How to Incorporate Accident Severity and Vehicle Occupancy into the Hot Spot Identification Process?

Luis F. Miranda-Moreno, Liping Fu, Satish Ukkusuri, and Dominique Lord

This paper introduces a Bayesian accident risk analysis framework that integrates accident frequency and its expected consequences into the hot spot identification process. The Bayesian framework allows the introduction of uncertainty not only in the accident frequency and severity model parameters but also in key variables such as vehicle occupancy levels and severity weighting factors. For modeling and estimating the severity levels of each individual involved in an accident, a Bayesian multinomial model is proposed. For modeling accident frequency, hierarchical Poisson models are used. How the framework can be implemented to compute alternative relative and absolute measures of total risk for hot spot identification is described. To illustrate the proposed approach, a group of highway–railway crossings from Canada is used as an application environment.

Because of the deficiencies of accident risk estimates based on raw data, the traffic safety community is interested in the development and application of the risk model–based approach, which makes use of statistical methods based on probability theory. The approach consists of a systematic analysis of the input crash data to develop accident frequency and consequence models from which ranking criteria are built (*1–3*). Once statistical models have been developed from the input data, several Bayesian ranking methods or criteria proposed in the literature can be applied to identify a list of hot spots (*1–8*). These criteria include the posterior expectation of accident frequency, the potential of accident reduction, and the posterior expectation of ranks. These measures usually are based on the assumption that the safety status of a site can be reflected by accident frequency, and severity is usually not incorporated in the analysis or is assumed to be fixed over locations (observed and unobserved severity heterogeneities are ignored across sites). In many applications, however, accident frequency may not completely reveal the total risk level of a site or capture the safety benefits that some safety countermeasures could introduce. For example, in highway–railway networks, some safety measures such as speed limits have a more significant effect on train–vehicle collision severity than on frequency.

Although most published work on hot spot identification focuses mainly on developing accident frequency and consequence models separately, few screening studies have proposed frameworks that integrate both elements in a two-dimensional risk approach including uncertainty in the analysis [examples are Nassar et al. (*9*) and Saccomanno et al. (*10*)]. This approach assumes that accident occurrence at a location is best represented by the product of accident frequency and severity. One way to incorporate accident severity in the analysis is to calibrate statistical models that relate accident consequences to factors such as location configuration, roadway alignment, speed limits, and surface conditions (*9–11*). In this stage, the aim is to identify factors that largely influence the likelihood of fatal or injury outcomes once an accident takes place. For the severity analysis, several statistical model settings have been suggested in the literature, such as the basic logistic regression, multinomial, ordered logit, and mixed logit models [examples are work by Milton et al. (*11*) and Eluru et al. (*12*)]. Alternatively, some studies incorporate accident consequences by simply classifying accident counts by severity type (e.g., fatal and injury and other accident types) [examples are work by Miaou and Song (*6*) and Park and Lord (*13*)]. In this case, a statistical multivariate model setting considering the different categories is implemented (multivariate analysis). Although this approach accounts for correlation among crash counts, the expected crash consequences (for drivers and passengers) do not vary across locations, and vehicle occupancy levels as an important determinant of overall risk exposure are ignored. Note that vehicle occupancy has been an important part of transportation management systems and is used for evaluating high-occupancy-vehicle lanes or congestion reduction strategies (*14*). However, vehicle occupancy levels as a determinant of traffic risk exposure have often been ignored in the implementation and evaluation of traffic safety strategies.

This paper introduces a new hierarchical Bayesian framework to integrate accident frequency, severity, and vehicle occupancy levels in the hot spot identification process. The primary intention is to illustrate the potential effect of incorporating accident severity on the result of the hot spot identification process. For this purpose, a group of highway–railway crossings from Canada is used as an application environment.

TOTAL RISK–BASED APPROACH

In this section, the elements of the proposed Bayesian risk-based methodology are defined, including severity score, accident consequence model, and hot spot identification criteria.

L. F. Miranda-Moreno, Department of Civil Engineering and Applied Mechanics, McGill University, Macdonald Engineering Building, 817 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada. L. Fu, Department of Civil and Environmental Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada. S. Ukkusuri, Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute, 4032 Jonsson Engineering Center, Troy, NY 12180. D. Lord, Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, 317 Gilchrist Building, College Station, TX 77843-3136. Corresponding author: L. F. Miranda-Moreno, luis.miranda-moreno@mcgill.ca.

Transportation Research Record: Journal of the Transportation Research Board, No. 2102, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 53–60.
DOI: 10.3141/2102-07

Severity Score Definition

Total risk is commonly defined as the product of the accident frequency and consequences [see, for example, Saccomanno et al. (*10*)]. That is,

$$\text{TR}_i = \theta_i \times C_i \quad (1)$$

where

TR_i = total risk at site i ($i = 1, \ldots, n$),
θ_i = mean number of accidents, and
C_i = expected consequence caused by an accident taking place at site i.

