Empirical Evaluation of Alternative Approaches in Identifying Crash Hot Spots

Naive Ranking, Empirical Bayes, and Full Bayes Methods

Helai Huang, Hoong Chor Chin, and Md. Mazharul Haque

This study proposes a framework of a model-based hot spot identification method by applying full Bayes (FB) technique. In comparison with the state-of-the-art approach [i.e., empirical Bayes method (EB)], the advantage of the FB method is the capability to seamlessly integrate prior information and all available data into posterior distributions on which various ranking criteria could be based. With intersection crash data collected in Singapore, an empirical analysis was conducted to evaluate the following six approaches for hot spot identification: (*a*) naive ranking using raw crash data, (*b*) standard EB ranking, (*c*) FB ranking using a Poisson-gamma model, (*d*) FB ranking using a Poisson-lognormal model, (*e*) FB ranking using a hierarchical Poisson model, and (*f*) FB ranking using a hierarchical Poisson (AR-1) model. The results show that (*a*) when using the expected crash rate–related decision parameters, all model-based approaches perform significantly better in safety ranking than does the naive ranking method, and (*b*) the FB approach using hierarchical models significantly outperforms the standard EB approach in correctly identifying hazardous sites.

Hot spot identification is an essential task in engineering programs to improve the safety situation of road networks. Although methods vary greatly, most studies have relied on historical traffic crash records to obtain an estimation of safety for diverse traffic entities. The majority of traditional methods estimate safety by using raw crash data. These methods include the crash frequency method (*1*), the crash rate method, the rate quality control method (*2*), the crash severity method, and the safety index method (*3*).

Despite their simplicity, however, these naive statistical methods using raw crash data have serious limitations, particularly in cases when data are only available for a short-term period (e.g., 2 to 3 years or less). A well-known limitation is the regression-to-the-mean problem (*4*). Furthermore, the naive method is subject to its incapability of examining the crash dispersion (*5*). In particular, it may be more useful to rank the sites by the magnitude by which the crash risk exceeds what is normal for such sites with similar characteristics. However, naive ranking methods are unable to obtain the "normal" expected crash rate by accounting for various heterogeneities associated with diverse site features.

To overcome the problems of naive ranking methods, several recent studies have proposed model-based ranking approaches for hot spot identification. Elvik (*5*) has presented an extensive review for the state-of-the-art approaches of road crash hot spot identification and concludes that the empirical Bayes approach (EB) (*4, 6–8*) is the most reliable method. Moreover, using simulation experiments (*9*) and innovative robust evaluation criteria (*10*), Cheng and Washington have also proved that the EB approach is the most consistent and reliable method for identifying sites with promise.

While the EB approach is demonstrably better suited to reliably estimate the safety of traffic sites than naive statistical methods, it may be subject to several imperfections under certain conditions. For example, in a standard before-and-after evaluation using the EB approach, an external (or previously obtained) safety performance function (SPF) is used for each candidate site to provide the prior estimate of safety. But in general hot spot identification studies, it may sometimes require all sites in a road network to be screened and the SPF to be fitted purely from the local data. In such a condition, the EB approach may be criticized for implicitly using the data twice. That is, the data are first used to estimate the parameters in SPF and, once these values are determined, the observed crash count of each site is utilized to make inference about the posterior estimation (e.g., *4, 11*). Another potential imperfection is that the EB approach may be inadequate to explicitly account for the "uncertainty" of associations of covariates and safety (*12*). In the EB approach, the uncertainty, which is used to indicate the confidence level of the estimated normal safety, is merely represented by a vague overdispersion term and it is identical for all sites in model prediction. Thus, once the SPF is calibrated, the point estimates of all covariate effects will be assumed to be true without any uncertainty. However, the covariate effects in the SPF are typically estimated from crash data and thus are definitely subject to uncertainty. Hence, the posterior estimate of safety may be more accurate if all uncertainties associated with each covariate effect on safety could be fully taken into account.

A promising alternative of the EB approach is the full Bayesian (FB) method. The essential characteristic of a FB analysis is its explicit use of probability for quantifying uncertainty in inferences based on statistical data analysis. Thus, by accommodating all uncertainties in the model, the normal expected crash rate could be represented by a fitted distribution and the final safety estimate for a specific site is obtained by averaging the possible values relative to the distribution (*12, 13*). It is clear that the EB approach is a special case of FB that

H. Huang, Department of Civil and Environmental Engineering, University of Central Florida, 301A Engineering Building 2, Orlando, FL 32816. H. C. Chin and M. M. Haque, Department of Civil Engineering, National University of Singapore, Singapore 117576. Corresponding author: H. Huang, huanghelai@hotmail.com.

Transportation Research Record: Journal of the Transportation Research Board, No. 2103, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 32–41.
DOI: 10.3141/2103-05

arises when an FB analysis is simplified, for instance, by assuming that the safety effect of covariates such as lane width is known without any uncertainty. A further advantage of the FB method is its flexibility in explicitly representing the hierarchical models. Hierarchical structures exist extensively in crash data because of the data collection and clustering process. Several recent studies have indicated that crash rate may be better fitted by explicitly accounting for the spatial and temporal heterogeneities using hierarchical modeling technique (*14–18*). Hence, there is a need to explore the use of FB hierarchical models in site ranking and hot spot identification.

This study aims at employing the FB approach to propose a methodological framework of hot spot identification. Different probabilistic model specifications are proposed with Bayesian model inference and model selection measurements. The impact of different risk models are compared with each other with respect to the safety ranking and hot spot screening. Comparison is also conducted with the naive ranking method and the standard EB approach to provide an empirical evaluation on alternative hot spot identification approaches.

FULL BAYESIAN HIERARCHICAL MODELING APPROACH IN HOT SPOT IDENTIFICATION

This section presents the proposed FB approach including model specification, comparison, and assessment, as well as safety estimate and decision parameters in ranking sites and identifying hot spots.

Model Specification

Central to the model-based hot spot identification methods is the establishment of a probabilistic functional form of crash frequency and diverse risk factors associated with the traffic sites (i.e., SPF or crash prediction model). Traditionally, a Poisson regression model has been applied for fitting crash rate. While the Poisson model requires the equality of mean and variance, many studies have found that the variance of crash frequency is significantly greater than the mean (e.g., *19, 20*). To accommodate for the overdispersion, some additional variance components could be incorporated into the mean of the standard Poisson model. The overdispersed Poisson model is generally expressed as follows:

$$y_{it} \mid \lambda_{it} \sim \text{Poisson}(\lambda_{it}) = \text{Poisson}(\mu_{it} e_{it})$$
$$\mu_{it} = \exp(\mathbf{X}_{it}'\boldsymbol{\beta} + \upsilon)$$

where y_{it} is the crash count at ith site in tth time period and Poisson distributed with the mean crash frequency λ_{it}. Given the exposure component e_{it} (normally a measure of traffic flow) and the observed risk factors $\mathbf{X}_{it}$ at a specific site, the mean crash rate μ_{it} varies by a stochastic component υ.

Conveniently, υ could be assumed independent of $\mathbf{X}_{it}$ and also independent of each other for different observations, denoted as ϵ_{it}. The standard Poisson–gamma (PG) (or negative binomial, NB) or Poisson–lognormal (PLN) models are obtained by specifying $\exp(\epsilon_{it})$ with a gamma or lognormal distribution, respectively.

