prediction. Unlike the two-stage mixed models, this approach is based on a joint (one-stage) modelling of crash count and crash severity, whereby a crash count model is jointly estimated with a crash severity model, and a correlation term is added to the overall model to account for the common factors between the two model components. The overall joint model has two important properties: 1) the parameters of the crash count model are influenced by the parameters of the crash severity model in a joint estimation process, and 2) the overall model parameters are moderated by the correlation between the two model components. Yasin and Eluru (2018) showed that the joint model of crash count and crash severity is superior to traditional crash count models in terms of statistical fit. However, the performance of this model was not evaluated in identification of high-risk sites mainly because a blackspot identification criterion for this model does not exist to date, and thus, its potential for blackspot identification remains unexplored. This study aims to bridge this gap by investigating the applicability and the performance of the recently developed joint model of crash count and crash severity in identifying highway segments with high risk of fatal and injury crashes.

2. literature review

Multivariate modelling of crash counts is one of the most popular approaches in the literature to consider crash severity while identifying high-risk sites. Aguero-Valverde and Jovani (2009) applied a full Bayesian multivariate Poisson lognormal regression to model crashes in Pennsylvania, United States and found a high correlation (from 0.47 to 0.97) between crash counts of different severity levels. They used the costs of different crash severity levels to convert the outcome of the multivariate model to expected crash cost and compared the expected crash cost with the observed crash cost to rank highway segments. They found that high-risk segments identified by the outcome of the multivariate model have consistently higher excess costs than those segments identified by the univariate model. The assumption of a lognormal distribution for the correlation between the crash counts of different severity levels in the multivariate setting resulted in crash counts following a Poisson lognormal distribution. This assumption, however, may not be entirely accurate as crash counts of different severity levels (fatal crashes with excess zero counts for example) may not be Poisson-lognormally distributed (Aguero-Valverde, 2013). In addition, the multivariate model does not consider the ordinal nature of crash severity levels and thus may lead to incorrect inferences about the effects of explanatory variables and result in misidentification of blackspots and inefficient allocation of resources.

EPDO is another popular approach to incorporate crash severity into the selection of high-risk sites. Washington et al. (2014) converted the crash counts of different severity levels to an equivalent property damage only count by applying weights based on average cost ratios by severity. They applied a non-parametric quantile regression on EPDO data in order to estimate the effects of covariates on various quantities of the population, rather than the population mean. Finally, they compared the outcome of the quantile regression model with the observed EPDO and ranked sites accordingly. They found that the high-risk sites identified by the EPDO approach places more emphasis on more harmful crashes than does the conventional count approach. As EPDO data are not distributed like any well-known distributions, non-parametric quantile regression models must be used in this approach. This requirement is a limitation of this EPDO approach as there is no likelihood function to be maximized in this approach and thus the complexity of crash data (e.g. unobserved heterogeneity) cannot be readily incorporated into the modelling process. Recently, some studies have used parametric regression models (e.g. lognormal hurdle and Tobit regression) to model EPDO rate as a function of exogenous covariates (Ma et al., 2016); however, the appropriateness of these distributional assumptions has not been tested.

Finally, two-stage mixed modelling of crash count and crash severity is another approach to consider crash severity for identifying high-risk sites. Miranda-Moreno et al. (2009) first introduced this approach and defined a total risk index as a multiplicative factor of crash count at a site and its expected consequences including expected number of fatal, serious, and minor injuries. They employed Bayesian framework to estimate a hierarchical Poisson model of crash count and a multinomial model of crash severity, independently. They constructed a total risk score per segment and compared it with a standard value established by decision makers and ranked highway segments accordingly. They found that the two-stage mixed model were able to identify high-risk highway segments based on crash count and crash severity and thus improved the effectiveness of allocating resources for safety improvements. In more recent studies, Wang et al. (2011) and Stipancic et al. (2019) used a similar two-stage mixed modelling approach to identify high-risk sites based on crash count and crash severity and found that this approach is superior to the traditional count based approaches. However, the two-stage (independent) modelling of crash count and crash severity in all of these studies crucially ignores the possible correlation between crash counts of different severity levels. This ignorance may result in biased parameters and in turn may lead to incorrect predictions of crash count and crash severity.

Recently, a joint model of crash count and crash severity has been proposed by Yasin and Eluru (2018) which is more flexible than the existing approaches in that it does not require the methodological/empirical assumptions of the above approaches. In particular, this approach is not constrained to any distributional assumptions for the extra-variation in crash counts over and above that accounted for by the Poisson density. The joint model is estimated using maximum likelihood estimation methods and so it can accommodate the complexities associated with crash data, e.g., unobserved heterogeneities. The crash count and crash severity model components are jointly estimated (in one stage) and the parameters of the joint model are moderated by the correlation between crash count and crash severity. As such, the joint model of crash count and crash severity represents a promising alternative to the existing approaches, but importantly, requires further exploration regarding its ability for blackspot identification.
3. Methodology

This section presents the structure of the joint model of crash count and crash severity which was first introduced by Yasmin and Eluru (2018). To describe the model structure more elaborately, the crash count and crash severity model components are presented separately. The joint model of crash count and crash severity is then presented, followed by the selection criteria for identifying highway segments with high risk of fatal and injury crashes. Let $Y_i$ represent the total observed crash count on the $i$th highway segment $(i = 1, 2, 3, \ldots, N)$ and in the $i$th year $(t = 1, 2, 3, \ldots, T)$.

3.1. Crash count model

The transport safety literature has shown that $Y_i$ follows a negative binomial (NB) distribution with mean $\mu_i$ and inverse dispersion parameter $\phi$:

$$Y_i \sim NB(\mu_i, \phi)$$

Assuming an exponential function for the mean of the negative binomial distribution, the total predicted crash count ($\mu_i$) on the $i$th segment and in the $i$th year can be expressed as a function of exogenous explanatory variables (Afghari et al., 2018a):

$$\mu_i = \exp(\sum \beta \phi X_{it} + \varepsilon_i)$$

(1)

where $X_{it}$ are other explanatory variables and $\varepsilon_i$ are random error terms. To account for unobserved heterogeneity, model parameters ($\beta$) are allowed to vary across highway segments (Anastasopoulos and Mannering, 2009). Note that parameters are fixed across time to account for multiple observations on a segment during different time periods (e.g. panel data setting). Such a model specification is referred to as grouped random parameters (Sarwar et al., 2017) in which model parameters are assumed to follow probabilistic distributions (e.g. Normal, Uniform, Triangular, etc.) across the observations within each group (or panel). The probability density function of the RPNB model is:

$$P(Y_{it} = y_{it} | \phi, \beta) = \frac{\Gamma(\theta_i + y_{it})}{\Gamma(\theta_i) \phi^{y_{it}} \phi^{\theta_i}} \left( \frac{\phi \mu_{it}}{\phi + \mu_{it}} \right)^{\theta_i} f(\beta)$$

(2)

where $\Gamma(\cdot)$ is the gamma function and $f(\beta)$ is the density function of the model parameters. The log-likelihood function ($LL$) of the model is obtained by integrating the probability density function of the model over the entire set of random parameters, applying the logarithm transformation and summing it over observations to yield:

$$LL = \sum_{i=1}^{N} \log \left( \int_{\beta} P(Y_{it} = y_{it} | \phi, \beta) d\beta \right)$$

(3)

To estimate this complex log-likelihood function, Maximum Simulated Likelihood Estimation is used where quasi random draws from Halton sequences are employed to simulate the densities of the random parameters (Bhat, 2001). It has been shown that this simulated maximum likelihood estimator is unbiased and consistent for a large number of draws (Munkin and Trivedi, 1999).