Previous research has mainly focused on how to estimate accident frequency; the work described here focuses on how to compute C_i. Under a Bayes framework, both θ_i and C_i can be considered as random variables. In addition, because an accident can result in different types of outcomes, such as fatal, major, and minor injuries and property damage, the total consequence should encapsulate at least the major types of outcomes. As a result, C_i is defined as a severity score integrating three components, as follows:

$$C_i = f_{1i} \cdot \omega_1 + f_{2i} \cdot \omega_2 + f_{3i} \cdot \omega_3 \quad (2)$$

where f_{1i}, f_{2i}, and f_{3i} are the expected number of fatal, serious, and minor injuries, respectively, under an accident taking place at site i; ω_1, ω_2, and ω_3 are estimated equivalent monetary costs (or prespecified weights) per fatality, serious injury, and minor injury, respectively. Three examples of these equivalent costs are given in Table 1 (*10, 15*). Note that, in the proposed integration approach, other costs, such as property damage, emergency services, and delays, are ignored—they are usually small or proportional to other injury costs.

The expected number of casualties of a given severity type in Equation 2 can be estimated by multiplying the expected number of passengers per vehicle (estimated motor vehicle occupancy) by the probability that a person involved in an accident suffers that type of severity:

$$f_{ki} = p_{ki} \times h_i \quad (3)$$

where p_{ki} is the probability that a passenger involved in a collision at site i with specific site attributes suffers an injury of type k ($k = 1$, 2, 3) and h_i refers to an average number of passengers involved in an accident. Since in some practical application h_i may be difficult to estimate, a common nominal value (h) could be assumed for different subgroups of sites. On the basis of past vehicle occupancy studies or vehicle occupancies reported in accident records, an analyst should be able to define various occupancy weighting values for various subgroups of sites. For example, heavy weights could be assigned to locations with a high percentage of vehicles with high levels of occupancy (*14*). This will allow vehicle occupancy level to be considered as an important determinant of the overall crash risk exposure in roadway facilities.

TABLE 1 Direct Severity Costs per Person and Weights Assumed in Analysis

Source	Cost Estimates (U.S. \$) Fatality	Serious Injury	Minor Injury	Weights[a] ω_1	ω_2	ω_3
1	2,710,000[b]	65,590	30,000	41	1.0	0.5
2	2,000,000[c]	65,590	30,000	30	1.0	0.5
3	1,000,000[d]	65,590	30,000	15	1.0	0.5

[a]The weights are relative to the serious injury cost. The serious and minor injury costs are the same as the ones proposed by Saccomanno et al. (*10*). The weights are obtained by dividing each cost by the serious injury cost.
[b]Cost employed by Saccomanno et al. (*10*).
[c]Cost reported by Zaloshnja et al. (*15*).
[d]Cost designated for U.S. agencies (e.g., National Safety Council, www.nsc.org/resources/issues/estcost.aspx).

Collision Consequence Modeling

To estimate the total accident consequences with Equations 2 and 3, the probability that a passenger involved in a collision will suffer each given injury type (p_{ki}) must be estimated. To do this, a Bayesian severity model setting is used, which assumes that the outcome of a collision follows a multinomial distribution. In this model, information about vehicle occupancy can be incorporated in which each person involved in a collision suffers one of four possible severity types: fatality, severe injury, minor injury, or no injury. Alternatively, an ordinal Bayesian model could be formulated for the injury outcomes; however, a comparative analysis is beyond the scope of this paper. In addition, a random effect at the site level could be introduced in the model to account for intrasite correlation, since accidents coming from the same site can be nested. However, only the basic model formulation is shown for illustrative purposes.

To model the severity of an accident, it is assumed that an accident can lead to K possible injury outcomes, denoted by $\boldsymbol{r} = \{r_1, \ldots, r_k\}$, where k indicates the type of injury, $k = 1, \ldots, K$. Suppose that there are h persons involved in an accident and that each person could suffer from a specific injury outcome k with probability p_k. Then, by assuming that $\boldsymbol{r}$ follows a multinomial distribution, an accident outcome can be modeled as

$$\boldsymbol{r} \mid \boldsymbol{p} \sim \text{multinomial}(h, \boldsymbol{p}) \quad (4)$$

where h is the number of persons involved in an accident (note that $r_1, \ldots, r_K$ are nonnegative integers) and $\boldsymbol{p}$ is the vector of probabilities [$\boldsymbol{p} = (p_1, \ldots, p_K)$, p_k being the probability that a person involved in an accident has an injury of type k ($p_k > 0$)].