Model 1. Poisson–gamma model (negative binomial model):

$$\mu_{it} = \exp(\mathbf{X}_{it}'\boldsymbol{\beta} + \epsilon_{it})$$
$$\exp(\epsilon_{it}) \sim \text{gamma}(a, a)$$

Model 2. Poisson–lognormal model:

$$\mu_{it} = \exp(\mathbf{X}_{it}'\boldsymbol{\beta} + \epsilon_{it})$$
$$\epsilon_{it} \sim \text{normal}(0, \sigma_\epsilon^2)$$

It is easy to derive the first two moments of the observed crash count y_{it} with these two model specifications as

Poisson–gamma model (negative binomial model):

$$E(y_{it}) = e_{it} \times \exp(\mathbf{X}_{it}'\boldsymbol{\beta}) \qquad \text{Var}(y_{it}) = E(y_{it})\left(1 + \frac{E(y_{it})}{a}\right)$$

Poisson–lognormal model:

$$E(y_{it}) = e_{it} \times \exp\left[\mathbf{X}_{it}'\boldsymbol{\beta} + \left(\frac{1}{2}\right)\sigma_\epsilon^2\right]$$
$$\text{Var}(y_{it}) = E(y_{it})\left[1 + E(y_{it})\left(\exp(\sigma_\epsilon^2) - 1\right)\right]$$

Clearly, both NB and PLN models have a relaxed variance structure which allows the variance to be larger than the marginal mean of y_{it}. In particular, the inverse dispersion parameters [i.e., $1/\alpha$ in the NB model and $(\exp[\sigma^2] - 1)$ in the PLN model] provide measures to accommodate the extra variability exhibited in crash data. While the NB model has been applied broadly in traffic safety studies (e.g., *19, 20*), Lord and Miranda-Moreno (*21*) recently found that the PLN model could be a better alternative in cases of small sample size and low sample mean. In the overdispersed Poisson models, a stochastic component is specified to the mean of the standard Poisson distribution to provide a vague measure to relax the equality constraint of mean and variance. However, they are incapable of taking into account any structured heterogeneities due to the spatial and temporal effects of the crash data (*14*). To explicitly model these structured heterogeneities introduced in the data collection and clustering process, hierarchical modeling technique has been found to be a better alternative in several recent traffic safety studies (*14, 17, 18*).

For this purpose, one common hierarchical specification could be proposed by replacing the cross-observation variation component ϵ_{it} with a cross-site random effects component α_i in the link function, which is,

Model 3. Hierarchical Poisson model

$$\mu_{it} = \exp(\mathbf{X}_{it}'\boldsymbol{\beta} + \alpha_i)$$
$$\alpha_i \sim \text{normal}(0, \sigma_\alpha^2)$$

The random effects α_i are independent of covariates $\mathbf{X}_{it}$ and also independent of each other for different sites. Clearly, this hierarchical Poisson (HP) model assumes that potential overdispersion identified in crash data may be explained by omitted or unobserved cross-site heterogeneities. Within each site, the observations in different time periods remain independent Poisson distributions.

Furthermore, it may also be interesting to investigate possible serial correlation between crash observations in successive time periods at a specific site. The serial correlation may arise from time-dependent factors that are unobservable or missing in the model and factors with time-dependent or autoregressive safety effect that are not explicitly specified in the model (*22*). This possible autocorrelation can be reflected in model by a serial variation δ_{it}. There are

various options to specify the distribution of δ_{it} (*23*). A common choice is to assume a lag-1 dependence in the errors (i.e., AR-1 model) based on the stationarity assumption, with γ as the autocorrelation coefficient:

Model 4. Hierarchical Poisson (AR-1) model

$$\mu_{it} = \exp\left(\mathbf{X}'_{it}\boldsymbol{\beta} + \delta_{it}\right)$$

$$\delta_{i,1} \sim \text{normal}\left(0, \frac{\sigma_\delta^2}{\left(1-\gamma^2\right)}\right)$$

$$\delta_{i,t} \sim \text{normal}\left(\gamma\delta_{i,t-1}, \sigma_\delta^2\right) \qquad \text{for } t > 1$$

As shown above, different models could be specified to explicitly accommodate various structured heterogeneities according to the specific crash data structure. In this study, the abovementioned four model specifications are employed to illustrate the FB approach in comparison with traditional hot spot identification methods.

Model Comparison and Assessment

In selecting an appropriate model, different stochastic components can be tested by the statistical significance of their hyperparameters. In principle, the most parsimonious model is preferred. For the purpose of model selection, deviance information criterion (DIC) is used in this study (*24*). Moreover, the goodness of fit of candidate models are assessed by mean absolute deviance (MAD) and mean-squared predictive error (MSPE):

$$\text{MAD} = \frac{1}{n}\sum_{\forall i,t}\left|y_{it}^{\text{pred}} - y_{it}^{\text{obs}}\right|$$

$$\text{MSPE} = \frac{1}{n}\sum_{\forall i,t}\left(y_{it}^{\text{pred}} - y_{it}^{\text{obs}}\right)^2$$

where y_{it}^{obs} is the observed crash number at site i in time period t while y_{it}^{pred} denotes the model prediction of expected crash number for the specific site.

Full Bayesian Estimation of "Safety" and Decision Parameters

Bayesian analysis is a process of fitting a probability model to the data set and summarizing the posterior probability distribution on the model parameters and on unobserved quantities (*13*). In the context of estimating the expected number of crashes at a specific traffic site, the FB method consists of three steps:

Step 1. Specifying prior distributions for each model parameter [i.e., Prior($\boldsymbol{\beta}$, υ)]. These prior distributions represent accumulated knowledge about, for example, effects of risk factors and dispersion parameters in crash data.

Step 2. Setting up a full probability model (M): Pr(y_{it}, $\boldsymbol{\beta}$, υ) = Pr(y_{it}|$\boldsymbol{\beta}$, υ)Pr($\boldsymbol{\beta}$, υ), which is a joint probability for all parameters as well as the unobservable expected number of crashes at certain sites.

Step 3. Conditioning on all observed data (D): Calculating the appropriate posterior distribution of all unknowns and then integrating this distribution over the model parameters to obtain the desired marginal distribution of the expected number of crashes at a specific site.

$$\begin{aligned}\Pr\left(y_{it}^{\text{pred}} \middle| \mathbf{X}_{it}, e_{it}, M, D, \text{prior}\left(\boldsymbol{\beta}, \upsilon\right)\right) &= \int \Pr\left(y_{it}^{\text{pred}} \middle| \mathbf{X}_{it}, e_{it}, M, D, \text{prior}\left(\boldsymbol{\beta}, \upsilon\right), \boldsymbol{\beta}, \upsilon\right) \\ &\quad \times \Pr\left(\boldsymbol{\beta}, \upsilon \middle| M, D, \text{prior}\left(\boldsymbol{\beta}, \upsilon\right)\right) d\left(\boldsymbol{\beta}, \upsilon\right)\end{aligned}$$

Markov chain Monte Carlo (MCMC) (*13*) is a general approach to implement Bayesian inference. For this study, simulations of model prediction for the crash number of each site (i.e., y_{it}^{pred}) were drawn from the posterior distribution. Use the sample draws to compute the posterior density of any functions of y_{it}^{pred} that may be of interest.

Hence, it is clear that as in the EB approach, information about similar sites is incorporated into the estimate for a single site in the FB analysis. Nevertheless, in the EB approach the model parameters are estimated externally. The point estimators of parameters are then used as if they were true values. All uncertainties associated with model prediction are integrated by a vague and identical overdispersion parameter as an indication of confidence level of observed crash data. This implies that, in the EB approach, uncertainty associated with estimating covariate coefficients may not be precisely taken into account in the estimate of safety for a specific site (*12, 13*). Comparatively, the FB approach provides a flexible and direct means to fully accommodate and measure uncertainties associated with model predictions of the expected crash number.