3.2. Crash severity model

To incorporate crash severity into the models, let $s$ ($s = 1, 2, 3, \ldots, S$) represent crash severity categories (e.g. property damage only crashes, minor injury crashes, serious injury crashes, fatal crashes, etc.) on highway segments. In ordered models, the actual proportion of total crashes by severity levels ($Y_{it}$) is associated with an underlying latent variable ($Y_{it}^*$. This latent variable is then mapped to the actual severity proportions by thresholds ($\tau$) and using the following linear function:

$$Y_{it}^* = x X_{it} + \delta_i \text{ and } Y_{it} = S \text{ if } \tau_{s-1} < Y_{it}^* < \tau_s$$

(4)

where $x$ is the vector of parameters, $X_{it}$ is the vector of covariates and $\delta_i$ is the random error term.

To estimate the latent propensity of crash severities, it is assumed that:

$$E(Y_{it} | X_{it}) = H_{it}()$$

(5)

where $H_{it}()$ is the probability density function for the severity category $s$. Depending on the distributional assumption for the probability of error terms, $H_{it}()$ can take standard normal or standard logistic probability density functions for the ordered probit or ordered logit models, respectively. The latter functional form is used in this study to construct an ordered logit model for crash severity. The probability of each crash severity category is then presented as:

$$P(Y_{it} = s) = \varphi(\tau_s - (sX_{it})) - \varphi(\tau_{s-1} - (sX_{it}))$$

(6)

where $\varphi(.)$ is the standard logistic cumulative probability density function. The corresponding quasi log-likelihood function is then expressed as:

$$LL = \sum_{i=1}^{N} \log \left( \prod_{s=1}^{S} P(Y_{it} = s)^{w_{is} dc} \right)$$

(7)

where $w_{is}$ is the fraction (proportion between 0 and 1) of crashes in severity category $s$ at road segment $i$ and time period $t$, and the rest of notations are as previously stated. These fractions sum to unity over the categories ($\sum_{s=1}^{S} w_{is} = 1$). This model is referred to as fractional split (Afghari et al., 2018b). Note that $w_{is}$ takes binary values (0 or 1) in conventional logit models; one for the chosen alternative and zero for the non-chosen alternative. Maximum simulated likelihood approach is used to estimate this log-likelihood function.

3.3. Joint model of crash count and crash severity

To generate the correlation between the above mentioned ordered severity model and the previously described count model, a correlation term ($\gamma_{it}$) is now defined as:

$$\gamma_{it} = \gamma m_i$$

(8)

where $\gamma$ is the vector of parameters and $m_i$ is the vector of covariates capturing the observed correlation between crash count and crash severity components. This correlation term is then added to the propensity function of the severity model and the predicted mean of the count model:

$$Y_{it}^* = x X_{it} + \delta_i + \gamma_{it}$$

(9)

$$\mu_{it} = \exp(\sum \hat{\beta} X_{it} + \varepsilon_i + \gamma_{it})$$

(10)

The log-likelihood function of the overall joint model of crash count and severity is then expressed as:

$$LL = \sum_{i=1}^{N} \log \left( \prod_{s=1}^{S} P(Y_{it} = y_{it} | \phi, \beta, \gamma, \eta) \times \prod_{s=1}^{S} P(Y_{it} = s)^{w_{is} dc} \right)$$

(11)

where $\Omega$ represents the vector of random parameters in the joint model ($\beta$ and $x$) and the rest of the notations are as previously stated. Maximum simulated likelihood approach is used to estimate the joint econometric model.

3.4. Weighted risk score to combine crash count and crash severity

The estimated crash count and crash severity models can be used to derive the crash count predictions of different severity levels on road segments by multiplying the total crash count and the proportions of
each severity level:

\[ \mu_{it} = \mu_i \times P(Y_{it} = s) \]  

(12)

A new weighted risk score (WRS) is now developed based on the cost ratio of crash severity levels:

\[ WRS_i = \sum_{s=1}^{s} c_r \mu_{it} \]  

(13)

where WRS, is the predicted weighted risk score for segment \( t \) and \( c_r \) is the ratio of the cost of a crash in severity level \( s \) to the cost of a reference crash severity level. This weighted risk score is analogous to the property damage only equivalency factor in the EPDO approach (Washington et al., 2014) but the weights are applied to the predicted crash counts post estimation and thus the distribution of observed crash counts are intact. This weighted risk score can now be used in ranking highway segments based on their crash count and crash severity.

3.5. Selection criteria: potential for improvement and excess weighted risk score

Two selection criteria are used in this study for identifying high-risk sites: potential for improvement (PFI) following the Empirical Bayesian (EB) method with the independent crash count model and excess weighted risk score for the joint model of crash count and crash severity. The EB approach has widely been used in the literature to identify high-risk sites (Afghari, 2019). The EB approach combines the predicted and observed number of crashes and accounts for the regression to the mean effect (Cheng and Washington, 2005; Washington et al., 2018). Many studies have shown that the EB approach is superior to alternative approaches in identifying high-risk sites (Montella, 2010). The EB estimator is a weighted sum of the predicted and the observed crash counts such that:

\[ EB_i = \omega_1 Y_i + \omega_2 \mu_{it} \]  

(14)

\[ \omega_1 = \frac{\mu_i}{(\mu_i + \frac{k}{\omega})} \quad \text{and} \quad \omega_2 = \frac{1}{k(\mu_i + \frac{k}{\omega})} \]

where \( k \) is the over-dispersion parameter estimated during the SPF calibration process, and \( \omega_1 \) and \( \omega_2 \) are weights calculated based on the mean and variance of the SPF estimate (Persaud et al., 2010). Potential for improvement (PFI) has been used as the selection criteria for identifying high-risk sites based on their EB estimate (Washington et al., 2018). PFI is defined as the difference between the EB estimate and the predicted mean of crash count at a site:

\[ PFI_i = EB_i - \mu_{it} \]  

(15)

Highway segments are ranked according to decreasing PFI and segments with higher PFI are identified as high-risk sites.

Despite its appealing properties, the EB approach rests on distributional assumptions about crash occurrence (Poisson-Gamma or negative Binomial) which do not hold for the weighted risk score approach in this study. As such, excess weighted risk score is used as the selection criteria for ranking highway segments based on crash count and crash severity. The excess weighted risk score (EWRS) is defined as:

\[ EWRS_i = WRS_i - WRS_{\text{avg}} \]  

(16)

where \( WRS_i \) and \( WRS_{\text{avg}} \) are observed and predicted weighted risk scores. Segments are ranked according to the decreasing \( EWRS_i \) and segments with higher \( EWRS_i \) are identified as high-risk sites.