Then the probability that an involved person will have a type k injury can be estimated by using the following logit regression:

$$p_{ik} = \Pr(\text{a passenger} = k) = \frac{\exp(\varphi_{ik})}{\sum_{k=1}^{K} \exp(\varphi_{ik})} \quad (5)$$

where φ_{ik} is a measure representing the propensity for a person involved in an accident with specific traits i to experience severity type k. Here, φ_{ik} is expressed as a linear function of site characteristics, environment, and individual attributes:

$$\varphi_{ik} = \gamma_{0k} + \gamma_{1k} z_{1i} + \cdots + \gamma_{mk} z_{mi} \quad (6)$$

where $\boldsymbol{z}_i = (z_{1i}, \ldots, z_{mi})$ represents site, vehicle, or other characteristic (e.g., speed limits, location characteristics where accident took place, vehicle type) and $\boldsymbol{\gamma}_k = (\gamma_{0k}, \ldots, \gamma_{mk})$ is a vector of regression parameters. In the model, a Gaussian noninformative prior is assumed on these parameters. Once the proposed model is calibrated with empirical data, Equation 3 [i.e., $f_{ki} = E(r_k) = p_{ki} \times h_i$] can be used to estimate, for example, the expected number of fatalities, serious and minor injuries, and noninjuries.

After computation of the expected number of casualties by type, the total severity score C_i can be estimated according to Equation 2, where the f-values vary across locations, depending on attributes such as posted road speeds, maximum train speeds, and levels of occupancy. Furthermore, ω_1, ω_2, and ω_3 are usually provided by insurance or governmental agencies. In the approach used here, these weights can be assumed to be fixed or to follow a known prior distribution with parameters fixed according to different cost estimates such as those reported in Table 1. Other components may be included in the cost per accident, such as costs for property damage, delays, and emergency services. Even so, there is often not enough information to incorporate these extra costs. In addition, the extra costs are usually smaller than the cost of fatalities and injuries.

This modeling setting is based on the idea that there are different collision configurations across sites. Thus, there are variations in the injury type probabilities that can be explained by site-specific factors such as roadway features (posted road speed, urban or rural site, surface width, etc.) as well as environmental conditions and passenger and vehicle characteristics. Obviously, the set of factors that can be included depends on data availability. Moreover, many unobservable factors affect injury levels (two passengers with the same observable characteristics who have a collision do not necessarily have the same injury level).

Ranking Criteria Based on Absolute and Relative Total Risk

Once the severity score is determined, a hot spot strategy based on the posterior distribution of the total risk (TR_i) can be specified as follows:

$$\upsilon_i^{TR} = \Pr\left(TR_i > c_T \,\middle|\, \text{data}\right) \geq \delta_0 \tag{7}$$

where c_T is a standard value established by decision makers and δ_0 is a threshold value or confidence level varying between 0 and 1. For the definition of δ_0, the Bayesian testing methods introduced in previous work (*16*) are used.

The decision as to whether a site should be considered a hot spot could also be made on the basis of its relative rank as compared with other sites under a given safety measure. The rank of a site i under the total risk (TR_i) is defined as follows:

$$r_{(TR)_i} = \sum_{j=1}^{n} I\left(TR_i \geq TR_j\right) \tag{8}$$

where I(condition) is an indicator function with the value 1 if the condition is met and 0 otherwise. The index TR in r stands for the relative comparison under total risk. Given that the safety measure TR_i is a random variable, the resulting rank $r_{(TR)_i}$ is also a random variable with its posterior distribution depending on the relative comparison of TR_i with respect to the others. Hot spots can then be identified on the basis of the ranks of the sites by computing the following posterior probability:

$$\upsilon_i^r = \Pr\left(r_{(TR)_i} > q \,\middle|\, \text{data}\right) \qquad \text{if } \upsilon_i^r > \delta_1, \text{ site } i \text{ is a hot spot} \tag{9}$$

where q is a standard or upper limit rank specified by the decision makers. For instance, q can be defined as a certain proportion of n, that is, $q = \tau \times n$, where τ is a percentage (e.g., 70%, 80%). This hot spot selection strategy can be used when the focus is on the identification of sites with ranks greater than a certain percentile value. Once υ_i is computed under Equation 9, the optimal cutoff value δ_1 can be determined by using any of the multiple testing procedures introduced by Miranda-Moreno et al. (*16*). Note also that other ranking criteria can be formulated according to any of the safety measures previously defined.

CASE STUDY

To illustrate the proposed approach, a sample of highway–railway intersections in Canada is considered as an application environment. For this case study, a group of public crossings with automatic gates as the main warning device is considered; they comprise 1,773 crossings. Automatic gates provide an additional control level and are usually found in conjunction with flashing lights. The gate arms are usually reflectorized and fully cover the approaching roadway to prevent motor vehicles from circumventing the gates, which are coordinated with the flashing lights (Figure 1). All crossings use