Following a FB estimate of safety, different decision parameters (Θ) could be selected to rank the sites considered for further engineering study of safety treatment. The choice of decision parameters needs to take into account the context under which the rank is to be used, especially the range of treatments to be performed. Without much loss of generality, two common decision parameters are illustrated:

$$\Theta_{i,1} = \sum_{\forall t}\hat{\lambda}_{it}$$

$$\Theta_{i,2} = \sum_{\forall t}\left(\hat{\lambda}_{it} - \hat{\lambda}_{it}^{\text{normal}}\right) \qquad \hat{\lambda}_{it}^{\text{normal}} = e_{it} \times \exp\left(\mathbf{X}'_{it}\hat{\boldsymbol{\beta}}\right)$$

While the first decision parameter ($\Theta_{i,1}$) provides a direct measure for expected crash number for the years of interest, the second ($\Theta_{i,2}$) is useful to identify sites that exhibit abnormally high unobserved random effects (i.e., higher accident reduction potential) so that some field studies may be conducted to figure out what might have contributed to these effects.

EMPIRICAL EVALUATION OF ALTERNATIVE HOT SPOT IDENTIFICATION APPROACHES

This section presents an empirical evaluation of proposed alternative FB models in terms of model fitting, site ranking, and hot spot identification by comparing the FB method with the naive statistical method and the state-of-the-art method using a standard EB approach.

Data Description and Analytical Strategy

In this study, a data set of all four-legged signalized intersections at Singapore was collected from a citywide survey. The data set contains historical crash data maintained by Singapore Traffic Police and major descriptive factors associated with each intersection. In order to circumvent special modeling treatment needed to account for zero

inflation in the crash data (*25, 26*), sites with four or more crashes during the past 10 years (1997–2006) were screened and used for analysis. A total of 582 sites were used. A cursory review of the selected sites showed that the samples cover almost all important intersections and are well spread out over the country.

In spite of considerable research efforts in exploring the statistical relationships between traffic crashes and traffic flows (e.g., *12*), there is no consistent crash exposure measurement for intersections. In order to select an appropriate exposure measure for this study, three traditional functional forms [see Chapman (*27*)] were preliminarily considered, which include the sum of all entering flows, the product of cross-flows at conflict points, and the square root of the product of the cross-flows. Correlation tests on each of these exposure measures showed that the product of traffic volumes (average daily traffic) from major and minor roads of each intersection is the most appropriate one with the strongest correlation with the observed crash numbers. Hence, it was used as the traffic exposure measure (i.e., e_{it}) in the current study. In other words, the analysis assumes that crash frequency for a specific intersection is proportional to the product of the flows at two intersecting roadways given all other risk factors equal.

In addition to the traffic volume, a variety of self-explanatory risk factors are considered as covariates ($\mathbf{X}'_{it}$), which include one-way road, number of lanes, uncontrolled left-turn lane, exclusive right-turn lane, presence of red light camera, presence of median, and presence of pedestrians crossing. The dummy variable technique was employed to code these categorical variables.

In the evaluation, crash data in the latest 3 years (2004–2006) were used to calibrate the alternative risk models and then make the FB estimation for the expected crash rate at each site. The hazardous sites could then be selected based on the ranking using the estimated "safety" under different decision parameters. To assess whether the hot spots were truly or falsely identified, average crash rates referring to the whole 10-year period (1997–2006) were used as empirical criteria for comparison. The assumption was that, without significant change for the intersection layout during the study period, the crash count from a long-term historical period should well represent the "true" value of safety. A descriptive statistics chart of crash frequencies for the latest 3 years as well as for the whole 10 years is shown in Figure 1.

Model Calibration and Diagnostics

The FB analysis was implemented in freeware WinBUGS package (*28*) using a MCMC algorithm. In the absence of strong prior information for factor effects and dispersion parameters, uninformative priors were assumed with normal distribution (0, 1,000) for all regression coefficients ($\boldsymbol{\beta}$), and with gamma distribution (0.001, 0.001) for hyperparameters associated with the disturbance terms [i.e., $1/\alpha$ in NB model, σ^2_ϵ in PLN model, σ^2_α in HP model, and σ^2_δ in HP(AR-1) model]. In model calibration, three chains of 15,000 iterations were set up for each model based on the convergence speed and the magnitude of the data set. Gelman–Rubin convergence statistics (*29*) were used to monitor the convergence of simulation chains. After ensuring the convergence, the first 5,000 samples were discarded as adaptation and burn-in, and only every 10th sample of the rest were retained for estimation to reduce autocorrelation, leaving a total of 3,000 posterior samples for each estimate.

To obtain the most parsimonious model, preliminary multicolinearity tests and backward stepwise methods were employed in selecting covariates. Results of parameter estimation and Bayesian credible interval (95% BIC) of significant factors in the final models are summarized in Table 1. In all four models, six identical factors are identified as significant to affect the crash risk. The results of covariate coefficients appear to be very robust regardless of which model is used. The robustness of the results implies that there would not be significant difference between selected models if they are used only for identifying crash-contributing factors. In particular, it was found that intersections with two-way traffic roads, more conflicting lanes,

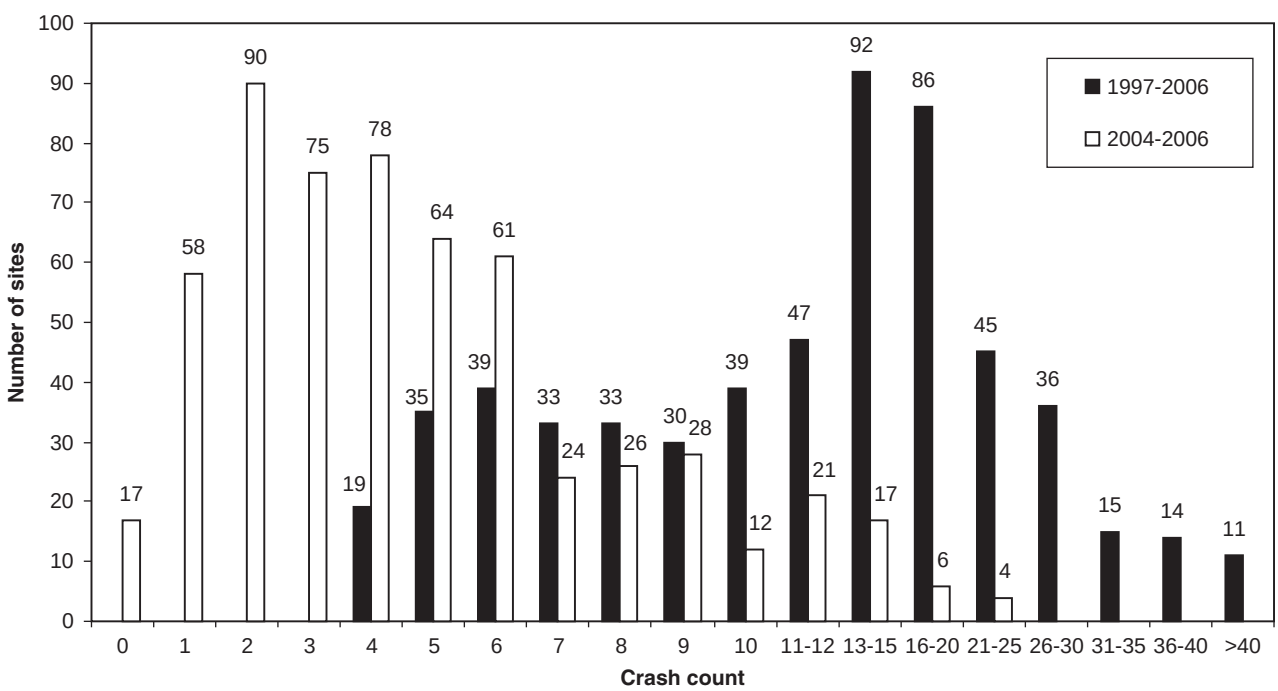

FIGURE 1 Distribution of crash counts in observations.