4. Empirical data

The data used to compare the performance of candidate models were collected for a random sample of highway segments and major arterial road sections (rural and urban) along the state-controlled roads in Queensland, Australia. The extent of the studied network is 1,477 km consisting of 521 road segments. The dataset includes crash data as well as roadway geometric and traffic characteristics data. Four years of crashes (from 2010 to 2013) were collected in three severity categories: fatal, serious injury and minor injury. The Queensland Department of Transport and Main Roads stopped collecting property damage only crash data in 2010 and so crashes in this category are not included in this study. Descriptive statistics of the crash data are presented in Table 1.

Roadway geometric and traffic operational characteristics include Annual Average Daily Traffic (AADT), percentage of heavy vehicle traffic, segment length, number of lanes, lane width, functional classification of the road (urban/rural), radius of horizontal curves, degree of horizontal curves, general terrain (vertical alignment), pavement seal conditions, speed limit, level of service, pavement roughness and rutting conditions. Furthermore, presence of shoulder, shoulder type (paved/unpaved), shoulder width, presence of shoulder marking, presence of median, median type (paved/unpaved), median width, and presence of median marking were extracted manually for these 521 segments and added to the dataset. Dummy values were assigned to the categorical variables for the functional classification of the road, presence of shoulder, shoulder type, presence of shoulder marking, presence of median, median type, presence of median marking, general terrain, speed limit, level of service and pavement seal conditions of road segments. The roadway geometric and traffic characteristics data were merged with crash data based on spatial coordinates of crashes and road segments. Table 2 presents descriptive statistics of explanatory variables used in this study.

5. Results and discussion

The crash count model (Eq. 3) and the crash severity model (Eq. 7) were first estimated to be used as the baseline for comparison purposes. The joint model of crash count and crash severity (Eq. 11) was then estimated and compared with the independent crash count and crash severity models. In all models, explanatory variables were tested for multicollinearity by computing the Pearson correlation coefficients, and the variables with unacceptably high (> 0.7) correlation coefficients were excluded from the models. In addition, AADT and segment length were used as the measures of exposure for both models. The relationship between these two variables and crash counts has been extensively studied in the road safety literature. While some research findings indicate that AADT and segment length have linear relationship with crash counts (Geedipally et al., 2009; Qin et al., 2005) and thus may be

Table 1

| Variable                  | Sum   | Mean  | St. D. | Minimum | Maximum |
|---------------------------|-------|-------|--------|---------|---------|
| Fatal crashes             | 84    | 0.040 | 0.202  | 0.000   | 2.000   |
| Fatal crashes in 2010     | 20    | 0.038 | 0.192  | 0.000   | 1.000   |
| Fatal crashes in 2011     | 16    | 0.031 | 0.173  | 0.000   | 1.000   |
| Fatal crashes in 2012     | 28    | 0.054 | 0.234  | 0.000   | 2.000   |
| Fatal crashes in 2013     | 20    | 0.038 | 0.202  | 0.000   | 2.000   |
| Serious injury crashes    | 5057  | 2.427 | 3.822  | 0.000   | 43.000  |
| Serious injury crashes in 2010 | 1245 | 2.390 | 3.713  | 0.000   | 35.000  |
| Serious injury crashes in 2011 | 1210 | 2.322 | 3.704  | 0.000   | 38.000  |
| Serious injury crashes in 2012 | 1227 | 2.355 | 3.636  | 0.000   | 35.000  |
| Serious injury crashes in 2013 | 1375 | 2.639 | 4.209  | 0.000   | 43.000  |
| Minor injury crashes      | 991   | 0.476 | 0.996  | 0.000   | 12.000  |
| Minor injury crashes in 2010 | 313  | 0.601 | 1.138  | 0.000   | 8.000   |
| Minor injury crashes in 2011 | 277  | 0.532 | 1.099  | 0.000   | 12.000  |
| Minor injury crashes in 2012 | 221  | 0.424 | 0.999  | 0.000   | 9.000   |
| Minor injury crashes in 2013 | 180  | 0.345 | 0.779  | 0.000   | 6.000   |
| Total crashes             | 7252  | 2.942 | 4.573  | 0.000   | 51.000  |
| Total crashes in 2010      | 1578  | 3.029 | 4.619  | 0.000   | 42.000  |
| Total crashes in 2011      | 1503  | 2.885 | 4.566  | 0.000   | 51.000  |
| Total crashes in 2012      | 1476  | 2.833 | 4.294  | 0.000   | 40.000  |
| Total crashes in 2013      | 1575  | 3.023 | 4.810  | 0.000   | 48.000  |
included in models as offset variables, other findings indicate that there is a non-linear relationship between these two variables and crash counts (Geedipally et al., 2009; Qin et al., 2004; Caliendo et al., 2007) and that the non-linear relationship is warranted by estimating parameters for these two variables. In the dataset used for this study, the scatterplots of ‘Crash counts versus AADT’ and ‘Crash counts versus segment length’, as presented in Fig. 1, confirm the possible non-linear relationships between these two variables and crash counts.

In addition, Anastasopoulos and Mannering (2009) have shown that estimating a parameter for segment length may reflect the boundary effect of road segmentation—crash counts may be clustered at the boundary of road segments because of a sudden change of roadway geometry. In accordance with the latter rationale, the logarithm of AADT and segment length are used as explanatory variables with estimable parameters (coefficient) in the models in this study to account for possible non-linear relationship between these two measures of exposure and crash count, and to account for potential boundary effects of road segmentation.

5.1. Baseline: independent models of crash count and crash severity

The grouped random parameters negative binomial crash count model and the fractional split ordered logit crash severity model were estimated separately using maximum simulated likelihood approach with 800 Halton draws. The required number of Halton draws was selected so that further increasing the number of draws does not change the estimates significantly. The estimated parameters of these models are presented in Table 3 and Table 4.

The results of the crash count model (Table 3) show that among all of the contributing factors to crash occurrence, ten factors are significant with at least 95 % confidence. The positive parameters of logarithm of AADT (0.286) and segment length (0.556) indicate that the likelihood of crash occurrence increases with increased exposure to crashes. However, the effect of AADT varies significantly across road segments with mean 0.286 and standard deviation 0.054. The positive parameters of number of lanes (0.283), presence of shoulder (0.063), level of service (0.087) and medium speed limit (0.598) indicate that these factors have increasing effects on total crash counts. The increasing effect of number of lanes might be due to aggregation of crashes by crash type in our study. Previous research has shown that the number of lanes have increasing effect on lane-changing related crashes along road segments (Venkataraman et al., 2014). In addition, the

Table 2
Descriptive statistics of explanatory variables used in this study.

