

1                   **Application of Bayesian Informative Priors to Enhance the**  
2                   **Transferability of Safety Performance Functions**

3  
4  
5  
6  
7  
8  
9                   **Ahmed Farid\***  
10                  **Mohamed Abdel-Aty**  
11                  **Jaeyoung Lee**  
12                  **Naveen Eluru**

13  
14  
15  
16  
17  
18  
19  
20  
21                  \*Corresponding Author

22  
23                  Department of Civil, Environmental and Construction Engineering  
24                               University of Central Florida  
25                               Orlando, Florida 32816-2450  
26                                       (407) 530-9360  
27                               Ahmedtf91@knights.ucf.edu

28  
29  
30  
31  
32  
33                  January 2017  
34  
35  
36  
37  
38

**ABSTRACT**

Safety performance functions (SPFs) are essential tools for highway agencies to predict crashes, identify hotspots and assess safety countermeasures. In the Highway Safety Manual (HSM), a variety of SPFs are provided for different types of roadway facilities, crash types and severity levels. Agencies, lacking the necessary resources to develop own localized SPFs, may opt to apply the HSM's SPFs for their jurisdictions. Yet, municipalities that want to develop and maintain their regional SPFs might encounter the issue of the small sample bias. Bayesian inference is being conducted to address this issue by combining the current data with prior information to achieve reliable results. It follows that the essence of Bayesian statistics is the application of informative priors, obtained from other SPFs or experts' experiences. In this study, we investigate the applicability of informative priors for Bayesian negative binomial SPFs for rural divided multilane highway segments in Florida and California. An SPF with non-informative priors is developed for each state and its parameters' distributions are assigned to the other state's SPF as informative priors. The performances of SPFs are evaluated by applying each state's SPFs to the other state. The analysis is conducted for both total (KABCO) and severe (KAB) crashes. As per the results, applying one state's SPF with informative priors, which are the other state's SPF independent variable estimates, to the latter state's conditions yields better goodness of fit (GOF) values than applying the former state's SPF with non-informative priors to the conditions of the latter state. This is for both total and severe crash SPFs.

**Keywords:** Bayesian Informative Priors, Negative Binomial Models, Markov Chain Monte Carlo Simulations, Highway Safety Manual, Transferability

## 1 INTRODUCTION

2  
3 Safety performance functions (SPFs), regression models used for predicting crash frequencies,  
4 are an indispensable constituent of traffic safety analyses. They are used to predict crashes of any  
5 type or severity. The HSM (AASHTO, 2010) provides SPFs by facility type, crash type and  
6 severity. The SPF development process involves regressing traffic crash counts with contributing  
7 factors. Modeling crashes using the ordinary linear least square regression framework is faulty  
8 since crash patterns are non-negative integers (Miaou et al., 1993; Miaou, 1994; Kim et al., 2005;  
9 Garber and Wu, 2001). Generalized linear models are applied for crash analysis studies (Sawalha  
10 and Sayed, 2006; Taylor et al., 2002; Harnen et al., 2004; Donnel and Mason, 2006). The  
11 Poisson model is proposed as it has a statistically appropriate structure for non-negative integer  
12 data. Yet, it offers a restriction that the mean crash frequency is equivalent to the variance. The  
13 condition in which the variance is greater than the mean is often observed in crash data and is  
14 termed over-dispersion. The negative binomial model, a type of Poisson-gamma models, relax  
15 the restrictive assumption of the Poisson model and is used as the conventional modeling  
16 framework in estimating SPFs in the literature (Lord et al., 2005). Often, municipalities develop  
17 and update jurisdiction specific SPFs with the least possible number of independent variables, or  
18 factors, that contribute to crashes. A typical problem encountered is the lack of available  
19 resources to gather data of sufficient sample sizes (Heydari et al., 2014). Application of the  
20 maximum likelihood estimation (MLE) method to obtain the optimal estimates of the model  
21 parameters produces inaccurate results under limited data whereas the full Bayesian method is an  
22 efficient alternative approach applicable to circumvent the limited sample size problem.

23 The essence of the full Bayesian method is that prior knowledge is taken into account by  
24 introducing informative prior distributions for the parameters. It is demonstrated that even for  
25 small sample sizes, the full Bayesian method yields unbiased estimates of the parameters  
26 (Heydari et al., 2014). The other advantage of the full Bayesian approach is that unlike the  
27 Frequentist approach, where the traditional MLE method is applied, the probability that an  
28 estimate of a parameter falls within a certain range is obtained by direct computation. In the  
29 Frequentist approach, it can be interpreted that the estimates are within confidence limits if  
30 multiple samples are collected (Heydari et al., 2014). Investigating the application of prior  
31 information to achieve accurate parameter estimates is a topic of interest for researchers in road  
32 safety as is discussed in the following section (Heydari et al., 2014, Lord and Miranda-Moreno,  
33 2008; Part et al., 2010; Miranda-Moreno et al., 2013; Yu and Abdel-Aty, 2013; Washington and  
34 Oh, 2006; Jang et al., 2010; Haleem et al., 2010). Yet, the main objective of these studies is to  
35 develop informative priors for modeling crashes for regions with limited data. The objective of  
36 this study is clearly distinguished from those of the previous studies. In our study, the goal is to  
37 examine the influence of informative priors in the transferability of SPFs. That is, when  
38 transferring an SPF from one jurisdiction to another, an investigation is made to check whether  
39 adopting informative prior information on the parameters from the jurisdiction where the SPF is  
40 being transferred will facilitate the transferability of the SPF. We aim at developing Bayesian  
41 negative binomial SPFs with non-informative priors for rural divided multilane highway  
42 segments in two states, which are Florida and California. Then, each state's data are provided  
43 with parameter estimates of the other state to re-develop SPFs with informative priors. We  
44 evaluate the improvement in model fit based on the consideration of informative priors.