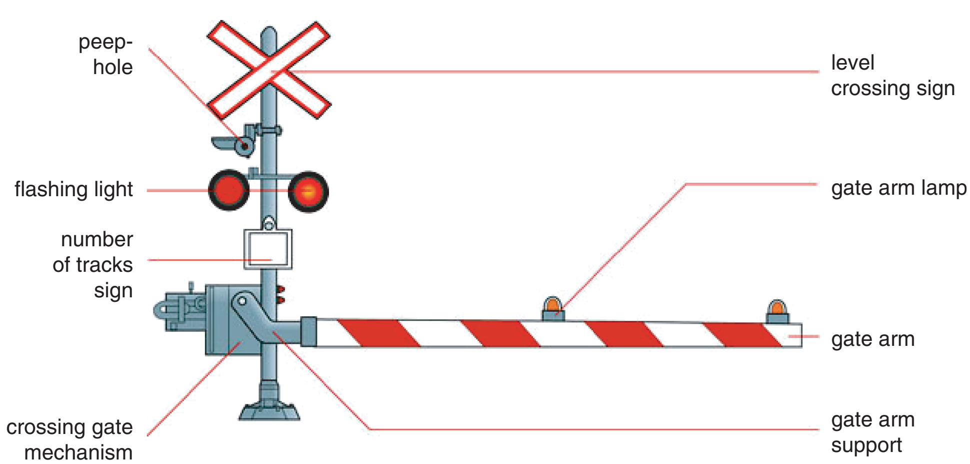

FIGURE 1 Standard level crossing with gates.

two-quadrant gates with dual gate arms, which block motor vehicles in each direction. Once a crossing of this type is identified as a hot spot, it can be further upgraded with new countermeasures such as four-quadrant gates; median separation, which can prevent vehicles from driving around lowered gates; or grade separation.

The main characteristics of this data set are summarized in Table 2. The table indicates that the data set is characterized by a high proportion of zero accidents (and low mean), which is a common characteristic in accident data. Before model calibration, an exploratory data analysis was carried out to identify high linear correlation among covariates and to detect observations with extreme values or missing information. The correlation among crossing attributes was moderate, and surface width is not included in the analysis because this information is missing for many crossings.

Frequency Model Calibration

For illustration purposes, two hierarchical Bayes models are discussed: the Poisson–gamma model with a multiplicative error term—$\exp(\epsilon_i) \sim \text{gamma}(\phi, \phi)$—and the Poisson–lognormal model with an additive error, $\epsilon_i \sim \text{normal}(0, \sigma^2)$. As described by Lord and Miranda-Moreno (*17*), the parameters a and b of the hyperprior distribution assumed on ϕ or σ^{-2} [$\phi \sim \text{gamma}(a, b)$ and $\sigma^{-2} \sim \text{gamma}(a, b)$] might be first specified. For the hierarchical Poisson–gamma model, noninformative priors with small values for these parameters could be assumed (e.g., $a = b = 0.01$ or $a = b = 0.001$). However, in this study, a more reasonable approach is followed: advantage is taken of the dispersion parameter estimate ($\hat{\phi}$) obtained by maximizing the negative binomial (NB) marginal likelihood (*16*). For example, for this data set, the NB model was calibrated first, which yielded a dispersion parameter of $\hat{\phi} = 0.64$. On the basis of the fact that the expectation of the gamma distribution assumed for ϕ is a/b and by fixing $a = 1$, it can be assumed that $E(\phi \mid b) = 1/b = 0.64$, from which $b = 1.56$. In this case study, vague hyperpriors are assumed for σ^{-2} with both parameters $a = b = 0.01$ or $a = b = 0.001$ for the hierarchical Poisson–lognormal model.

Once the hyperparameters are fixed, posterior distributions are sampled by using the statistical software WinBUGS. In this study, 6,000 simulation iterations were carried out for each parameter of interest, with the first 2,000 samples used as burn-in iterations. To select the crossing attributes to be included in the final model, the posterior expected values of all the regression coefficients were obtained first, along with their standard deviations and 95% confidence intervals. From those, only the attributes whose regression coefficients did not contain 0 in the 95% confidence intervals (i.e., regression parameters significantly different from 0 at the 95% confidence level) were selected. The crossing attributes are as follows:

1. Road type, represented as a binary variable (road type = 1 for arterials or collectors, 0 otherwise);
2. Posted road speed (km/h); and
3. Traffic exposure (E_i), computed as a function of daily road vehicle traffic (AADT_i) and number of daily trains (t_i) (i.e., $E_i = \text{AADT}_i \times t_i$).

The posterior summary of the covariate coefficients (β) along with the dispersion parameters ϕ and σ^2 were computed for the Poisson–gamma and Poisson–lognormal models, respectively. The results are presented in Table 3. The posterior mean of the regression coefficients is positive, except β_0, which makes sense from a safety point of view and confirms the results obtained in previous work (*7, 18*). In addition, the deviance information criterion (DIC) results presented in the same table indicate that a better fit to the data was obtained by applying the Poisson–gamma model. As expected, the DIC value computed with the Poisson–gamma model with semi-informative prior is smaller than the one obtained with the Poisson–lognormal model with vague hyperprior on the dispersion parameter. As stated earlier, a sensitivity analysis on alternative prior specifications is always recommended to identify the best alternative model. Given the large sample ($n = 1{,}773$) used in this exercise, the model outcome is not very sensitive to the prior assumptions. On the basis of the results of this analysis, it was decided to use Poisson–gamma in the subsequent analysis.