| Variable                          | Mean     | St. D.   | Minimum | Maximum |
|-----------------------------------|----------|----------|---------|---------|
| Average annual daily traffic      | 23482.960| 28038.900| 34.000  | 146357.000|
| Fraction of heavy vehicle traffic| 0.098    | 0.086    | 0.000   | 0.965   |
| Length (km)                       | 2.834    | 3.195    | 0.080   | 20.903  |
| Number of lanes                   | 3.070    | 0.420    | 2.000   | 5.000   |
| Lane width (m)                    | 0.910    | 1.470    | 0.000   | 15.000  |
| Median width (m)                  | 4.000    | 4.950    | 0.000   | 38.000  |
| Radius of horizontal curve (m)    | 12096.480| 5880.050 | 127.800 | 41218.000|
| Degree of horizontal curve        | 0.817    | 1.047    | 0.053   | 10.517  |
| Roughness (mm/km)                 | 50.490   | 33.960   | 0.000   | 150.000 |
| Rutting                           | 2.810    | 2.010    | 0.000   | 9.600   |

| Categorical Variables            | Observation | Frequency | Sample Share |
|-----------------------------------|-------------|-----------|--------------|
| Functional classification of road | 172         | rural     | 33 %         |
| Presence of shoulder              | 318         |           | 61%          |
| Shoulder type – paved             | 500         |           | 96%          |
| Presence of shoulder marking      | 77          |           | 15%          |
| Presence of divided median        | 232         |           | 44%          |
| Presence of median marking        | 97          |           | 19%          |
| High speed limit (> 100 km/h)     | 103         |           | 20%          |
| Medium speed limit (> 50 and < 100 km/h) | 292      |           | 56%          |
| Low speed limit (< 50 km/h)       | 126         |           | 24%          |
| Terrain – rolling and/or mountainous | 57       |           | 11%          |
| Pavement seal conditions – sealed | 396        |           | 76%          |
| LOS – E or F                      | 79          |           | 15%          |

**NOTE:** MJ = megajoules.

Fig. 1. Scatterplots of Average Annual Daily Traffic (AADT) and segment length versus crash counts in the dataset used for this study.
increasing effect of presence of shoulder might be related to unsealed surface of the shoulder and other shoulder characteristics (e.g. edge drop). Past research has shown that unsealed shoulder is associated with increased crash counts (Cairney and McGann, 2000; Meuleners drop). Past research has shown that unsealed shoulder is associated with increased crash counts (Cairney and McGann, 2000; Meuleners drop). Past research has shown that unsealed shoulder is associated with increased crash counts (Cairney and McGann, 2000; Meuleners drop).

Segment length, number of lanes, rural roads, presence of shoulder, median width and level of service are the statistically significant variables within the crash count model component with their parameter estimates slightly moderated than their counterparts in the independent count model. In addition, the terrain of the road is statistically significant with a negative parameter (-0.063) indicating that rolling and mountainous terrain is associated with less number of crashes and implying that drivers are more cautious in such circumstances. Sealed pavement, radius of horizontal curves and rural roads are the statistically significant variables within the crash severity model component. The negative parameter of sealed pavement indicates that sealed pavement is associated with decreased severity of crashes whereas the positive parameter of rural roads indicates that these roads are associated with increased severity of crashes. The positive parameter of radius of horizontal curves indicates that larger radius of curves (i.e. sharper curves) is associated with more injuries. This finding might suggest that drivers are more cautious along sharper curves. Numerous studies (e.g., Schneider et al., 2009; Anastasopoulos et al., 2012; Fitch and Hanowski, 2015; Oviedo-Trespalacios et al., 2018, 2019) have reported that the complexity of the road geometry triggers risk-compensating behaviour among drivers and reduces crash risk.

AADT, percentage of heavy vehicle traffic, medium speed limit and pavement roughness are the statistically significant explanatory variables that are common between the two model components and have plausible parameter estimates. The positive parameters of the logarithm of AADT (0.237), medium speed limit (0.393) and roughness (0.149) imply that these variables have increasing effects on crash count and crash severity. The negative parameter of percentage of heavy vehicle traffic (-0.222), rural functional classification (-0.243), median width (-0.095) and sealed pavement (-0.200) have decreasing effects on total crash counts. The varying parameter for percentage of heavy vehicle traffic (mean = -0.222 and standard deviation = 0.395) shows that this factor has heterogeneous effect across road segments. Finally, the inverse dispersion parameter indicates that total number of crashes are over dispersed and are correctly modelled by negative binomial specification.

The results of the crash severity model (Table 4) show that among all factors, only functional classification as the rural road and posted speed limit more than 100 km/h are significantly associated with crash severities. The positive parameters of these two variables are intuitive, indicating that these variables have increasing effect on the severity of crashes along road segments.

### 5.2. Joint model of crash count and crash severity

The joint model of crash count and crash severity was also estimated using 800 Halton draws in the maximum simulated likelihood estimation. Again, the number of Halton draws was selected to guarantee stability of the estimates. The estimated parameters of this joint model are presented in Table 5.

Results show that the joint model of crash count and crash severity consists of a different combination of explanatory variables compared to the crash count and crash severity models. Among the statistically significant explanatory variables, some are unique to each model component and some are common between the two model components (capturing their correlation).

On the contrary, percentage of heavy vehicle traffic (-0.222), rural functional classification (-0.243), median width (-0.095) and sealed pavement (-0.200) have decreasing effects on total crash counts. The varying parameter for percentage of heavy vehicle traffic (mean = -0.222 and standard deviation = 0.395) shows that this factor has heterogeneous effect across road segments. Finally, the inverse dispersion parameter indicates that total number of crashes are over dispersed and are correctly modelled by negative binomial specification.

The results of the crash severity model (Table 4) show that among all factors, only functional classification as the rural road and posted speed limit more than 100 km/h are significantly associated with crash severities. The positive parameters of these two variables are intuitive, indicating that these variables have increasing effect on the severity of crashes along road segments.

### 5.3. Comparison of goodness-of-Fit

The goodness-of-fit measures associated with the count model and
the count component of the joint model shows that the latter model has lower MSPE and MAD (8.287 and 1.825, reduced from 43.230 and 1.990 respectively) and thus has improved statistical fit. In addition to MSPE and MAD, the cumulative residual plots are also plotted against increasing order to AADT to shed more light on the statistical fit of the two models. Cumulative residual plots are helpful tools in demonstrating a model fit with respect to its covariates and identifying potential and systematic bias e.g. over/under prediction (Hauer, 2015). A superior fit occurs when the plots oscillate close to zero. Excess oscillations above/under the zero axis, on the other hand, are a sign of under/over prediction. In addition, a less biased model has an approximately equal amount of positive and negative residuals. Fig. 2 presents the cumulative residual plots (adjusted to terminate at zero) for the two models and shows that the crash count component of the joint model oscillates substantially closer to zero, maintaining more balance between the positive and negative sides and staying closer within the 95% boundaries of cumulative residuals.

The substantially improved goodness-of-fit and cumulative residuals plot of the count component of the joint model suggests that for this sample of data, at least, the joint model is preferred for predicting crash counts for the identification of high-risk sites.

### 5.4. Identification of high-risk sites

To identify high-risk sites across the network, the estimated models were used by three blackspot identification methods: (i) potential for improvement approach with a crash count model, (ii) excess weighted risk score approach with a two-stage mixed model of crash count and crash severity, and (iii) excess weighted risk score approach with a joint model of crash count and severity (Fig. 3).

More specifically, the crash count model was first used to determine the total predicted crash counts at individual road segments. These

*Numbers inside brackets are the estimates for the standard deviations of random parameters.