## 1 LITERATURE REVIEW

2  
3 Few studies are focused on the application of Bayesian priors in traffic crash modeling. Lord and  
4 Miranda-Moreno (2008) address the low sample mean and size problem encountered by  
5 transportation authorities when collecting data for developing SPFs. The authors simulate and  
6 model crash data using negative binomial and Poisson-lognormal formulations. Specifically,  
7 several datasets are simulated with varying sample sizes and means. Non-informative and  
8 informative priors are tested for Bayesian modeling of the simulated data. For the models with  
9 informative priors, a shape parameter of 0.1 and a scale of 1 are assigned to the inverse  
10 dispersion parameter which is assumed to follow a gamma distribution. According to the  
11 authors' findings, modeling small datasets with low means leads to an inaccurate estimation of  
12 the inverse dispersion parameter's distribution mean. This is particularly true in the case where  
13 non-informative priors are applied. The chances of encountering the poor accuracy problem are  
14 reduced when informative priors are provided for the inverse dispersion parameter.

15 Park et al. (2010) address the unobserved heterogeneity for datasets of low sample means and  
16 sizes. As described earlier, several datasets are simulated with different sample means and sizes.  
17 Negative binomial, finite mixture negative binomial and finite mixture Poisson models are  
18 applied. The analysis is conducted using Bayesian inference once with non-informative priors  
19 and once with informative priors for the inverse dispersion parameter that follows a gamma  
20 distribution. The priors are 0.5 and 0.1 for the shape and scale respectively. Finite mixture  
21 negative binomial models exhibit the best fit since they account for the unobserved heterogeneity  
22 more accurately than the traditional negative binomial models. Nevertheless, the results of the  
23 finite mixture negative binomial models with non-informative priors demonstrate that for  
24 datasets of low means and sample sizes, the estimated posterior distribution mean of the inverse  
25 dispersion parameter is biased. Based on the findings, the authors suggest minimum sample sizes  
26 of 300, 500 and 1,500 sites for datasets with high, medium and low mean crash frequencies,  
27 respectively.

28 In another study by Miranda-Moreno et al. (2013), the authors investigate the benefit of  
29 incorporating knowledge of previous experiences in the Bayesian inference framework when  
30 having limited data. The study is aimed at identifying hot spots in rural three-leg intersections in  
31 California and four-lane highway segments in Texas. The data include both divided and  
32 undivided segments. Several specifications of the prior distributions of the inverse dispersion  
33 parameter are tested for different sample sizes. The informative prior distributions, applied, are  
34 the uniform distribution, Christiansen's distribution and the gamma distribution. Results of the  
35 Bayesian models with the gamma and uniform informative prior distributions are superior to  
36 those of the models with the informative prior distribution of Christiansen for data of 50 sites.  
37 Furthermore, under low sample sizes, informative prior models outperform non-informative prior  
38 models. However, for data of large samples of more than 100 sites, the benefit of assigning  
39 informative priors to the inverse dispersion parameter is less pronounced.

40 Heydari et al. (2014) offer a practical Bayesian methodology to estimate and update SPFs by  
41 making use of both data of limited samples and the national HSM's SPF parameters. The study is  
42 conducted for rural undivided multilane highways. A sensitivity analysis is undertaken by  
43 proposing 15 combinations of informative and non-informative prior specifications for both the  
44 independent variables and the inverse dispersion parameter. According to the results, there is no  
45 single combination that yields the best estimate for each regression coefficient including that of

1 the inverse dispersion parameter. Yet, two combinations outperform the others in terms of  
2 overall model fit. One is a combination in which a variance of 10 is assumed for the independent  
3 variables while the HSM's provided parameters are assumed as the means. The ratio of the  
4 squared mean to the variance of the inverse dispersion parameter is assigned as the shape prior of  
5 the inverse dispersion parameter that follows a gamma distribution. The scale, assigned, is the  
6 ratio of the mean to the variance of the inverse dispersion parameter. The second best  
7 combination is similar to the aforementioned one. However, the inverse dispersion parameter is  
8 exponentially distributed and assigned the inverse of its mean as an informative prior.