Severity Model Calibration

To calibrate the parameters of the proposed severity model, the collision database for the period 1997–2004 was utilized. A total of 941 highway–railway grade crossing collisions were included (see Table 4). Alternative specifications were attempted for the function φ_{ik}, from which it was found that maximum train speed and posted road speed are the main salient factors significantly influencing collision severity at a crossing—that is, $\varphi_{ik} = \gamma_{0k} + \gamma_{1k} \cdot z_{1i} + \gamma_{2k} \cdot z_{2i}$,

TABLE 2 Variables and Statistics for Crossing Data Set with Gates

Variable	Unit	Average/%	St. Dev.	Max	Min
Road class	Arterial/collector = 1, 0 others	44.0%			
Track number	Number	1.9	0.9	8.0	1.0
Track angle	Degree	72.5	18.4	120.0	0.0
Train maximum speed	mph	56.4	24.3	100.0	5.0
Road posted speed	km/h	59.3	16.1	100.0	15.0
Number of daily trains (F_1)	Trains/day	22.3	22.8	338.0	1.0
AADT (F_2)	Vehicles/day	4,162.7	6,041.7	48,000.0	10.0
Traffic exposure	$\text{Ln}(F_1 \times F_2)$	10.1	1.7	15.9	3.9
Whistle prohibition	If prohibited = 1, 0 otherwise	35.2%			
Surface width	ft	11.4	7.2	75.0	0.0
Number of collisions	Number (5-year period, 1997–2001)	0.12	0.4	4.0	0.0

TABLE 3 Posterior Estimates of Model Parameters

Hierarchical Model	Attributes		Posterior Mean	Std. Dev.	Markov Chain Error	Conf. Interval (2.50%–97.50%)
Poisson–gamma	Intercept	β_0	−6.429	0.717	0.076	(−7.764, −4.955)
$a = 1.56$	Road type	β_1	0.499	0.164	0.007	(0.171, 0.815)
	Posted road speed	β_2	0.011	0.005	0.000	(0.001, 0.021)
	Traffic exposure	β_3	0.323	0.054	0.005	(0.214, 0.429)
	ϕ		0.691	0.237	0.025	(0.381, 1.332)
					DIC = 1,191.25	
Poisson–lognormal	Intercept	β_0	−7.041	0.72	0.08	(−8.316, −5.657)
$a = 0.001$	Road type	β_1	0.506	0.17	0.01	(0.167, 0.842)
$b = 0.001$	Posted road speed	β_2	0.011	0.01	0.00	(0.001, 0.022)
	Traffic exposure	β_3	0.327	0.05	0.00	(0.228, 0.421)
	σ		1.016	0.14	0.01	(0.706, 1.296)
					DIC = 1,220.40	

where z_{1i} and z_{2i} are maximum train and road posted speeds, respectively. The calibration results are shown in Table 5, where the parameters for Severity Type 1 (fatality) are not shown, since Severity Type 1 is set as the base type with parameters equal to 0. The selection of this model was supported by the DIC. Figure 2 shows how p_{1i} varies as a function of maximum train and posted road speeds. The figure shows that the outcomes of a highway–railway crossing collision are sensitive to maximum train speeds.

Hot Spot Identification Using Total Risk

On the basis of the same group of highway–railway crossings with gates as a main warning device (n = 1,773 intersections) and the hierarchical Poisson–gamma model defined above, the υ-values were computed for the hot spot identification. To do so, a Bayesian approach was implemented by using a Markov chain Monte Carlo (MCMC) framework. One of the advantages of this framework is that different sources of information and uncertainty can be incorporated into the analysis. The multiple sources of information (parameters) in this decision process are illustrated in Figure 3. In the framework used here, not only are the model parameters (β, ϕ, θ_i) assumed random but also the uncertainty with C_i can be introduced by defining prior distributions in different model parameters, including ω, h_i, and γ.

In this demonstrative example, a value of $c_T = 1$ is defined according to the weights defined in Table 1, which is equivalent to the cost of a serious injury. In addition, an average level of occupancy (O_i) equal to 1.29 (which corresponds to the average vehicle occupancy of the recorded collisions involved in this analysis) is used. Once the various model parameters are fixed, MCMC algorithms can be used for the computation of υ-values.

Finally, a Bayesian testing approach is used for the definition of δ_0 (*16*). Fixing $c_T = 1$ and controlling the false-discovery rate at 10% (α_D = 10%) result in the optimal thresholds and hot spot list sizes presented in Table 6. Use of a Bayesian test with weights for $\kappa_0 = 3$ and $\kappa_1 = 1$ results in threshold values and hot spot list sizes given in the same table. The codes for computing the model parameters and υ-values are provided by Miranda-Moreno (*18*). For estimating the parameters of the Bayesian Poisson and multinomial logit models and the posterior υ-values, the software package WinBUGS was used. Written codes are provided by Miranda-Moreno (*18*).