TABLE 1 Parameter Estimation and Model Diagnostics

	NB Model		PLN Model		HP Model		HP (AR-1) Model	
	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI	Mean	95% BCI
Intercept	–0.892	(–1.06, –0.71)	–0.971	(–1.14, –0.80)	–0.908	(–1.13, –0.67)	–0.930	(–1.17, –0.70)
One-way traffic	–0.226	(–0.31, –0.15)	–0.230	(–0.31, –0.15)	–0.238	(–0.34, –0.13)	–0.241	(–0.35, –0.13)
No. of lanes	–0.184	(–0.21, –0.16)	–0.185	(–0.21, –0.16)	–0.194	(–0.23, –0.16)	–0.191	(–0.23, –0.16)
Uncontrolled left-turn lane	–0.216	(–0.27, –0.17)	–0.217	(–0.27, –0.17)	–0.225	(–0.29, –0.16)	–0.228	(–0.30, –0.16)
Pedestrian crossing	0.222	(0.18, 0.26)	0.223	(0.18, 0.26)	0.225	(0.17, 0.28)	0.227	(0.17, 0.28)
Red-light camera	–0.286	(–0.36, –0.22)	–0.286	(–0.36, –0.21)	–0.290	(–0.40, –0.19)	–0.295	(–0.40, –0.181)
Speed limit < 50 km/h	–0.582	(–0.99, –0.19)	–0.583	(–1.02, –0.17)	–0.566	(–1.10, –0.07)	–0.580	(–1.09, –0.07)
Variance of υ[a]	0.171	(0.22, 0.13)	0.182	(0.14, 0.22)	0.219	(0.178, 0.27)	0.011	(0.00, 0.04)
Serial correlation coefficient (γ)							0.775	(0.72, 0.80)
DIC	7,661		7,655		7,344		7,339	
MAD	0.84		0.82		0.63		0.51	
MSPE	1.42		1.34		1.12		1.03	

[a]Variance of υ: $1/a$ in NB model, σ_ϵ^2 in PLN model, σ_α^2 in HP model, and σ_δ^2 in HP(AR-1) model.

higher speed limit, and the presence of a pedestrian crossing are associated with a higher crash rate. However, the use of an uncontrolled left-turn lane (left-side driving) and the installation of a red light camera may help reduce the crash rate at signalized intersections.

The comparison results of model diagnostic criteria estimated from the models are also shown in Table 1. The results indicate that the consideration of site-specific effects, resulting in HP and HP (AR-1) models, can significantly improve the goodness of fit in comparison with standard NB and PLN models. Specifically, judged by MAD values in the NB model and the HP model, the model-fitting performance is improved by about 25%. Although it is very difficult to theoretically distinguish between overdispersion (NB and PLN models) and within-site correlation (HP and HP (AR-1) models), the results imply that modeling site-specific effects significantly improves the model fitting.

The results also show that the PLN model provides a slightly better model fitting for the subject data set than the NB model (DIC from 7,661 to 7,655; MAD from 0.84 to 0.82; and MSPE from 1.42 to 1.34). It may be reasoned that the heavier tail in the lognormal distribution may better accommodate for the overdispersion exhibited in the data (*21*). However, it should be noted that, generally, if the difference in DIC between models is less than about 10, one would not readily conclude that one model is better than the other. Thus, even though MAD and MSPE support the difference between the two models, the model-fitting improvement is not conclusive.

Likewise, the slightly decreased DIC value in the HP (AR-1) model [DIC from 7,344 in the HP model to 7,339 in the HP(AR-1) model] should not be used to conclude a substantial model-fitting improvement. Nevertheless, the significant serial correlation coefficient (γ = 0.775) estimated in the HP (AR-1) model implies that autocorrelation on crash rate may really exist in the subject data set. Furthermore, in comparison with the HP model, the MAD (0.51) and MSPE (1.03) in the HP(AR-1) model also lend support to the improvement of goodness of fit. Wang and Abdel-Aty (*22*) have proved that the autoregression model is effective in accounting for time-dependent uncertainties in Florida intersection crash data. The limited improvement from the HP model to the HP (AR-1) model in this study may probably become more significant if more serial observations per site are used in the model.

Results of Empirical Evaluation in Hot Spot Identification

To evaluate alternative Bayesian models in hot spot identification, the expected crash numbers for a specific site ($\hat{\lambda}_{it}$), and the normal expected crash numbers ($\hat{\lambda}_{it}^{\text{normal}}$) for similar sites with equal values of observed covariates were monitored at the same simulation process of parameter estimation. Apart from the four proposed FB models, safety estimators using the means of crash counts observed during the 3 years (2004–2006), $\hat{\lambda}_i^3$, and the 10 years (1997–2006), $\hat{\lambda}_i^{10}$, were calculated. While $\hat{\lambda}_i^3$ is used as raw crash count for the naive ranking, $\hat{\lambda}_i^{10}$, as the long-term average of crash rate, is assumed to empirically represent the "true" values of safety. Moreover, a standard EB estimation for the expected crash frequencies was also implemented for comparison purpose. Specifically, the EB estimate is calculated as follows (*4*):

$$\hat{\lambda}_{it}^{\text{EB}} = \omega_{it}\hat{\lambda}_{it}^{\text{normal}} + \left(1-\omega_{it}\right)y_{it}^{\text{obs}} \qquad \text{with } \omega_{it} = \frac{1}{1+\hat{\lambda}_{it}^{\text{normal}}/\hat{a}}$$

where $\hat{\lambda}_{it}^{\text{normal}}$ is estimated in the NB model and $1/a$ has the value of 0.171 as shown in Table 1. Using the two decision parameters defined previously ($\Theta_{i,1}$ and $\Theta_{i,2}$), comparisons are conducted between the six safety ranking approaches: (*a*) naive ranking using $\hat{\lambda}_i^3$; (*b*) EB ranking using $\hat{\lambda}_{it}^{\text{EB}}$; (*c*) FB ranking using NB model $\hat{\lambda}_{it}^{\text{NB}}$; (*d*) FB ranking using PLN model $\hat{\lambda}_{it}^{\text{PLN}}$; (*e*) FB ranking using HP model $\hat{\lambda}_{it}^{\text{HP}}$; and (*f*) FB ranking using HP(AR-1) model $\hat{\lambda}_{it}^{\text{HP(AR-1)}}$.

For the purpose of comparison, false identification test (FIT) is commonly used in practice (*5, 9*). Generally in FIT, some critical crash rates (e.g., four or more crashes per year) or cutoff percentage of total number of sites (e.g., top 1%, 2.5%, 5%, or 10% of all sites) are predetermined to establish hazardous sites. Compared with the list of true safe sites, the binary result (i.e., safe or unsafe) given by different methods can be analyzed as correct positive (CP: correctly claim a safe site as safe), false positive (FP: falsely claim a safe site as unsafe), correct negative (CN: correctly claim an unsafe site as unsafe), and false negative (FN: falsely claim an unsafe site as safe). Two diagnostic criteria have been established in previous studies (*5*): sensitivity and specificity.

$$\text{sensitivity} = \frac{\text{number of correct positives identified } (\text{CP})}{\text{total number of true positives } (\text{TP})}$$

$$\text{specificity} = \frac{\text{number of correct negatives identified } (\text{CN})}{\text{total number of true negatives } (\text{TN})}$$

Clearly, "sensitivity" measures the ability of a specific method to correctly identify true hazardous sites, whereas "specificity" indicates the ability to distinguish safe sites from hazardous sites. Thus, the methods with higher values of sensitivity and specification are preferred.