9 Yu and Abdel-Aty (2013) examine four approaches to formulate Bayesian informative priors for  
10 negative binomial and Poisson-lognormal models of crashes in a freeway section in Colorado.  
11 One approach is the two-stage Bayesian updating approach where past data are modeled while  
12 non-informative priors are assigned to all independent variables and the inverse dispersion  
13 parameter. The output means and variances of the variables' posterior distributions act as the  
14 informative priors for modeling the training data. The second approach is the MLE approach  
15 where the likelihood function is maximized and the estimated variable coefficients are used as  
16 informative priors for Bayesian modeling. Third, the method of moments is one where the  
17 dataset's independent variables' means and variances are used as informative priors for the  
18 analysis and the variables are assumed to be normally distributed. The last approach is the expert  
19 experience method which is based on experts' judgments on crash contributing factors provided  
20 by means of surveys. The authors conclude that the two-stage updating approach is the superior  
21 one. If, however, past data are limited, the authors recommend the method of moments approach  
22 since its performance is not substantially different from that of the MLE approach. Overall, it is  
23 concluded that the performances of the models with informative priors are superior to those with  
24 non-informative priors.

25 Washington and Oh (2006) develop crash modification factors (CMFs) for railroad crossing  
26 crashes in South Korea while relying on expert knowledge. Jang et al. (2010) estimate zero-  
27 inflated Poisson and zero inflated negative binomial models with the likelihood function raised  
28 to an exponent which is an informative prior. Also, in a study by Haleem et al. (2010), a  
29 reliability analysis is undertaken using the Bayesian updating approach with informative priors to  
30 accurately estimate crash frequencies at three-leg and four-leg un-signalized intersections in  
31 Florida. Overall, there is a general interest in the application of Bayesian inference particularly to  
32 address the issue of developing SPFs in cases where data are of limited sample size. As  
33 previously discussed, historical data are implemented as the informative prior information and  
34 the limited current data, collected, are used for Bayesian analysis. Indeed, the analysis is efficient  
35 since the small sample size problem is circumvented.

36 In this study, Bayesian negative binomial SPFs are developed to estimate KABCO and KAB  
37 crashes for rural divided multilane highway segments in Florida and California. Each state's SPF  
38 is developed with non-informative priors. Then, each state's SPF parameters are assigned the  
39 estimated regression coefficients' distributions of the other state's model as informative priors  
40 and the SPFs are re-developed. The aim of this study is similar to those of Yu and Abdel-Aty  
41 (2013) and Heydari et al. (2014). However, the advantages of the proposed research are two-fold.  
42 The research not only checks the efficiency of Bayesian informative priors but also assesses  
43 whether informative priors from elsewhere are applicable to the jurisdiction of interest. This  
44 study is a follow up on a previous investigation by Farid et al. (2016) where the transferability of  
45 SPFs, taking the functional form of those of the HSM, are explored among the same

1    aforementioned states. In the previous study, it is concluded that Florida and California's SPF of  
2    multiple vehicle crashes are to an extent transferable. This study is of great value for roadway  
3    agencies not having the resources to collect and process data in that Bayesian SPFs can be  
4    developed swiftly. The SPFs can be estimated efficiently and at a reduced cost by borrowing  
5    informative priors from other regions as opposed to spending excessively on retrieving and  
6    preparing historical data. In the following sections, the preparation of the data, details of the  
7    research methodology and analysis results are discussed.

## 8    **DATA PREPARATION**

9  
10   The data of KABCO and KAB crashes at rural divided multilane highway segments in Florida  
11   and California are employed in our study. Geometric characteristics are included as well.  
12   Specifically, Florida's road geometric data are obtained from the Roadway Characteristics  
13   Inventory of the Florida Department of Transportation (FDOT). The crash data of Florida are  
14   retrieved from the Crash Analysis Reporting System (CARS), which also belongs to FDOT.  
15   There are 436 homogenous segments with a total of 1,114 crashes sampled from Florida. On the  
16   other hand, the California data, including crash and road geometric characteristics, are obtained  
17   from the Highway Safety Information System (HSIS). It is a database that includes traffic data  
18   on multiple states including not only California but also Ohio, Washington, Minnesota, Illinois,  
19   North Carolina and Maine (Highway Safety Information System). In California, the samples,  
20   obtained, are 1,153 segments with 3,887 crashes. Note that Florida's crash data are from the  
21   years 2009 to 2011 while those of California are from 2009 to 2010. The descriptive statistics of  
22   both states' data are shown in Table 1. The minimum segment length, sampled, is not less than  
23   0.1 mi to be consistent with the HSM standards. The AADT, in Table 1, is the average annual  
24   daily traffic.

1

**Table 1 Descriptive Statistics of the States' Data**

<b>State</b>	<b>Florida (N=436)</b>				<b>California (N=1,153)</b>			
<b>Variable</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Maximum</b>	<b>Mean</b>	<b>Standard Deviation</b>	<b>Minimum</b>	<b>Maximum</b>
<b>Segment Length (mi)</b>	0.804	1.459	0.100	18.078	0.516	0.572	0.100	5.329
<b>AADT (veh/day)</b>	12,681.930	8,709.680	2,500	49,500	21,243.260	14,642.420	2,325	79,500
<b>Lane Width (ft)</b>	12.008	0.205	10	13	11.995	0.204	11	13
<b>Shoulder Width (ft)</b>	4.752	1.669	1.5	12	6.370	2.013	0	11
<b>Median Width (ft)</b>	40.429	23.504	8	140	42.699	30.238	5	99
	<b>Crashes Per Hundred Million Vehicle Miles Traveled Per Year</b>							
<b>KABCO</b>	22.704				43.712			
<b>KAB</b>	8.213				7.928			

2

1 Large differences in geometric characteristics may inhibit improvements in model fits after  
 2 incorporating informative priors. From Table 1, it can be observed that Florida's shoulders are  
 3 narrower than those of California, as indicated by the means. The variation in shoulder widths in  
 4 Florida is marginally less than that of shoulder widths of California as well. It should also be  
 5 noted that the variation in median widths is larger for California's segments. Also, California  
 6 experiences a greater KABCO crash rate while the opposite is the case for KAB crashes.