**Log-Likelihood of the count component within the joint model.

### Table 5

Model results of the joint model of crash count and crash severity.

| Variable | Crash Count Model Component | Crash Severity Model Component |
|----------|-----------------------------|-------------------------------|
|          | Estimate | Standard Error | t-stat | p-value | Estimate | Standard Error | t-stat | p-value |
| Unique Parameters | | | | | | | | |
| Constant | 0.588 | 0.034 | 17.474 | 0.000 | - | - | - | - |
| Threshold Parameters: | | | | | | | | |
| Threshold: minor injury and serious injury | - | - | - | - | -1.925 | 0.084 | -22.946 | 0.000 |
| Threshold: serious injury and fatal | - | - | - | - | 4.210 | 0.200 | 21.007 | 0.000 |
| Segment length | 0.535 | 0.025 | 21.468 | 0.000 | - | - | - | - |
| Number of lanes | 0.250 | 0.030 | 7.566 | 0.000 | - | - | - | - |
| Rural road | -0.354 | 0.030 | -11.990 | 0.000 | - | - | - | - |
| Presence of shoulder | 0.073 | 0.024 | 3.012 | 0.000 | - | - | - | - |
| Median width | -0.076 | 0.026 | -2.891 | 0.000 | - | - | - | - |
| Level of service = E or F | 0.048 | 0.023 | 2.131 | 0.033 | - | - | - | - |
| Terrain: mountainous and rolling | -0.063 | 0.022 | -2.861 | 0.004 | - | - | - | - |
| Sealed pavement | - | - | - | - | -0.418 | 0.094 | -4.436 | 0.000 |
| Radius of horizontal curve | - | - | - | - | 0.324 | 0.084 | 3.870 | 0.000 |
| Inverse dispersion parameter (φ) | 4491.761 | 104.043 | 43.172 | 0.000 | - | - | - | - |
| Common Parameters | | | | | | | | |
| Log(average annual daily traffic) | 0.237 (0.054) | 0.029 (0.020) | 8.126 (2.705) | 0.000 (0.007) | - | - | - | - |
| Percent of heavy vehicle traffic | -0.186 (0.248) | 0.032 (0.070) | -5.870 (3.565) | 0.000 (0.000) | - | - | - | - |
| Medium speed limit (> 60 km/hr and < 100 km/h) | 0.393 (0.650) | 0.036 (0.028) | 10.835 (-23.209) | 0.000 (0.000) | - | - | - | - |
| Roughness | 0.149 | 0.041 | 3.641 | 0.000 | - | - | - | - |
| Measures of Fit | | | | | | | | |
| Log-Likelihood | -4686.048 [-3898.283**] | 8.287 | 1.825 | 0.000 | - | - | - | - |

![Fig. 2. Cumulative residual plots against Average Annual Daily Traffic (AADT) for the independent count model vs count component of the joint model.](image)

**Fig. 2.** Cumulative residual plots against Average Annual Daily Traffic (AADT) for the independent count model vs count component of the joint model.
predicted crash counts were then used with the PFI approach to identify high-risk sites. The crash severity model was then used to determine the predicted probabilities of each severity level. The product of these probabilities and the previously mentioned predicted crash counts were then calculated to determine the predicted crash counts by severity while taking into account the correlation between different severity levels, and were used with the excess weighted risk score approach to identify high-risk sites. The weighted risk score for predicted crash counts of different severity levels (either from the two-stage mixed model or from the joint model) were calculated using the human capital costs (i.e., social costs) associated with Queensland crashes collected from Austroads Guide to Road Safety Part 8 (Austroads, 2015). These costs vary depending on the functional classification of the road and thus their weighted averages were used for the road network in this study. Table 6 shows the original crash costs and their weighted average across the network.

Road segments were ranked by total decreasing potential for improvement (based on the crash count model) and total decreasing excess weighted risk score (based on the two-stage mixed model and the joint model of crash count and crash severity) for the period of 2010-2013.

### 5.4.1. Potential for improvement vs excess weighted risk score

To illuminate the importance of incorporating crash severity into blackspot identification, the PFI approach is first compared with the EWRS approach. Table 7 shows the top 20 road segments (out of 521 road segments) identified as high-risk sites by the two approaches. The results show that 7 out of 20 road segments (shaded cells in Table 7) are unique to the blackspot identification approach and are not commonly identified by the two approaches. The sites that are unique to the risk score approach have higher number of fatal crashes compared to the sites that are unique to the potential for improvement approach. As reported in Table 7, the ranking of sites by two approaches is different. As expected, the risk score approach puts more emphasis on fatal crashes in ranking sites, whereas the potential for improvement approach puts more emphasis on the total crashes. To obtain a more tangible understanding of the performance of these two approaches in identifying high-risk sites, the total number of crashes identified across the top 20 sites were also used for the comparison between the two approaches (Afghari et al., 2016).

The PFI approach identified 1356 crashes whereas the excess weighted risk score approach identified 1178 crashes along the top 20 sites. However, the risk score approach identified 19 fatal crashes whereas the potential for improvement approach identified 14 fatal crashes along those sites. In addition, the potential for improvement approach identified higher number of minor and serious injury crashes compared to the excess weighted risk score approach. These findings show that, in general, the independent crash count model together with the Empirical Bayesian approach leads to the identification of higher number of crashes in the top ranked sites. However, if the rationale is to identify high-risk sites with high fatal crashes, then the joint model of crash count and severity together with the risk score approach represents a better alternative. Another important consideration in blackspot identification is that excess crashes are most treatable (Cheng and Washington, 2008). To shed more light on this finding, excess crashes are calculated by subtracting predicted crashes from observed crashes along the top 20 sites (Table 8)—suggesting that installing improvements to reduce crashes in excess of ‘typical’ have much higher chance of success than expecting crash improvements beyond ‘typical’

### Table 6

Human capital cost of crashes by injury severity level for the state of Queensland in Australia (Australian dollars).

| Injury Severity | Rural Cost | Urban Cost | Weighted Average Cost | Weights |
|-----------------|------------|------------|-----------------------|---------|
| Fatal crash     | $2,728,617 | $2,456,691 | $2,546,427            | 104.189 |
| Serious injury  | $642,034   | $595,802   | $611,057              | 25.002  |
| Minor injury    | $25,822    | $23,760    | $24,440               | 1.000   |

### Table 7

Observed crashes along top 20 high-risk sites identified by the Potential for Improvement (PFI) versus Excess Weighted Risk Score (EWRS) approach.