However, the FIT has certain limitations. As argued by Cheng and Washington (*10*), the standard FIT is not informative enough to provide adequate insight into the relative performances of hot spot identification methods. One disadvantage of FIT is that, by using several predetermined thresholds, it is not able to fully utilize the ordinal rank data produced in hot spot identification. For example, suppose that a cutoff level of 5% of all sites is given in the current study, resulting in the top 29 sites as hot spots (582∗5%). In this case, for the worst site (top 1) in reality, the standard FIT is not able to provide any differentiation in terms of predictive performance for this specific site to be perfectly ranked first by one method or 29th by another. Another potential drawback of FIT is that each identification is weighted equally. In other words, once a site is identified as positive or negative, there is no quantitative comparison between the true crash rates and model predictions.

To overcome the major drawbacks of FIT, a continuous set of thresholds were tested in this study to monitor the performance of alternative approaches with respect to ordinal rank data. The percentage of dispersion deviations produced by different methods is also calculated in a continuous numerical scale. It is clear that with a continuous scale of critical crash rate or cutoff level, all information contained in the ordinal ranking data can be reflected in the analysis. This treatment is supposed to have a comparable effect to the site consistency test suggested by Cheng and Washington (*10*). A unique advantage of this treatment is its convenience in making graphical analysis, as shown later. Furthermore, for the purpose of quantitative comparison, a measure of MAD is utilized to indicate the quantitative discrepancy of each model prediction to the true safety. The empirical comparison results are detailed as follows.

False Identification Test

Two common types of FITs were conducted to compare the alternative methods using decision parameter $\Theta_{i,1}$, that is, by giving critical crash rates or cutoff number of sites (i.e., cutoff percentage of all sites). Specifically, FIT by giving critical crash rates means that the sites with predicted crash rates equal to or higher than the given critical crash rates are identified as hot spots. Likewise, in FIT by giving cutoff number of sites, the top sites with ranks up to the cutoff number are screened as hazardous.

Using a continuous scale of critical crash rates, Figures 2 and 3 show the performance of the six alternative approaches in terms of sensitivity and specificity, respectively. Clearly, the naive ranking method using only raw data is the one with the worst performance in correctly identifying both hazardous sites (see Figure 2) and nonhazardous sites (see Figure 3). Among the five model-based methods, it is apparent that the HP and HP (AR-1) models outperform EB, PG, and PLN models. This indicates that the models that account for site-specific effects are superior to the nonhierarchical models. The discrepancy becomes larger when the critical crash rate is increasing. In usual practice, because of the constraint of resource or budget, only

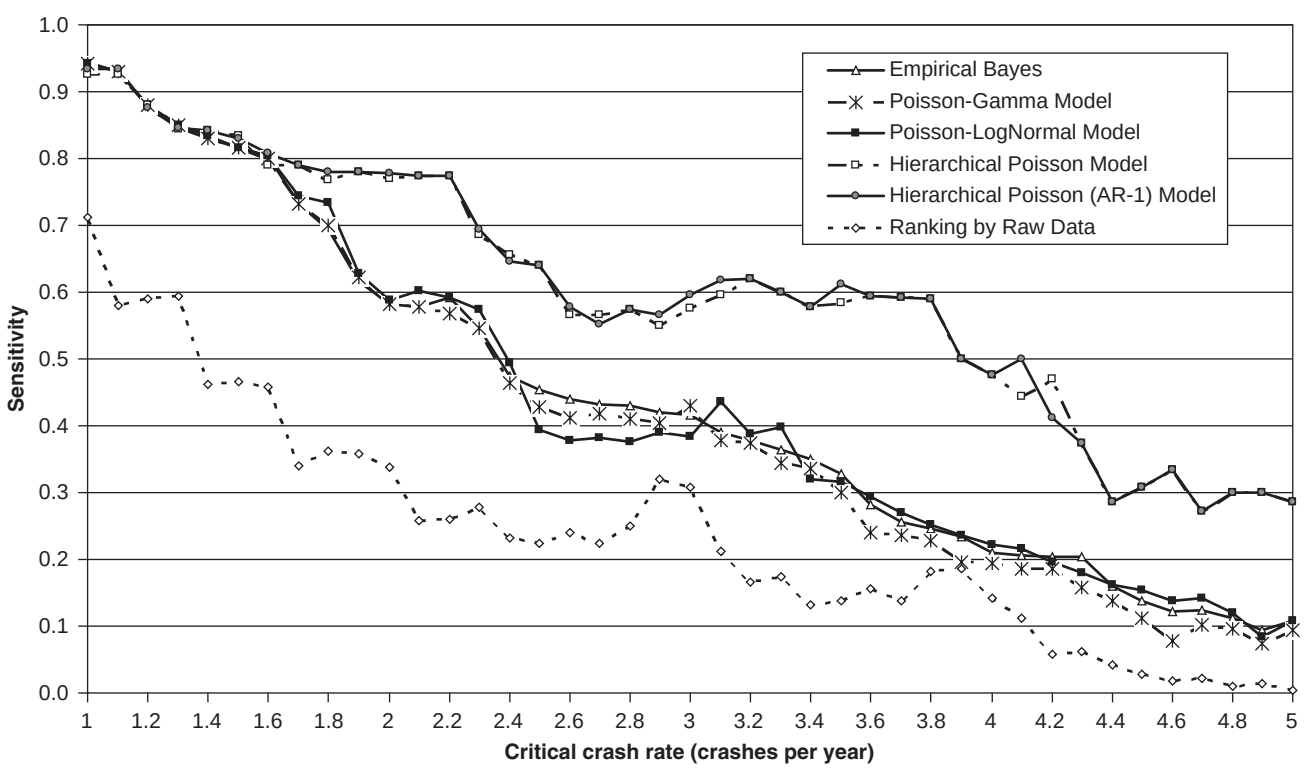

FIGURE 2 Sensitivity of alternative ranking approaches given critical crash rate.