## 7 RESEARCH METHODOLOGY

8  
 9 The negative binomial framework is used as the traditional structure for developing SPFs. That is  
 10 because the over-dispersion is accommodated (Lord et al., 2005). The research approach,  
 11 employed, is the application of Bayesian inference to estimate negative binomial models. Under  
 12 the negative binomial model, the probability,  $p$ , of a crash at site,  $i$ , is given by the following  
 13 function (Lord and Miranda-Moreno, 2008):

$$14 \quad p_i = \frac{\Gamma(y_i + \varphi)}{\Gamma(\varphi)y_i!} \times \left[ \frac{\varphi}{\mu_i + \varphi} \right]^\varphi \times \left[ \frac{\mu_i}{\mu_i + \varphi} \right]^{y_i} \quad (1)$$

15 The term,  $\varphi$ , denotes the inverse dispersion parameter, while  $y_i$  is the observed crash frequency  
 16 and  $\mu_i$  is the mean crash frequency at site  $i$ . The negative binomial model converges to the  
 17 Poisson model as  $\varphi \rightarrow \infty$ . The mean crash frequency per site is a function of the estimated  
 18 independent variable coefficients,  $\beta_j$ 's, the site characteristics whether roadway geometrics,  
 19 demographics or any other characteristics,  $X_{ij}$ 's, and finally the inverse dispersion parameter's  
 20 function,  $\theta$ . The mean crash frequency is defined as follows (Lord and Miranda-Moreno, 2008):

$$21 \quad \mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_j X_{ij}) \times \theta_i \quad (2)$$

22 The parameter,  $\theta$ , is a function of the inverse dispersion parameter,  $\varphi$ , both of which are assumed  
 23 to follow gamma distributions as follows (Lord and Miranda-Moreno, 2008).

$$24 \quad \theta \sim \Gamma(\varphi, \varphi) \quad (3)$$

$$25 \quad \varphi \sim \Gamma(a, b) \quad (4)$$

26 The terms,  $a$  and  $b$ , are the shape and scale parameters of the gamma distribution. In this study,  
 27 the mean crash frequency function, which is the prediction equation, is considered as follows.

$$28 \quad \hat{\mu}_i = \exp[\hat{A} + \hat{B} \times \ln(AADT) + \ln(\text{segment length} \times \text{year})] \quad (5)$$

29 The last term is not associated with a coefficient to be consistent with the HSM's SPFs for rural  
 30 divided multilane highway segments. The year is an offset variable that takes into consideration  
 31 the fact that crashes occurred throughout multiple years and that the mean function yields the  
 32 crash counts per year.

33 According to Bayes' theorem, which forms the basis of the Bayesian inference, both the  
 34 likelihood of a crash and the parameters' prior distributions are required. The posterior  
 35 probability of the vector of parameters,  $\beta$ , can be interpreted as the following (Heydari et al.,  
 36 2014):

$$37 \quad f(\beta, \gamma | y) \propto f(y | \beta) \times f(\beta | \gamma) \times f(\gamma) \quad (6)$$



1 The vectors,  $\mathbf{y}$  and  $\boldsymbol{\gamma}$ , are those of the crashes, observed, and parameters of the inverse dispersion  
 2 parameter which follows the gamma distribution. Thus, the posterior probability is proportional  
 3 to the product of the likelihoods of crashes in the current dataset,  $f(\mathbf{y}/\boldsymbol{\beta})$ , the prior knowledge of  
 4 the independent variables,  $f(\boldsymbol{\beta}/\boldsymbol{\gamma})$ , and that of the inverse dispersion parameter,  $f(\boldsymbol{\gamma})$ . The posterior  
 5 distribution cannot be obtained by direct computation. Integrations with a large number of  
 6 dimensions are required. Markov Chain Monte Carlo (MCMC) simulations including the  
 7 Metropolis-Hastings and the Gibbs methods are applicable (Heydari et al., 2014; Cowles, 2013).