| Rank | Site ID | Total excess crashes | Minor injury excess crashes | Serious injury excess crashes | Fatal excess crashes | Rank | Site ID | Total excess crashes | Minor injury excess crashes | Serious injury excess crashes | Fatal excess crashes |
|------|---------|----------------------|----------------------------|-------------------------------|----------------------|------|---------|----------------------|----------------------------|----------------------------|----------------------|
| 1    | 62      | 130                  | 23                         | 106                           | 1                    | 1    | 333     | 29                   | 151                         | 106                        |                     |
| 2    | 482     | 90                   | 16                         | 73                            | 1                    | 2    | 214     | 20                   | 166                         | 73                         |                     |
| 3    | 333     | 181                  | 29                         | 151                           | 1                    | 3    | 482     | 16                   | 106                         | 73                         |                     |
| 4    | 214     | 129                  | 20                         | 166                           | 3                    | 4    | 400     | 6                    | 47                          | 41                         |                     |
| 5    | 191     | 72                   | 8                          | 64                            | 0                    | 5    | 191     | 8                    | 72                          | 64                         |                     |
| 6    | 155     | 76                   | 15                         | 58                            | 3                    | 6    | 155     | 15                   | 76                          | 58                         |                     |
| 7    | 21      | 49                   | 14                         | 34                            | 1                    | 7    | 21      | 14                   | 49                          | 34                         |                     |
| 8    | 218     | 49                   | 8                          | 41                            | 0                    | 8    | 21      | 8                    | 49                          | 34                         |                     |
| 9    | 250     | 64                   | 12                         | 52                            | 0                    | 9    | 218     | 8                    | 49                          | 34                         |                     |
| 10   | 455     | 58                   | 10                         | 48                            | 0                    | 10   | 147     | 6                    | 34                          | 27                         |                     |
| 11   | 227     | 46                   | 8                          | 38                            | 0                    | 11   | 401     | 7                    | 31                          | 24                         |                     |
| 12   | 460     | 48                   | 5                           | 43                            | 0                    | 12   | 271     | 1                    | 25                          | 23                         |                     |
| 13   | 167     | 45                   | 7                           | 38                            | 0                    | 13   | 253     | 1                    | 22                          | 21                         |                     |
| 14   | 36      | 44                   | 10                          | 34                            | 0                    | 14   | 35      | 6                    | 58                          | 51                         |                     |
| 15   | 422     | 54                   | 10                          | 43                            | 1                    | 15   | 22      | 9                    | 33                          | 23                         |                     |
| 16   | 400     | 47                   | 6                            | 41                            | 0                    | 16   | 437     | 3                    | 26                          | 23                         |                     |
| 17   | 375     | 40                   | 5                            | 35                            | 0                    | 17   | 153     | 3                    | 19                          | 13                         |                     |
| 18   | 35      | 58                   | 6                            | 51                            | 1                    | 18   | 460     | 5                    | 48                          | 43                         |                     |
| 19   | 147     | 34                   | 6                            | 27                            | 1                    | 19   | 167     | 7                    | 45                          | 38                         |                     |
| 20   | 15      | 42                   | 10                           | 31                            | 1                    | 20   | 61      | 1                    | 14                          | 11                         |                     |

Shaded: sites that are uniquely identified by each approach.
Shaded: sites that are uniquely identified by each approach.

The results show that out of 1356 crashes identified by the potential for improvement approach, only 408 crashes are excess (30.0 %) whereas out of 1178 crashes identified by the excess weighted risk score approach, 618 are excess (52.4 %). This finding shows that the excess weighted risk score approach leads to the identification of more treatable crashes along the top ranked sites. The ranking of sites also paints a similar picture of model performance. The top ranked sites identified by the risk score approach have substantially higher excess crashes than the top ranked sites identified by the potential for improvement approach (for example, see sites of rank 1, rank 3, rank 4 and rank 5 in Table 8). Moreover, the excess weighted risk score approach provides additional information about excess crashes by severity levels which cannot be obtained by the other approach. This additional information indicate that the majority of the excess crashes (498 out of 618) are fatal and serious injury crashes.

5.4.2. Two-stage mixed model vs joint model
As illustrated previously, the PFI approach identifies higher total crashes in high rank sites while the EWRS approach identifies crashes with higher severities. However, this comparison may not provide a thorough understanding of the blackspot identification performance of the joint model within the EWRS approach because an approach that only looks into total crashes is not going to identify as many fatal and severe injury crashes compared to an approach that emphasizes severe crashes, regardless of model specification. As a result, the EWRS approach is selected as the blackspot identification approach and the performance of the joint model is now compared with that of a two-stage (independent) mixed model of crash count and crash severity. Table 9 and Table 10 show the top 20 road segments identified by the EWRS approach using these two model specifications.

The results show that 4 out of 20 road segments (shaded cells in Table 9 and Table 10) are unique to the model specification and are not commonly identified by the two models. The sites that are unique to the joint model have higher number of fatal crashes compared to the sites that are unique to the two-stage mixed model. In addition, the ranking of sites by these two models is different. The two-stage mixed model identified 1257 crashes whereas the joint model identified 1178 crashes along the top 20 sites. The joint model identified 19 fatal crashes whereas the two-stage mixed model identified 14 fatal crashes along those sites. Similar to the PFI approach with the crash count model, the EWRS approach with the two-stage mixed model identified higher number of minor and serious injury crashes compared to the EWRS approach with the joint model. However, a closer look at ‘excess crashes’ revealed the true benefit of the joint model (Table 10). Out of 1257 crashes identified by the two-stage mixed model, only 586 crashes are excess (46.6 %), whereas out of 1178 crashes identified by the joint model, 618 are excess (52.4 %). This finding shows that the joint model leads to the identification of more treatable crashes along the top ranked sites. The ranking of sites also paints a similar picture of model performance.

The above findings together illustrate the true benefit of the joint model of crash count and crash severity when using within the excess weighted risk score approach in that it ultimately leads to the identification of higher number of treatable fatal and serious injury crashes. This benefit is the direct consequence of applying weights to the predicted crash counts by severity, and accounting for the correlation between the predicted crash counts of different severity which is in turn the consequence of employing the joint model of crash count and crash severity.

6. Conclusions
Total crash counts at transport locations have widely been used to identify high-risk sites across the network. However, the societal impact of crashes and their harm to the society is not accounted for in the existing approaches because crash severity has not been well incorporated into the traditional blackspot identification method. This study aimed to fill this gap by using the joint modelling approach to incorporate crash severity into crash count prediction and identify road segments with high risk of fatal and serious injury crashes.

Findings indicate that the incorporation of crash severity into the crash count model and capturing the correlation between crash counts of different severity levels improve the accuracy (i.e. statistical fit) of crash count predictions. In addition, accounting for such correlations influences the over dispersion of crash data and the statistical significance of explanatory variables. The joint model of crash count and crash severity provides unique information about the probability of
Table 9
Observed crashes along top 20 high-risk sites identified by the Excess Weighted Risk Score (EWRS) approach using two-stage mixed vs joint models of crash count and crash severity.