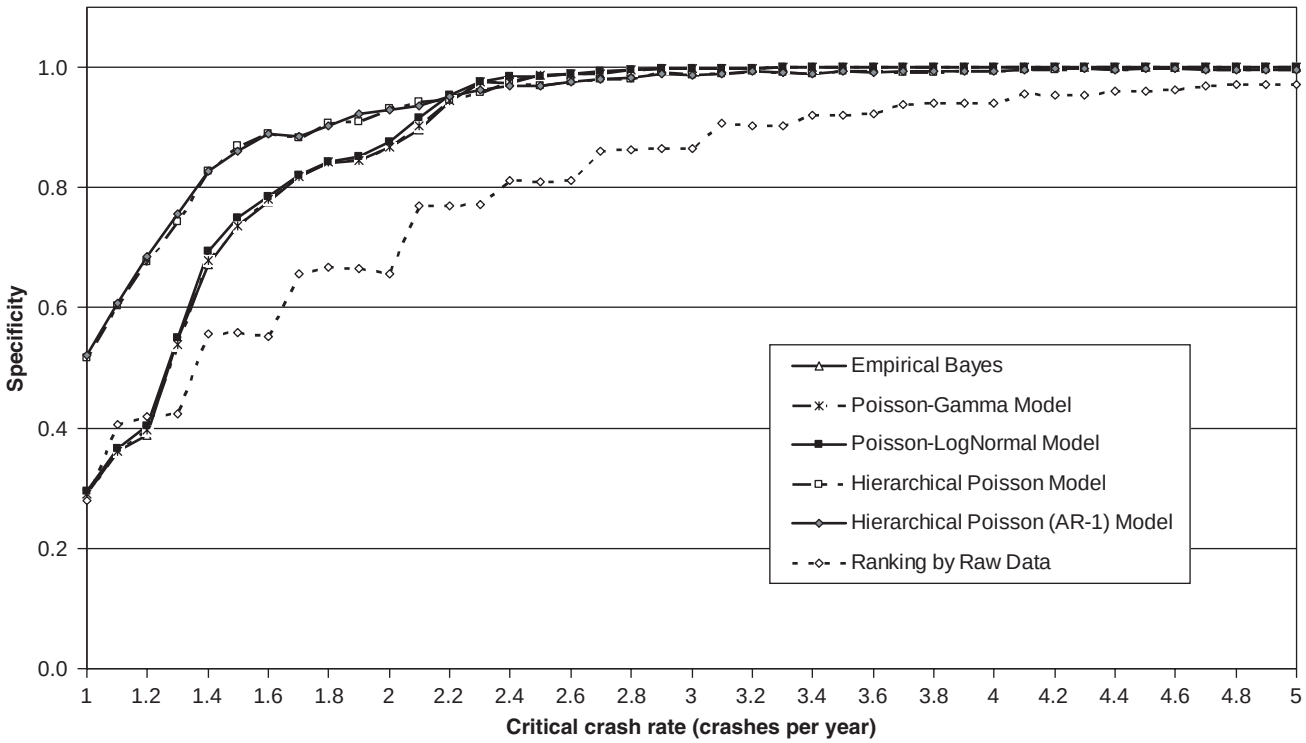

FIGURE 3 Specificity of alternative ranking approaches given critical crash rate.

sites with very high crash risk will be selected for further investigation and remedial treatment. Hence, the outperformance of the FB approach with hierarchical models becomes more important, particularly for practice purposes. In addition, as shown in Figures 2 and 3, there seems to be no obvious discrepancy among EB, PG, and PLN models as well as between HP and HP (AR-1) models.

Furthermore, the results of sensitivity for cutoff number of sites from 10 to 300 are illustrated in Figure 4. It should be noted that, for FIT by giving cutoff number of sites, there is no need to duplicate reporting the specificity, because given a fixed number of sites for TP, the FP will be equal to the FN. As in the aforementioned result, the ranking criterion by raw crash data in the FIT by giving cutoff number of sites has the worst performance too. Compared with other model-based approaches, the discrepancy becomes larger when the number of sites selected is small (e.g., <100). Likewise, EB and FB using the NB model again have comparable capability in correctly identifying hot spots. A slight outperformance of FB using the PLN model over EB and FB using the NB model is found, which is more significant when the number of hot spots selected is between 20 and 80 sites, approximately 3% to 14% of the overall sites. The HP and HP (AR-1) models have equivalently the best performance in correctly identifying hot spots. Compared with standard EB approaches, this outperformance is on average 25% when fewer than 200 sites (34% of all sites) are screened as hot spots for further engineering study.

Percentage of Dispersion Deviations

To further examine the deviations between alternative model-based ranking approaches, the percentage of dispersion deviations using the decision parameter $\Theta_{i,2}$ was calculated for alternative FB models compared with the standard EB approach, which is shown in Figure 5. As mentioned previously, the decision parameter $\Theta_{i,2}$ is able to measure the dispersion to which traffic sites exhibit abnormally high unobserved random effects by comparing the "average" expected values at similar sites. The result shows that the alternative FB approach using different crash prediction models may also lead to considerably different lists of hot spots using deviation as the ranking decision parameter. The deviation is the most significant for the FB estimate using the PLN model.

Mean Absolute Deviations

Finally, to yield a quantitative understanding about the discrepancy of performance of alternative approaches, the MADs were calculated by comparing model prediction at a specific site ($\hat{\lambda}_{it}$) to the mean of the crash number observed in the 10 years ($\hat{\lambda}_i^{10}$):

$$\text{MAD} = \frac{1}{582}\sum_{\forall i}\left|\frac{1}{3}\sum_{t=1,2,3}\hat{\lambda}_{it} - \hat{\lambda}_i^{10}\right|$$

The results shown in Figure 6 indicate that HP and HP (AR-1) models have the best predictive performance to the long-term mean of crash frequency [MAD: 0.394 in the HP model and 0.383 in the HP (AR-1) model]. Naive estimation using a simple mean of crash count in the 3 years has an average deviation from the long-term mean of 0.845 crash per intersection. Again, EB and FB ranking using the NB model have equivalent performances (MAD: 0.567 in the EB approach and 0.566 in the NB model). It is not surprising because both the point safety estimates and overdispersion parameter used in the EB approach were estimated from the NB model.

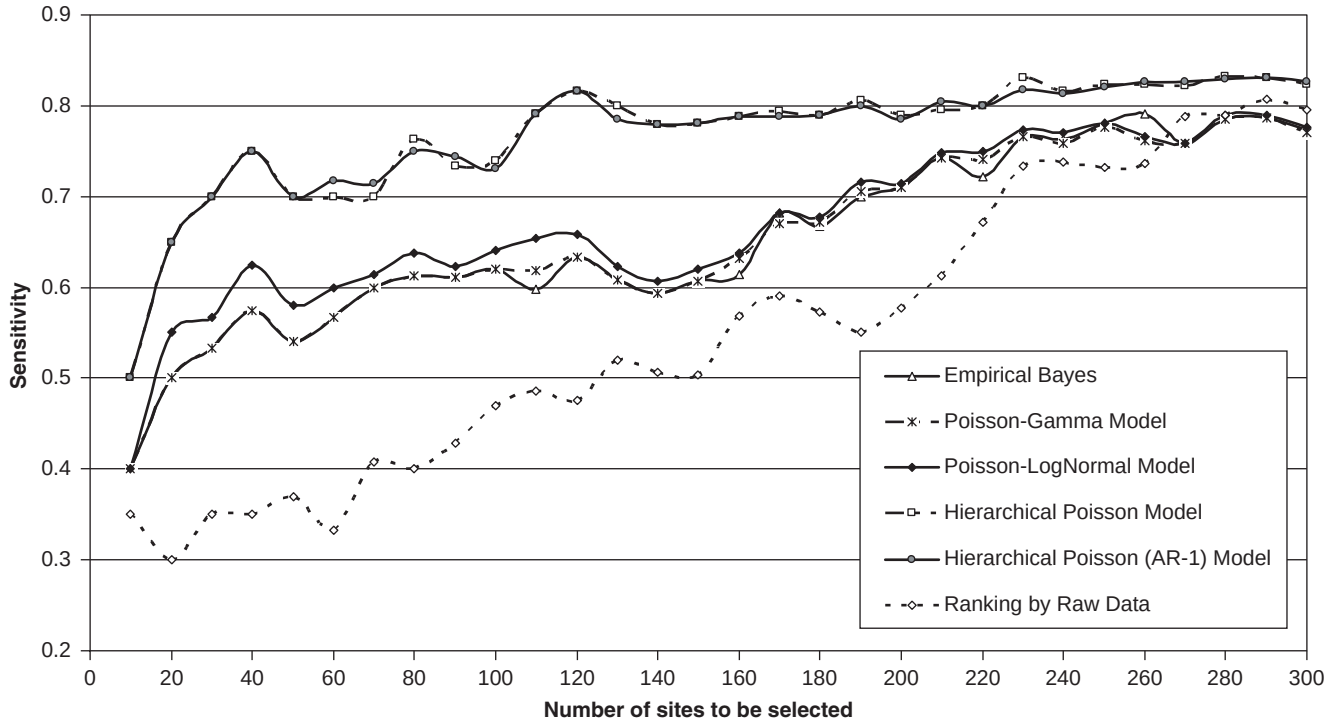

FIGURE 4 Sensitivity of alternative ranking approaches given cutoff number of sites.