8 First, a model with non-informative priors for the independent variables is developed for each  
 9 state. That is, each variable is assigned a prior of 0 and a precision factor, which is the inverse of  
 10 the variance, of 0.001 to indicate no prior information. The variables are assumed to follow  
 11 normal distributions. Likewise, the dispersion parameter is assigned a shape and a scale of 0.01  
 12 each. The models are run in R software (R Core Team) and WinBUGS open source software  
 13 (Lunn et al., 2000) using the R package, R2WinBUGS (Sturtz, 2005), for MCMC simulations.  
 14 The model runs comprise of 3 chains of 120,000 iterations, of which 12,000 are burn-in  
 15 iterations, and 120,000 simulations. The results are recorded and checked for model convergence  
 16 by inspecting the Kernel density plots and trace plots (Cowles, 2013). The estimated parameter  
 17 means and standard deviations of each state's model are recorded and assigned to the other state  
 18 as informative priors. Subsequently, the models are rerun with the informative priors. It should  
 19 be noted that no prior information is assumed for the dispersion parameter throughout all models.  
 20 Each state's models, including those with informative priors and those with non-informative  
 21 priors, are applied to the other state for validation. The deviance information criterion (DIC) is  
 22 used for evaluating the fit of Bayesian SPFs (Spiegelhalter et al., 2002) and is defined as shown  
 23 in Equation (7).

$$24 \quad DIC = \bar{D} + p_D \quad (7)$$

25 The term,  $\bar{D}$ , is the average of the SPF's deviance while  $p_D$  is the effective number of  
 26 parameters. Furthermore, other GOF measures are computed. They are the mean absolute  
 27 deviation (MAD) and mean square prediction error (MSPE). Defined as follows, the GOF  
 28 measures, applied in this study, are dependent on the difference between the observed and  
 29 predicted crash frequency per site.

$$30 \quad MAD = (1/n) \sum_{i=1}^n |N_{SPF i} - N_{obs i}| \quad (8)$$

$$31 \quad MSPE = (1/n) \sum_{i=1}^n (N_{SPF i} - N_{obs i})^2 \quad (9)$$

32 The terms,  $n$ ,  $N_{SPF i}$  and  $N_{obs i}$  are the number of sites, predicted number of crashes, which is the  
 33 mean crash frequency, and observed number of crashes per site  $i$ , respectively. The GOF  
 34 measures are calculated to check whether the informative prior models are superior to the non-  
 35 informative prior models when applied elsewhere.

## 36 **EMPIRICAL ANALYSIS**

37

38 The analysis is conducted by running 4 Bayesian SPFs for KABCO crashes and 4 for KAB  
 39 crashes. According to the Kernel density and trace plots, all models converged successfully.  
 40 Results of each state's KABCO crash SPFs with non-informative priors are shown in Table 2  
 41 while results of those with informative priors are presented in Table 3. Similarly, the KAB crash  
 42 SPFs with non-informative priors are shown in Table 4 and those with informative priors are

1 shown in Table 5. As previously mentioned, the means and precisions (i.e., inverse of variances)  
 2 of the parameters, not including those of the inverse dispersion parameter, are used as the  
 3 informative priors. Summaries of the results of the application of the models with informative  
 4 priors for validation are presented in Table 6. Note that in Table 6, results of applying  
 5 jurisdiction specific SPFs, whether with non-informative or informative priors, to local  
 6 conditions are not presented. Instead, we are interested in transferring SPFs to other state(s) and  
 7 examining the effect of informative priors, retrieved from the state of which the SPF is  
 8 transferred to, on the transferability of the SPF.  
 9

10 **Table 2 States' Non-Informative Prior Model Results for KABC0 Crashes**

Model	Parameter	Mean	Standard Deviation	Bayesian Credible Interval				
				Lower 95%	Lower 90%	Median	Upper 90%	Upper 95%
Florida (N=436)	Constant	-6.448	0.983	-8.355	-8.043	-6.483	-4.744	-4.418
	Ln(AADT)	0.709	0.105	0.493	0.528	0.712	0.879	0.912
	$\phi$	1.147	0.161	0.871	0.907	1.134	1.432	1.499
	Deviance	1,144	28.880	1,089	1,097	1,144	1,192	1,202
	pD	171.535						
	DIC	1,315.520						
California (N=1,153)	Constant	-11.140	0.488	-12.110	-11.970	-11.140	-10.360	-10.210
	Ln(AADT)	1.247	0.049	1.154	1.169	1.246	1.330	1.344
	$\phi$	1.440	0.107	1.243	1.272	1.436	1.624	1.663
	Deviance	3,523	45.280	3,435	3,449	3,522	3,598	3,613
	pD	521.498						
	DIC	4,044.400						

11

12

13

**Table 3 States' Informative Prior Models for KABC0 Crashes**

Model	Parameter	Mean	Standard Deviation	Bayesian Credible Interval				
				Lower 95%	Lower 90%	Median	Upper 90%	Upper 95%
Florida with California's Informative Priors (N=436)	Constant	-10.840	0.325	-11.490	-11.370	-10.840	-10.310	-10.210
	Ln(AADT)	1.179	0.035	1.111	1.122	1.179	1.236	1.248
	$\phi$	1.041	0.142	0.795	0.828	1.030	1.291	1.349
	Deviance	1,142	29.000	1,087	1,096	1,142	1,191	1,201
	pD	176.175						
	DIC	1,318.400						
California with Florida's Informative Priors (N=1,153)	Constant	-9.502	0.408	-10.300	-10.170	-9.503	-8.833	-8.723
	Ln(AADT)	1.083	0.041	1.005	1.016	1.083	1.149	1.162
	$\phi$	1.394	0.103	1.205	1.232	1.390	1.570	1.607
	Deviance	3,513	44.940	3,426	3,440	3,512	3,588	3,602
	pD	526.824						
	DIC	4,039.570						