As Table 6 and Figure 4 indicate, the hot spot list size is sensitive to the weight assigned to the fatalities (ω_1). For example, for a given value of c_T and a specific control level (α_D), the hot spot list size increases in a nonlinear way as ω_1 increases. The designation of a weight (or monetary value) for a human life may be controversial. However, in the hot spot identification activity, this helps target locations where not only the accident frequency but also the consequences will be high. In the case of highway–railway crossings, intersections with high maximum train and posted road speeds will be pushed up in the ranking, since the accident severity at these sites will be higher.

TABLE 4 Summary of Collisions by Severity

Variable	Total	Average (no./collisions)	Max	Min
No. of accidents	941	—	—	—
No. of occupants (persons involved)	1,217	1.29	21	1.0
Fatalities	137	0.15	3.0	0.0
Serious injuries	189	0.20	3.0	0.0
Minor injuries	241	0.26	7.0	0.0

Practical Application of the Proposed Methodology

For practitioners, the implementation of the modeling framework introduced above may not be straightforward. It demands advanced statistical and computational knowledge, which could significantly hinder its application in addressing practical problems. Therefore, a web-based decision support tool called GradeX has been developed (details are available at www.gradex.ca/). The tool makes some state-of-the-art risk-based methodologies for hot spot identification, such as the one introduced in this paper, accessible to practitioners. GradeX is used by Transport Canada and all of its regional offices to identify grade crossing hot spots and analyze alternative countermeasures for safety improvements. It integrates a rich set of accident prediction models and risk assessment methodologies, including the multinomial logit model presented in this paper. GradeX also offers state-of-the-art methodologies for countermeasure effectiveness analysis. It provides users with a convenient interface to define

TABLE 5 Calibration Results of Consequence Models

Severity Type	Variable	Coefficient	Posterior Mean	Std. Dev.	Markov Chain Error	Conf. Interval (2.50%–97.50%)
Fatal		(Base type)		$\gamma_{01} = \gamma_{11} = \gamma_{21} = 0$		
Major injury	Intercept	γ_{02}	1.462	0.305	0.025	(0.833, 2.026)
	Train speed	γ_{12}	−0.021	0.005	0.000	(−0.030, −0.010)
Minor injury	Intercept	γ_{03}	2.072	0.306	0.026	(1.412, 2.698)
	Train speed	γ_{13}	−0.028	0.005	0.000	(−0.038, −0.018)
No injury	Intercept	γ_{04}	4.459	0.319	0.028	(3.832, 5.134)
	Train speed	γ_{14}	−0.042	0.004	0.000	(−0.051, −0.033)
	Road speed	γ_{24}	−0.014	0.003	0.000	(−0.020, −0.008)

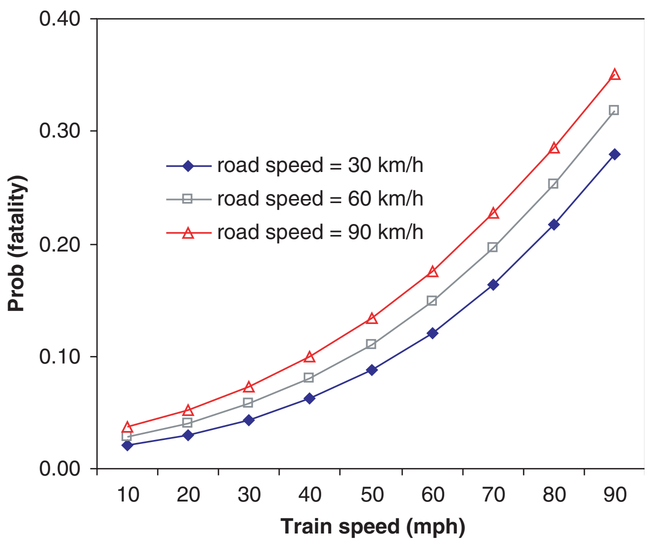

FIGURE 2 Probability of fatality for various road and train speeds.

a set of crossings to be investigated, which facilitates analysis of crossings located within any geographical area, such as region, municipality, and corridor [details are given by Fu et al. (*19*)].