| Rank | Site ID | Total excess crashes | Minor injury excess crashes | Serious injury excess crashes | Fatal excess crashes | Top sites identified by the two-stage mixed model | Total excess crashes | Minor injury excess crashes | Serious injury excess crashes | Fatal excess crashes |
|------|---------|----------------------|-----------------------------|------------------------------|---------------------|-----------------------------------------------|---------------------|-----------------------------|------------------------------|---------------------|
| 1    | 62      | 130                  | 23                          | 106                          | 1                   | 1                                             | 333                 | 181                         | 29                           | 151                 |
| 2    | 214     | 129                  | 20                          | 106                          | 3                   | 2                                             | 214                 | 129                         | 20                           | 106                 |
| 3    | 482     | 90                   | 16                          | 73                           | 1                   | 3                                             | 62                  | 130                         | 23                           | 106                 |
| 4    | 400     | 47                   | 6                           | 41                           | 0                   | 4                                             | 482                 | 90                          | 16                           | 73                  |
| 5    | 191     | 72                   | 8                           | 64                           | 0                   | 5                                             | 400                 | 47                          | 6                            | 41                  |
| 6    | 155     | 76                   | 15                          | 58                           | 3                   | 6                                             | 191                 | 72                          | 8                            | 64                  |
| 7    | 333     | 181                  | 29                          | 151                          | 1                   | 7                                             | 155                 | 76                          | 8                            | 64                  |
| 8    | 21      | 49                   | 14                          | 34                           | 1                   | 8                                             | 21                  | 49                          | 8                            | 41                  |
| 9    | 218     | 49                   | 8                           | 41                           | 0                   | 9                                             | 218                 | 49                          | 8                            | 41                  |
| 10   | 460     | 48                   | 5                            | 43                           | 0                   | 10                                            | 460                 | 48                          | 5                            | 43                  |
| 11   | 401     | 31                   | 7                            | 24                           | 0                   | 11                                            | 401                 | 31                          | 7                            | 24                  |
| 12   | 147     | 34                   | 6                            | 27                           | 1                   | 12                                            | 147                 | 34                          | 6                            | 27                  |
| 13   | 227     | 46                   | 8                            | 38                           | 0                   | 13                                            | 227                 | 46                          | 8                            | 38                  |
| 14   | 271     | 25                   | 1                            | 23                           | 1                   | 14                                            | 271                 | 25                          | 1                            | 23                  |
| 15   | 35      | 58                   | 6                            | 51                           | 1                   | 15                                            | 35                  | 58                          | 6                            | 51                  |
| 16   | 167     | 45                   | 7                            | 38                           | 0                   | 16                                            | 167                 | 45                          | 7                            | 38                  |
| 17   | 253     | 22                   | 1                            | 21                           | 1                   | 17                                            | 253                 | 22                          | 1                            | 21                  |
| 18   | 340     | 31                   | 4                            | 27                           | 0                   | 18                                            | 340                 | 31                          | 4                            | 27                  |
| 19   | 375     | 40                   | 5                            | 35                           | 0                   | 19                                            | 375                 | 40                          | 5                            | 35                  |
| 20   | 422     | 54                   | 10                           | 43                           | 1                   | 20                                            | 422                 | 54                          | 10                           | 43                  |

Sum of observed crashes across top 20 sites identified by the two-stage mixed model: 1257, 199, 1044, 14.

Table 10
Excess crashes along top 20 high-risk sites identified by the Excess Weighted Risk Score (EWRS) approach using two-stage mixed vs joint models of crash count and crash severity.

| Rank | Site ID | Total excess crashes | Minor injury excess crashes | Serious injury excess crashes | Fatal excess crashes | Top sites identified by the joint model | Total excess crashes | Minor injury excess crashes | Serious injury excess crashes | Fatal excess crashes |
|------|---------|----------------------|-----------------------------|------------------------------|---------------------|----------------------------------------|---------------------|-----------------------------|------------------------------|---------------------|
| 1    | 333     | 181                  | 29                          | 151                          | 1                   | 1                                       | 333                 | 181                         | 29                           | 151                 |
| 2    | 214     | 129                  | 20                          | 106                          | 3                   | 2                                       | 214                 | 129                         | 20                           | 106                 |
| 3    | 482     | 90                   | 16                          | 73                           | 1                   | 3                                       | 482                 | 90                          | 16                           | 73                  |
| 4    | 400     | 47                   | 6                           | 41                           | 0                   | 4                                       | 400                 | 47                          | 6                            | 41                  |
| 5    | 191     | 72                   | 8                           | 64                           | 0                   | 5                                       | 191                 | 72                          | 8                            | 64                  |
| 6    | 155     | 76                   | 15                          | 58                           | 3                   | 6                                       | 155                 | 76                          | 15                           | 58                  |
| 7    | 333     | 181                  | 29                          | 151                          | 1                   | 7                                       | 333                 | 181                         | 29                           | 151                 |
| 8    | 21      | 49                   | 14                          | 34                           | 1                   | 8                                       | 21                  | 49                          | 14                           | 34                  |
| 9    | 218     | 49                   | 8                           | 41                           | 0                   | 9                                       | 218                 | 49                          | 8                            | 41                  |
| 10   | 460     | 48                   | 5                            | 43                           | 0                   | 10                                      | 460                 | 48                          | 5                            | 43                  |
| 11   | 401     | 31                   | 7                            | 24                           | 0                   | 11                                      | 401                 | 31                          | 7                            | 24                  |
| 12   | 147     | 34                   | 6                            | 27                           | 1                   | 12                                      | 147                 | 34                          | 6                            | 27                  |
| 13   | 227     | 46                   | 8                            | 38                           | 0                   | 13                                      | 227                 | 46                          | 8                            | 38                  |
| 14   | 271     | 25                   | 1                            | 23                           | 1                   | 14                                      | 271                 | 25                          | 1                            | 23                  |
| 15   | 35      | 58                   | 6                            | 51                           | 1                   | 15                                      | 35                  | 58                          | 6                            | 51                  |
| 16   | 167     | 45                   | 7                            | 38                           | 0                   | 16                                      | 167                 | 45                          | 7                            | 38                  |
| 17   | 253     | 22                   | 1                            | 21                           | 1                   | 17                                      | 253                 | 22                          | 1                            | 21                  |
| 18   | 340     | 31                   | 4                            | 27                           | 0                   | 18                                      | 340                 | 31                          | 4                            | 27                  |
| 19   | 375     | 40                   | 5                            | 35                           | 0                   | 19                                      | 375                 | 40                          | 5                            | 35                  |
| 20   | 422     | 54                   | 10                           | 43                           | 1                   | 20                                      | 422                 | 54                          | 10                           | 43                  |

Sum of excess crashes across top 20 sites identified by the joint model: 1178, 188, 971, 19.

Shaded: sites that are uniquely identified by each approach.

Each severity level on road segments.

From the pragmatic point of view and in terms of identifying high-risk sites, the joint model of crash count and crash severity enables the analyst to predict crash counts of different severity levels and provides more information about crashes compared to the traditional count model. Such information can be readily used to rank transport locations not only based on the total crashes but also based on the severe crashes such as fatal and serious injury crashes. However, it is important to note that determining which blackspot identification approach is more "useful" in practice may require additional considerations. For example, economic analysis (determining overall benefit and overall return on investment given a fixed budget) of the final selection of high-risk sites and the required countermeasures may change the superiority of one approach over another (Gross et al., 2016). Nonetheless, the weighted risk score approach presented in this study is not a replacement but may serve as a complement to the existing approaches because it provides additional information for identifying high-risk sites.

This study is not without limitations. An implicit assumption in the specification of the joint model of crash count and crash severity is that different severity levels have the same amount of over dispersion. Future research should test whether this assumption is valid. In addition, the performance of the joint model of crash count and crash severity.
severity has not been compared with the other state-of-the-practice models including multivariate count models and equivalent property damage only models. Testing the performance of these models against each other is a worthy research direction.

CRediT authorship contribution statement

Amir Pooyan Afghari: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Software, Validation.
Mdz Mazharul Haque: Conceptualization, Methodology, Supervision, Writing - review & editing.
Simon Washington: Conceptualization, Methodology, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the Queensland University of Technology for providing the funding sources for this study. The authors would also like to acknowledge the Queensland Department of Transport and Main Roads (QTMR) for their support in data collection for this study.