1

**Table 4 States' Non-Informative Prior Model Results for KAB Crashes**

Model	Parameter	Mean	Standard Deviation	Bayesian Credible Interval				
				Lower 95%	Lower 90%	Median	Upper 90%	Upper 95%
Florida (N=436)	Constant	-3.700	1.098	-5.768	-5.446	-3.744	-1.845	-1.440
	Ln(AADT)	0.294	0.117	0.054	0.097	0.299	0.481	0.515
	$\phi$	1.985	0.645	1.101	1.191	1.867	3.169	3.572
	Deviance	771.5	26.980	719.5	727.6	771.3	816.300	825
	pD	80.836						
	DIC	852.351						
California (N=1,153)	Constant	-11.090	0.737	-12.510	-12.310	-11.090	-9.894	-9.640
	Ln(AADT)	1.067	0.073	0.923	0.948	1.067	1.188	1.208
	$\phi$	2.214	0.528	1.447	1.531	2.127	3.194	3.478
	Deviance	1,804	42.020	1,723	1,736	1,803	1,874	1,887
	pD	166.889						
	DIC	1,970.470						

2

3

4

**Table 5 Results of States' Informative Prior Models for KAB Crashes**

Model	Parameter	Mean	Standard Deviation	Bayesian Credible Interval				
				Lower 95%	Lower 90%	Median	Upper 90%	Upper 95%
Florida with California's Informative Priors (N=436)	Constant	-9.879	0.476	-10.810	-10.670	-9.876	-9.100	-8.949
	Ln(AADT)	0.954	0.051	0.854	0.870	0.953	1.038	1.054
	$\phi$	1.399	0.373	0.844	0.905	1.342	2.084	2.282
	Deviance	766.800	27.270	714.100	722.400	766.600	812.200	820.900
	pD	92.188						
	DIC	859.019						
California with Florida's Informative Priors (N=1,153)	Constant	-7.683	0.545	-8.769	-8.573	-7.686	-6.779	-6.594
	Ln(AADT)	0.727	0.054	0.619	0.637	0.728	0.816	0.836
	$\phi$	1.943	0.438	1.288	1.363	1.880	2.733	2.966
	Deviance	1,800	42.870	1,717	1,730	1,799	1,871	1,885
	pD	178.209						
	DIC	1,977.890						

5

6

7

1

**Table 6 Model Validation Results**

SPF of KABCO Crashes	Application Data			
	Florida (N=436)		California (N=1,153)	
	MAD	MSPE	MAD	MSPE
Florida with Non-Informative Priors	NA		2.384	27.986
Florida with California's Informative Priors	NA		2.134	20.886
California with Non-Informative Priors	3.294	80.938	NA	
California with Florida's Informative Priors	3.367	70.291	NA	
<b>SPF of KAB Crashes</b>				
Florida with Non-Informative Priors	NA		0.619	1.073
Florida with California's Informative Priors	NA		0.613	0.935
California with Non-Informative Priors	0.820	2.777	NA	
California with Florida's Informative Priors	0.808	2.079	NA	

2

### 3 *Remarks on the SPFs with Non-Informative Priors*

4 The AADT is associated with crash occurrence as can be shown in Tables 2 and 4. The  
5 relationship is nearly linear for California. Also, it is worth noting that several types of SPF  
6 specifications are attempted besides the model structure of Equation (5). In addition to the  
7 natural log transformation of the AADT, the lane width, shoulder width and median width  
8 variables are included as independent variables. However, the Bayesian credible interval  
9 includes 0 for these variables' values prompting the variables' removal.

### 10 *Remarks on the SPFs with Informative Priors*

11 When applying a state's KABCO crash SPF to another state while borrowing the latter state's  
12 independent variable values as informative priors improves the GOF relative to applying the  
13 former state's SPF with non-informative priors to the latter state as shown in Table 6. Florida's  
14 KABCO crash SPF with informative priors from California exhibits a lower MAD and MSPE  
15 than that with non-informative priors when both SPFs are applied to California. On the other  
16 hand, California's KABCO SPF with Florida's informative priors outperforms its respective SPF  
17 with non-informative priors when both SPFs are applied to Florida. Even though, the MAD  
18 increased by 0.073 (3.294 to 3.367), the MSPE dropped considerably by 10.647 (80.938 to  
19 70.291). The MSPE is a more effective measure than the MAD since the MSPE is more sensitive  
20 to large residuals, which are deviations between observed and predicted crash counts needed for  
21 calculation of both measures. It is demonstrated in the SPF transferability study by Farid et al.  
22 (2016) that Florida's SPFs, for crashes involving at least two vehicles, are to an extent  
23 transferable to California and vice versa possibly due to similarities in demographics, topography  
24 and weather conditions in both states. Florida and California are characterized by tourism even  
25 though Florida experiences more rainfall than California. In addition, it is critical to note the  
26 unobserved difference attributed to crash reporting thresholds in both states. The thresholds in  
27 Florida and California are \$500 (Florida Statutes) and \$750 (Xie et al., 2011) worth of damage to  
28 property, respectively.