CONCLUSIONS AND FUTURE WORK

One of the common approaches to hot spot identification is first to rank candidate sites on the basis of a safety measure and then to select the top sites according to a critical value. However, little research has been conducted in literature on how to incorporate heterogeneities across locations in the severity and occupancy levels at the hot spot identification stage. In this paper, a systematic full Bayesian framework for estimating the total risk of a given site as the product of accident frequency and its expected consequences was proposed. The Bayesian framework allows the introduction of severity uncertainty, not only in the model parameters but also in key factors such as vehicle occupancy levels and severity weighting factors. The proposed framework also allows identification of hot spots under relative or absolute measures of total risk, with an

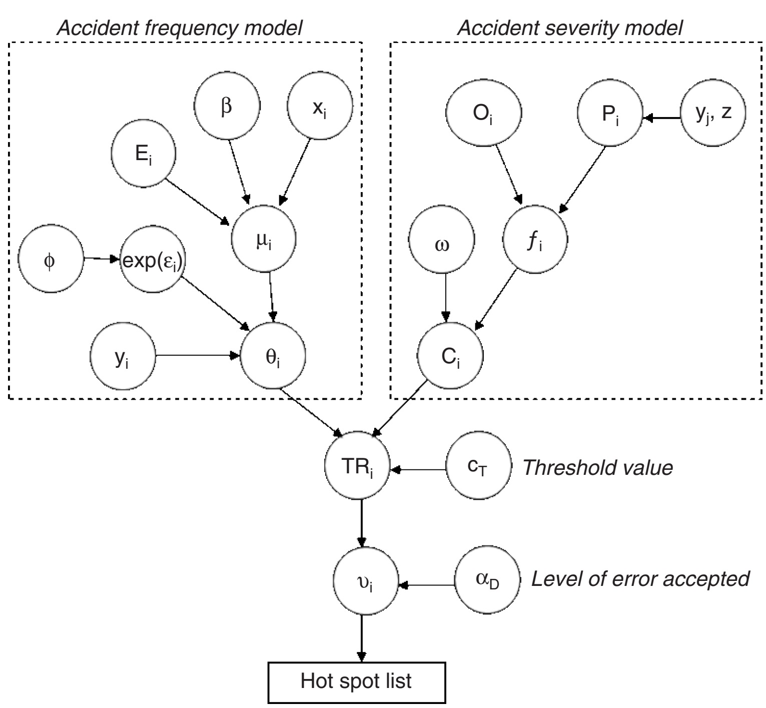

FIGURE 3 Modeling framework for computation of total risk.

TABLE 6 Threshold Values and Hot Spot List Size Under Total Risk

	FDR Test ($\alpha_D = 10\%$)		Bayesian Test with Weights ($\kappa_0 = 3$ and $\kappa_1 = 1$)	
Severity Weights	Threshold (δ_0)	No. of Hot Spots	Threshold (δ_0)	No. of Hot Spots
$\omega_1 = 15$, $\omega_2 = 1$, $\omega_3 = 0.5$	0.83	32	0.75	52
$\omega_1 = 30$, $\omega_2 = 1$, $\omega_3 = 0.5$	0.80	66	0.75	77
$\omega_1 = 41$, $\omega_3 = 1$, $\omega_3 = 0.5$	0.76	105	0.75	111

NOTE: FDR = false-discovery rate.

appropriate control on global error rates, such as a false positive error rate.

The applicability of the framework is illustrated by using an accident data set from Canadian highway–railway crossings with automatic gates. To estimate total accident consequences, the probability that a passenger involved in a collision is fatally or seriously injured is estimated by using a Bayesian multinomial model. In this model, information with regard to vehicle occupancy can be incorporated in which each person involved in a collision has several possible severity outcomes, such as fatality, severe or minor injury, and no injury. In addition, hierarchical Poisson models with additive and multiplicative model errors are used to model accident frequency. For this particular case, the Poisson–gamma model fits the observed data better than does the Poisson–lognormal. Finally, multiple Bayesian tests are implemented to control the proportion of false positives in the hot spot list.

Considering the number of persons involved in an accident in concert with the number of crashes is expected to improve the effectiveness of allocating resources to various safety programs. It is recognized that the inclusion of vehicle occupancy in road safety analysis may represent some challenges in practical applications—obtaining occupancy data for each location involved in the analysis can be an expensive and time-consuming task. However, in safety studies in which detailed occupancy information is not available, sites can be classified according to the proportion of high-occupancy vehicles, such as transit and school buses. Alternatively, as shown in this work, average vehicle occupancy levels can be approximately determined for various subgroups of locations on the basis of vehicle occupancies reported in accident data. The premise is that, with the proposed model, locations with a higher average vehicle occupancy would have a better chance of being included in the hot spot list.

As part of the research effort, a safety measure that estimates the "anticipated" cost–benefit ratio is being developed. Since a hot spot selection strategy aims to direct safety improvement efforts toward sites where maximum cost-effectiveness can be achieved, it would be of great value if the process could take into account both the costs and the safety benefits of remedy projects that could be introduced at the sites under consideration (*4*). Hierarchical ordered models are to be developed and integrated into the risk-based framework. The comparative performance of relative versus absolute measures of risk will be part of future research. It is also necessary to explore the use of new technologies to improve existing methods for occupancy data collection. The evolution of vehicle occupancy and its implications for road safety also deserve further investigation.