References

Afghari, A.P., 2019. Detecting Motor Vehicle Crash Blackspots Based on Their Underlying Behavioural, Engineering, and Spatial Causes. Doctoral dissertation. University of Queensland, Australia.
Afghari, A.P., Haque, M.M., Washington, S., Smyth, T., 2016. Bayesian latent class safety performance function for identifying motor vehicle crash black spots. Transp. Res. Rec. 2601 (1), 90–98.
Afghari, A.P., Haque, M.M., Washington, S., 2018a. Applying fractional split model to examine the effects of roadway geometric and traffic characteristics on speeding behavior. Traffic Inf. Prev. 19 (8), 860–866.
Afghari, A.P., Washington, S., Haque, M.M., Li, Z., 2018b. A comprehensive joint econometric model of motor vehicle crashes arising from multiple sources of risk. Anal. Methods Accid. Prev. 18, 1–14.
Aguro-Valverde, J., 2013. Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: comparing the precision of crash frequency estimates. Accid. Anal. Prev. 50, 289–297.
Aguro-Valverde, J., Jovais, P.P., 2009. Bayesian multivariate Poisson lognormal models for crash severity modeling and site ranking. J. Transp. Res. Board 2136, 82–91.
Anastasopoulos, P.C., Mannerling, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameter counts models. Accid. Anal. Prev. 41, 153–159.
Anastasopoulos, P.C., Shankar, V.N., Haddock, J.E., Mannerling, F.L., 2012. A multivariate tobit analysis of highway accident-injury-severity rates. Accid. Anal. Prev. 45, 110–119.
Austroads, 2015. Austroads Guide to Road Safety Part 8: Treatment of Crash Locations. Austroads, Sydney, Australia.
Bhat, C.R., 2001. Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transp. Res. Part B Methodol. 35, 677–693.
Cairney, P., McGann, A., 2000. Relationship Between Crash Risk and Geometric Characteristics of Rural Highways (No. AP-R162/00).
Caliendo, C., Guida, M., Parisi, A., 2007. A crash-prediction model for multilane roads. Accid. Anal. Prev. 39, 657–670.
Cheng, W., Washington, S.P., 2005. Experimental evaluation of hotspot identification methods. Accid. Anal. Prev. 37, 870–881.
Cheng, W., Washington, S., 2008. New criteria for evaluating methods of identifying hot spots. Transp. Res. Rec. 2083, 76–85.
Fitch, G.M., Hanowski, R.J., Guo, F., 2015. The risk of a safety-critical event associated with mobile device use in specific driving contexts. Traffic Inj. Prev. 16 (2), 124–132.
Geddipally, S.R., Lord, D., Park, B.-J., 2009. Analyzing different parameterizations of the varying dispersion parameter as a function of segment length. Transp. Res. Rec. 2103, 108–118.
Gross, F.B., Harmon, T., Albee, M., Himes, S., Srinivasan, R., Carter, D., Dugas, M., 2016. Evaluation of Four Network Screening Performance Measures. Report No. FHWA-SA-16-103. Federal Highway Administration. Office of Safety, United States.
Harmon, T., Bahar, G., Gross, F., 2018. Crash Costs for Highway Safety Analysis. Federal Highway Administration Report No. FHWA-SA-17-071.
Hauer, E., 2015. The Art of Regression Modeling in Road Safety. Springer, New York, NY.
Ma, L., Yan, X., Wei, C., Wang, J., 2016. Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework. Anal. Methods Accid. Prev. 11, 48–61.
Meulener, L.B., Hendrie, D., Lee, A.H., 2011. Effectiveness of sealed shoulders and audible edge lines in Western Australia. Traffic Inf. Prev. 12 (2), 201–205.
Miranda-Moreno, L.F., Pu, L., Ukkusuri, S., Lord, D., 2009. How to Incorporate Accident Severity and Vehicle Occupancy into the Hot Spot Identification Process? Transp. Res. Rec. 2102, 53–60.
Montella, A., 2010. A comparative analysis of hotspot identification methods. Accid. Anal. Prev. 42, 571–581.
Munkin, M.K., Trivedi, P.K., 1999. Simulated maximum likelihood estimation of multivariate mixed-Poisson regression models, with application. Econom. J. 2, 29–48.
Oviedo-Trespalacios, O., Haque, M.M., Kim, M., Demmel, S., 2018. Driving behaviour while self-regulating mobile phone interactions: a human-machine system approach. Accid. Anal. Prev. 118, 253–262.
Oviedo-Trespalacios, O., Haque, M.M., Kim, M., Washington, S., 2019. “Matel! I’m running 10 min late”: an investigation into the self-regulation of mobile phone tasks while driving. Accid. Anal. Prev. 122, 134–142.
Persaud, B., Lan, B., Lyon, C., Bhim, R., 2010. Comparison of empirical Bayes and full Bayes approaches for before-after road safety evaluations. Accid. Anal. Prev. 42, 38–43.
Qin, X., Ivan, J.N., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. Accid. Anal. Prev. 36, 183–191.
Qin, X., Ivan, J.N., Ravishanker, N., Liu, J., 2005. Hierarchical Bayesian estimation of safety performance functions for two-lane highways using Markov chain Monte Carlo modeling. J. Transp. Eng. 131, 345–351.
Starw, M.T., Anastasopoulos, P.C., Golshani, N., Hulme, K.F., 2017. Grouped random parameters bivariate probit analysis of perceived and observed aggressive driving behavior: a driving simulation study. Anal. Methods Accid. Prev. Res. 13, 52–64.
Schneider, W.H., Savolainen, P.T., Zimmerman, K., 2009. Driver injury severity resulting from single-vehicle crashes along horizontal curves on rural two-lane highways. Transp. Res. Rec. 2102 (1), 85–92.
Shan, M.R.R., Qin, X., Afghari, A.P., Washington, S., Haque, M.M., 2019. Incorporating behavioral variables into crash count prediction by severity: a multivariate multiple risk source approach. Accid. Anal. Prev. 129, 277–288.
Stipancic, J., Miranda-Moreno, L., Saunier, N., Labbe, A., 2019. Network screening for large urban road networks: using GPS data and surrogate measures to model crash frequency and severity. Accid. Anal. Prev. 125, 290–301.
Tay, R., 2002. Prioritising road safety initiatives: fatality versus social cost. Appl. Health Econ. Health Policy 1, 129–134.
Venkataraman, N., Shankar, V., Ullarson, G.F., Deptuch, D., 2014. A heterogeneity-in-means count model for evaluating the effects of interchange type on heterogeneous influences of interstate geometrics on crash frequencies. Anal. Methods Accid. Res. 2, 12–20.
Wang, C., Quddus, M.A., Ison, S.G., 2011. Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accid. Anal. Prev. 43, 1979–1990.
Washington, S., Haque, M.M., Oh, J., Lee, D., 2014. Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. Accid. Anal. Prev. 66, 136–146.
Washington, S., Afghari, A.P., Haque, M.M., 2018. Detecting High-risk Accident Locations. Safe Mobility: Challenges, Methodology and Solutions. Emerald Publishing Limited, Bingley, UK.
Yasmin, S., Eluru, N., 2018. A joint econometric framework for modeling crash counts by severity. Transp. A Transp. Sci. 14, 230–255.