29 The KAB crash SPF results are consistent with those of KABCO crash SPFs. The MADs and  
30 MSPEs are reduced when applying Florida's KAB crash SPF with California's informative  
31 priors to California compared with applying Florida's KAB crash SPF with non-informative

1 priors to California. On the other hand, assigning Florida's independent variables as informative  
2 priors to the California KAB crash SPF and applying the SPF to Florida improves the GOF  
3 values relative to applying the California KAB crash SPF with non-informative priors to Florida.  
4 Nevertheless, it is crucial to mention that differences in reporting thresholds affect the count of  
5 KABCO crash records but are highly unlikely to affect those of KAB crashes.

6

## 7 **SUMMARY AND CONCLUSIONS**

8

9 Employing informative priors for modeling crash frequencies is an area that is explored to a  
10 limited extent. Recent studies contribute to the solution of accurately estimating crashes with  
11 data having limited sample sizes by the application of informative priors within the context of  
12 the Bayesian inference. In our study, the use of informative priors from Florida and California is  
13 tested where each state is assigned the priors of the other state. When transferring an SPF from  
14 one jurisdiction to another, borrowing informative priors from the latter, when developing the  
15 SPF for the former, facilitates the transferability of the developed SPF to the latter. The Florida  
16 KABCO crash SPF with California's priors exhibits a better GOF than the Florida SPF with non-  
17 informative priors when both SPFs are applied to California. The opposite is also true. The crash  
18 predictions of California's KABCO SPF with Florida's informative priors, applied to Florida's  
19 conditions, are more accurate than those of California's KABCO SPF with non-informative  
20 priors applied to Florida's conditions.

21 The KAB crash SPF results are consistent with those of KABCO crash SPFs. Florida's KAB  
22 crash SPF with California's informative priors demonstrates better performance when applied to  
23 California's conditions compared to Florida's KAB crash SPF with non-informative priors  
24 applied to California's conditions. Also, the GOF results are improved when applying the  
25 California KAB SPF with Florida's informative priors to Florida relative to applying California's  
26 KAB SPF with non-informative priors to Florida. Ultimately, the findings of this study indicate  
27 that when developing SPFs, roadway agencies, lacking staff with expertise, may provide  
28 available data for another region's municipality to extract the informative priors. The  
29 municipality, elsewhere, may develop SPFs with the priors, extracted, and transfer the SPFs to  
30 the locality of roadway agencies. Preferably, the agencies would provide their data to  
31 municipalities of regions with similar roadway design characteristics, weather trends,  
32 demographics and topographic characteristics.

33 It is imperative to discuss the future work to build upon this study. The analysis, undertaken, is  
34 for distant states specifically for rural divided multilane highways. The investigation can be  
35 conducted for neighbor states and for other types of roadway facilities. Furthermore, throughout  
36 the entire analysis procedure, non-informative priors are assigned to the inverse dispersion  
37 parameter even though informative priors are applied for the independent variables. Using  
38 informative priors for the inverse dispersion parameter, obtained from another region's SPF, is a  
39 viable option. It should also be noted that the HSM's SPFs' dispersion parameter for rural  
40 divided multilane highway segments is a function with an estimable coefficient rather than a  
41 fixed parameter. That is, every site's dispersion parameter is different than that of the other sites,  
42 which is not the case of this study. Modifying the dispersion parameter structure is a convenient  
43 technique to reduce bias. However, care should be taken to avoid including a large number of  
44 parameters in the dispersion parameter formula especially if the model size is large (Mitra and

1 Washington, 2007). Another issue to caution is the application of uniform prior distributions,  
2 known as diffuse priors, because inaccuracies will result in the estimated posterior distributions  
3 (Natarajan and McCulloch, 1998). Also, the independent variables used were limited to the daily  
4 traffic even though the lane width, shoulder width and median width are attempted and exhibit  
5 Bayesian intervals that include 0. One issue to investigate for future analysis is the addition of  
6 more variables such as the roadside hazard rating, presence of horizontal curvature and street  
7 lighting, among others that are not necessarily available in each state. Another is the inclusion of  
8 socio-demographic and topographic variables since crashes are not only affected by road  
9 geometrics (Abdel-Aty et al., 2013; Lee et al., 2014a; Lee et al., 2014b; Lee et al., 2015a; Lee et  
10 al., 2015b). Finally, the spatial correlation among crashes can and ought to be accounted for in  
11 future studies (Ma et al., 2008).

12

### 13 **ACKNOWLEDGMENTS**

14

15 The authors would like to express their gratitude for NCHRP 17-62's funds. Special thanks are  
16 conveyed to the research team headed by Drs. John Ivan, Raghavan Srinivasan and Bhagwant  
17 Persaud. All opinions, stated in this manuscript, are those of the authors only.

18

### 19 **REFERENCES**

20

21 Farid, A., Abdel-Aty, M., Lee, J., Eluru, N., Wang, J.-H., 2016. Exploring the Transferability of  
22 Safety Performance Functions. *Accident Analysis and Prevention* 94, 143-152.

23

24 Jang, H., Lee, S., Kim, S., 2010. Bayesian Analysis for Zero-Inflated Regression Models with  
25 the Power Prior: Applications to Road Safety Countermeasures. *Accident Analysis and*  
26 *Prevention* 42, 540-547.