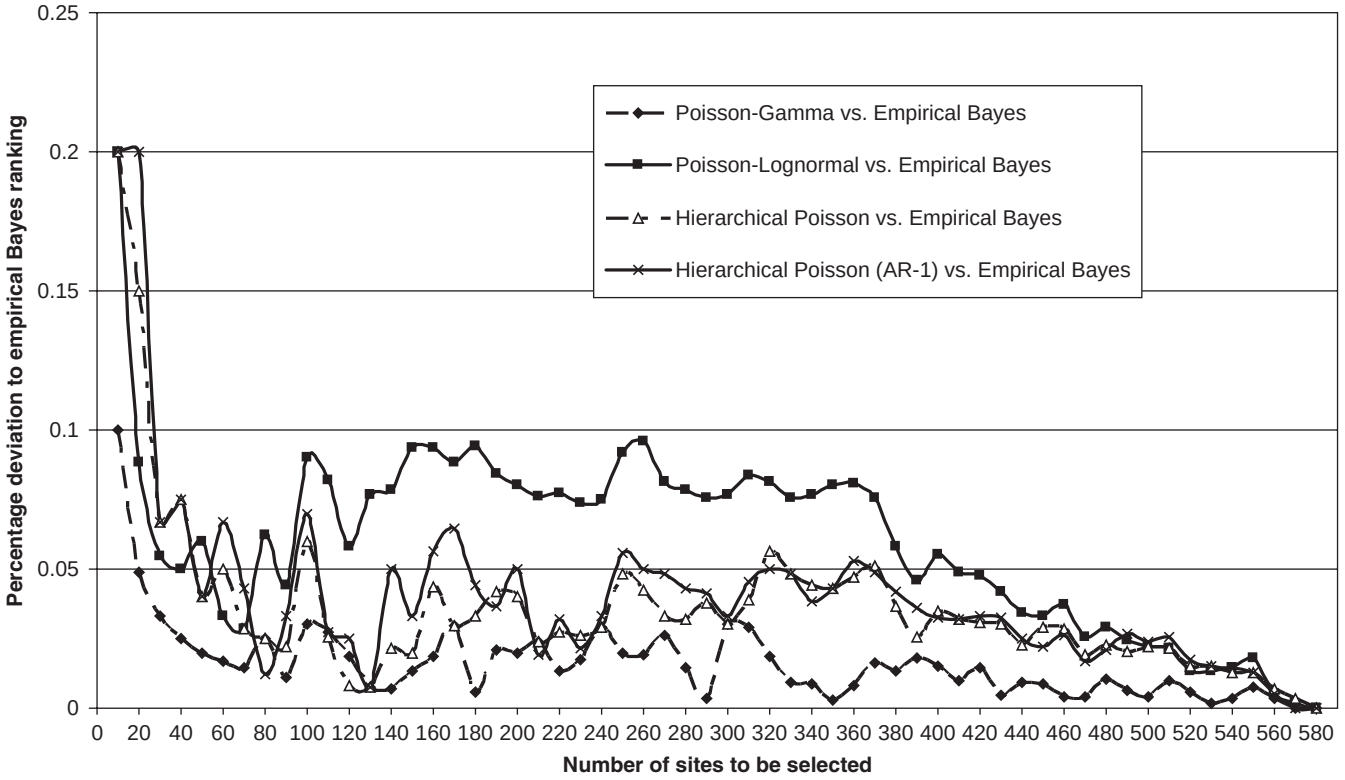

FIGURE 5 Percentage of dispersion deviations of alternative FB models to EB approach.

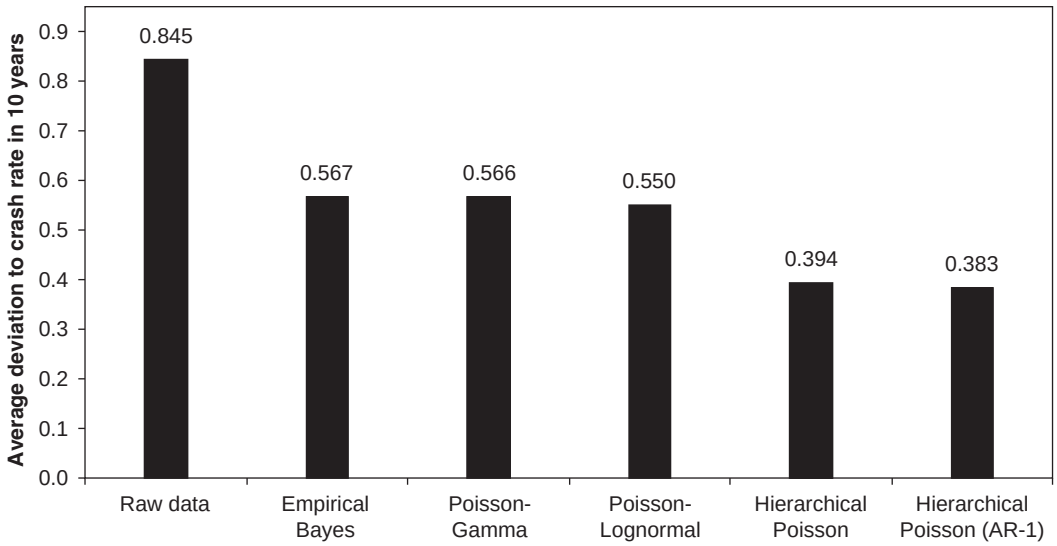

FIGURE 6 Mean absolute deviations of alternative estimates to long-term crash rate.

Moreover, as in the model-fitting diagnostic, there is only a slight difference in model prediction accuracy between the PLN model (MAD: 0.550) and the NB model.

CONCLUSIONS, LIMITATIONS, AND FUTURE RESEARCH

This research proposed a framework for the model-based hot spot identification method by applying the FB technique. Compared with the standard EB approach, an essential theoretical advantage of the FB approach is the capability to seamlessly integrate prior information and all available data into posterior distributions of unknowns. Particularly in the context of hot spot identification, all uncertainties of associations of crash risk and various risk factors could be accommodated to estimate a posterior distribution representing the predicted site safety on which various ranking criteria could be based. Moreover, by use of hierarchical model specification, the FB approach is able to flexibly take into account various heterogeneities of crash occurrence due to spatial and temporal effects on traffic safety.

Using four-legged signalized intersection data collected in Singapore, a comparison study was conducted to empirically evaluate the performance of model fitting and hot spot identification of alternative approaches. With respect to model fitting, results showed that the hierarchical models [HP model and HP (AR-1) model] with accommodation for site-specific effect and serial correlation have better goodness of fit than the nonhierarchical models (NB model and PLN model).