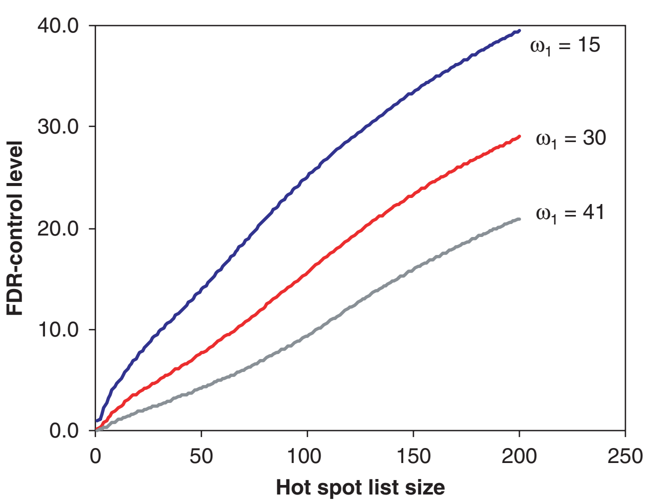

FIGURE 4 Sensitivity of list size to fatality weights.

REFERENCES

1. Schluter, P. J., J. J. Deely, and A. J. Nicholson. Ranking and Selecting Motor Vehicle Accident Sites by Using a Hierarchical Bayesian Model. *Statistician,* Vol. 46, No. 3, 1997, pp. 293–316.
2. Heydecker, B. G., and J. Wu. Identification of Sites for Accident Remedial Work by Bayesian Statistical Methods: An Example of Uncertain Inference. *Advances in Engineering Software,* Vol. 32, 2001, pp. 859–869.
3. Hauer, E., J. Kononov, B. Allery, and M. S. Griffith. Screening the Road Network for Sites with Promise. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 27–32.
4. Hauer, E., and B. N. Persaud. How to Estimate the Safety of Rail–Highway Grade Crossings and the Safety Effects of Warning Devices. In *Transportation Research Record 1114,* TRB, National Research Council, Washington, D.C., 1987, pp. 131–140.
5. Persaud, B., C. Lyon, and T. Nguyen. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1665,* TRB, National Research Council, Washington, D.C., 1999, pp. 7–12.
6. Miaou, S. P., and J. J. Song. Bayesian Ranking of Sites for Engineering Safety Improvement: Decision Parameter, Treatability Concept, Statistical Criterion and Spatial Dependence. *Accident Analysis and Prevention,* Vol. 37, 2005, pp. 699–720.
7. Miranda-Moreno, L. F., L. Fu, F. F. Saccomanno, and A. Labbe. Alternative Risk Models for Ranking Locations for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 1–8.

8. Cheng, W., and S. P. Washington. Experimental Evaluation of Hotspot Identification Methods. *Accident Analysis and Prevention,* Vol. 37, 2005, pp. 870–881.
9. Nassar, S., F. F. Saccomanno, and J. H. Shortreed. Disaggregate Analysis of Road Accident Severities. *International Journal of Impact Engineering,* Vol. 15, No. 6, 1994, pp. 815–826.
10. Saccomanno, F. F., L. Fu, and L. F. Miranda-Moreno. Risk-Based Model for Identifying Highway–Rail Grade Crossing Blackspots. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1862,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 127–135.
11. Milton, J. C., V. N. Shankar, and F. L. Mannering. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention,* Vol. 40, 2008, pp. 260–266.
12. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis and Prevention,* Vol. 40, No. 3, 2008, pp. 1033–1054.
13. Park, E. S., and D. Lord. Multivariate Poisson–Lognormal Models for Jointly Modeling Crash Frequency by Severity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2019,* Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–6.
14. Levine, N., and M. Wachs. Factors Affecting Vehicle Occupancy Measurement. *Transportation Research A,* Vol. 32, No. 3, 1998, pp. 215–229.
15. Zaloshnja, E., T. Miller, F. Council, and B. Persaud. Crash Cost in the United States by Crash Geometry. *Accident Analysis and Prevention,* Vol. 38, 2006, pp. 644–651.
16. Miranda-Moreno, L. F., A. Labbe, and L. Fu. Multiple Bayesian Testing Procedures for Selecting Hazardous Sites. *Accident Analysis and Prevention,* Vol. 39, No. 6, 2007, pp. 1192–1201.
17. Lord, D., and L. F. Miranda-Moreno. Effects of Low Sample Mean Values and Small Sample Sizes on the Dispersion Parameter Estimation of Poisson–Gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective. *Safety Science,* Vol. 46, No. 5, 2008, pp. 751–770.
18. Miranda-Moreno, L. F. *Statistical Models and Methods for Identifying Hazardous Locations for Safety Improvements.* PhD thesis. University of Waterloo, Waterloo, Ontario, Canada, 2006.
19. Fu, L., F. Saccomanno, L. F. Miranda-Moreno, and P. Y.-J. Park. GradeX: Decision Support Tool for Hot Spot Identification and Countermeasure Analysis of Highway–Railway Grade Crossings. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.

The Safety Data, Analysis, and Evaluation Committee sponsored publication of this paper.