In hot spot identification, a comprehensive comparison was conducted between six alternative approaches: naive ranking using raw data, EB ranking, FB ranking using the NB model, FB ranking using the PLN model, FB ranking using the HP model, and FB ranking using the HP (AR-1) model. Without knowing a priori which sites are truly hazardous, the empirical estimation using the long-term crash rate was assumed to be the true safety situation at specific sites. For the purpose of comparison, several evaluation measurements, including the improved FIT, percentage of dispersion deviations, and MAD, were employed and demonstrated to be useful. Specifically, it was found that all model-based approaches perform significantly better in safety ranking than the naive approach using a raw crash count. This result is consistent with most relevant studies. It confirms the necessity of taking into account the regression-to-the-mean effect and various heterogeneities associated with diverse site features. Furthermore, the FB hierarchical models [i.e., HP model and HP (AR-1) model] were found to outperform the standard EB approach in correctly identifying hot spots. This finding indicates that the flexibility in model specification of the FB approach can generate great potentials to improve the accuracy of identifying hazardous sites by explicitly accounting for various structured heterogeneities in crash data.

Although this study has dealt with a number of aspects in applying the FB approach for hot spot identification, two limitations are worth mentioning and future research expanding from this study is recommended. First, although the serial correlation coefficient in the HP (AR-1) model was found to be significant, it failed to provide a remarkable improvement of model fitting and safety ranking over the HP model. Nevertheless, it is still reasonable to suspect the presence of potential serial correlation because time-dependent traffic factors naturally exist in reality. Some of these factors may be unobservable, or may not be easily specified in the model. Essentially, serial correlation has already been identified using crash data in several recent studies (e.g., *22*). Further evaluation study is needed to investigate the necessity of accommodating the possible time-dependent uncertainties by using longer time series crash data.

Second, no clear distinction of predictive performance was identified in this study between the EB approach and the FB approach with nonhierarchical models (i.e., NB model and PLN model). In other words, the potential benefit of the FB approach in better accounting for uncertainties associated with parameter estimation of SPF was not adequately demonstrated in this empirical evaluation. Moreover, although empirical data have the advantage of essentially representing the realism, the study could be criticized for using the empirical estimation (i.e., long-term crash rate in this study) to establish the true "safety" of candidate sites for the purpose of evaluation. Alternatively, several previous studies have argued the advantage of using simulated data in evaluating hot spot identification methods (*5, 9, 10, 30*). Future research could be conducted using simulated

data to experimentally investigate the application potential of the FB approach in traffic safety evaluation.

ACKNOWLEDGMENTS

The authors thank Singapore Traffic Police and TSM Consultancy Pte Ltd for providing crash and geometric data of intersections used in this study.

REFERENCES

1. Deacon, J. A., C. V. Zegeer, and R. C. Deen. Identification of Hazardous Rural Highway Locations. In *Transportation Research Record 543,* TRB, National Research Council, Washington, D.C., 1975, pp. 16–33.
2. Stokes, R. W., and I. M. Mutabazi. Rate–Quality Control Method of Identifying Hazardous Road Locations. In *Transportation Research Record 1542,* TRB, National Research Council, Washington, D.C., 1996, pp. 44–48.
3. Tamburri, T. N., R. N. Smith, J. P. Mills, and V. J. Perini. The Safety Index: A Method of Evaluating and Rating Safety Benefits. In *Highway Research Record 307,* HRB, National Research Council, Washington, D.C., 1970, pp. 28–43.
4. Hauer, E. *Observational Before–After Studies in Road Safety.* Pergamon–Elsevier Science, Inc., Tarrytown, N.Y., 1997.
5. Elvik, R. *State-of-the-Art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks.* Report 883. Institute of Transport Economics, Oslo, Norway, 2007.
6. Hauer, E. Identification of Sites with Promise. In *Transportation Research Record 1542,* TRB, National Research Council, Washington, D.C., 1996, pp. 54–60.
7. 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.
8. Hauer, E., and B. N. Persaud. Problem of Identifying Hazardous Locations Using Accident Data. In *Transportation Research Record 975,* TRB, National Research Council, Washington, D.C., 1984, pp. 36–43.
9. Cheng, W., and S. P. Washington. Experimental Evaluation of Hotspot Identification Methods. *Accident Analysis & Prevention,* Vol. 37, 2005, pp. 870–881.
10. Cheng, W., and S. P. Washington. New Criteria for Evaluating Methods of Identifying Hot Spots. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2083,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 76–85.
11. Carlin, B., and T. Louis, *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, New York, 2000.
12. Miaou, S.-P., and D. Lord. Modeling Traffic-Flow Relationships at Signalized Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes Methods. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 31–40.
13. Gelman, A., J. B. Carlin, and H. S. Stern. *Bayesian Data Analysis,* 2nd ed. Chapman and Hall, New York, 2003.
14. 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.
15. Song, J. J., M. Ghosh, S. Miaou, and B. Mallick. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis,* Vol. 97, 2006, pp. 246–273.
16. MacNab, Y. C. A Bayesian Hierarchical Model for Accident and Injury Surveillance. *Accident Analysis and Prevention,* Vol. 35, 2003, pp. 91–102.
17. Huang, H. L., H. C. Chin, and M. M. Haque. Severity of Driver Injury and Vehicle Damage in Traffic Crashes at Intersections: A Bayesian Hierarchical Analysis. *Accident Analysis and Prevention,* Vol. 40, 2008, pp. 45–54.
18. Huang, H., H. C. Chin, and M. M. Haque. Bayesian Hierarchical Analysis of Crash Prediction Models. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
19. Miaou, S. P. The Relationship Between Truck Accidents and Geometric Design of Road Section: Poisson Versus Negative Binomial Regression. *Accident Analysis and Prevention,* Vol. 26, 1994, pp. 471–482.
20. Shankar, V. N., F. L. Mannering, and W. Barfield. Effect of Roadway Geometric and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis and Prevention,* Vol. 27, 1995, pp. 371–389.
21. Lord, D., and L. F. Miranda-Moreno. Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-Gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective. *Safety Science,* Vol. 46, 2008, pp. 751–770.
22. Wang, X., and M. Abdel-Aty. Temporal and Spatial Analyses of Rear-End Crashes at Signalized Intersections. *Accident Analysis and Prevention,* Vol. 38, 2006, pp. 1137–1150.
23. Congdon, P. *Applied Bayesian Modelling.* John Wiley and Sons Inc., New York, 2003.
24. Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and V. D. Linde. Bayesian Measures of Model Complexity and Fit (with discussion). *Journal of the Royal Statistical Society,* Vol. B64, No. 4, 2003, pp. 583–616.
25. Lord, D., S. P. Washington, and J. N. Ivan. Poisson, Poisson-Gamma and Zero-Inflated Regression Models for Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and Prevention,* Vol. 37, 2005, pp. 35–46.
26. Lord, D., S. P. Washington, and J. N. Ivan. Further Notes on the Application of Zero-Inflated Models in Highway Safety. *Accident Analysis and Prevention,* Vol. 39, 2007, pp. 53–57.
27. Chapman, R. The Concept of Exposure. *Accident Analysis and Prevention,* Vol. 5, 1973, pp. 95–110.
28. Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn. *WinBUGS Version 1.4.1 User Manual.* MRC Biostatistics Unit, Cambridge, United Kingdom, 2003.
29. Brooks, S. P., and A. Gelman. Alternative Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics,* Vol. 7, 1998, pp. 434–455.
30. Maher, M. J., and L. J. Mountain. The Identification of Accident Blackspots: A Comparison of Current Methods. *Accident Analysis and Prevention,* Vol. 20, 1988, pp. 143–151.

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