27

28 Park, B.-J., Lord, D., Hart, J., 2010. Bias Properties of Bayesian Statistics in Finite Mixture of  
29 Negative Binomial Regression Models in Crash Data Analysis. *Accident Analysis and*  
30 *Prevention* 42 (2), 741-749.

31

32 AASHTO, 2010. Highway Safety Manual. American Association of State Highway and  
33 Transportation Officials, Washington D.C.

34

35 Abdel-Aty, M., Lee, J., Siddiqui, C., Choi, K., 2013. Geographical Unit Based Analysis in the  
36 Context of Transportation Safety Planning. *Transportation Research Part A: Policy and Practice*  
37 49, 62-75.

38

39 Cowles, M., 2013. *Applied Bayesian Statistics with R and OpenBUGS Examples*. Springer, New  
40 York.

41

42 Donnell, E., Mason, J., 2006. Predicting the Frequency of Median Barrier Crashes on  
43 Pennsylvania Interstate Highways. *Accident Analysis and Prevention* 38 (3), 590-599.

- 1  
2 Florida Statutes. Title XXIII – Motor Vehicle. The Florida Legislature, Tallahassee, Florida.  
3 [http://www.leg.state.fl.us/Statutes/index.cfm?App\\_mode=Display\\_Statute&Search\\_String=&UR](http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=0300-0399/0316/Sections/0316.065.html)  
4 [L=0300-0399/0316/Sections/0316.065.html](http://www.leg.state.fl.us/Statutes/index.cfm?App_mode=Display_Statute&Search_String=&URL=0300-0399/0316/Sections/0316.065.html). Accessed Oct. 28, 2015.  
5
- 6 Garber, N., Wu, L., 2001. Stochastic Models Relating Crash Probabilities with Geometric and  
7 Corresponding Traffic Characteristics Data. Presented to National Intelligent Transportation  
8 Systems Implementation Research Center of the Department of Transportation, Charlottesville,  
9 Virginia.  
10
- 11 Haleem, K., Abdel-Aty, M., Mackie, K., 2010. Using a Reliability Process to Reduce  
12 Uncertainty in Predicting Crashes at Unsignalized Intersections. *Accident Analysis and*  
13 *Prevention* 42 (2), 654-666.  
14
- 15 Harnen, S., Umar, R., Wong, S., Hashim, W., 2004. Development of Prediction Models for  
16 Motorcycle Crashes at Signalized Intersections on Urban Roads in Malaysia. *Journal of*  
17 *Transportation and Statistics* 7 (2), 27-39.  
18
- 19 Heydari, S., Miranda-Moreno, L., Lord, D., Fu, L., 2014. Bayesian Methodology to Estimate and  
20 Update Safety Performance Functions under Limited Data Conditions: a Sensitivity Analysis.  
21 *Accident Analysis and Prevention* 64, 41-51.  
22
- 23 Highway Safety Information System. State Data. Highway Safety Research Center, University of  
24 North Carolina, Chapel Hill, North Carolina. <http://www.hsisinfo.org/>. Accessed Sept. 28, 2015.  
25
- 26 Kim, S., Chung, S., Song, K., Chon, K., 2005. Development of an Accident Prediction Model  
27 using GLIM (Generalized Log-linear Model) and EB method: A case of Seoul. *Journal of the*  
28 *Eastern Asia Society for Transportation Studies* 6, 3669-3682.  
29
- 30 Lee, J., Abdel-Aty, M., Choi, K., 2014a. Analysis of Residence Characteristics of at-Fault  
31 Drivers in Traffic Crashes. *Safety Science* 68, 6-13.  
32
- 33 Lee, J., Abdel-Aty, M., Choi, K., Huang, H., 2015a. Multi-Level Hot Zone Identification for  
34 Pedestrian Safety. *Accident Analysis and Prevention* 76, 64-73.  
35
- 36 Lee, J., Abdel-Aty, M., Jiang, X., 2014b. Development of Zone System for Macro-Level Traffic  
37 Safety Analysis. *Journal of Transport Geography* 38, 13-21.  
38
- 39 Lee, J., Abdel-Aty, M., Jiang, X., 2015b. Multivariate Crash Modeling for Motor Vehicle and  
40 Non-Motorized Modes at the Macroscopic Level. *Accident Analysis and Prevention* 78, 146-  
41 154.  
42

- 1 Lord, D., Miranda-Moreno, L., 2008. Effects of Low Sample Mean Values and Small Sample  
2 Size on the Estimation of the Fixed Dispersion Parameter of Poisson-Gamma Models for  
3 Modeling Motor Vehicle Crashes: a Bayesian Perspective. *Safety Science* 46 (5), 751-770.  
4
- 5 Lord, D., Washington, S., Ivan, J., 2005. Poisson, Poisson-gamma and Zero-Inflated Regression  
6 Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory. *Accident Analysis and  
7 Prevention* 37 (1), 35-46.  
8
- 9 Lunn, D., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS—a Bayesian Modelling  
10 Framework: Concepts, Structure, and Extensibility. *Statistics and Computing* 10 (4), 325–337.  
11
- 12 Ma, J., Kockelman, K., Damien, P., 2008. A Multivariate Poisson-Lognormal Regression Model  
13 for Prediction of Crash Counts by Severity, Using Bayesian Methods. *Accident Analysis and  
14 Prevention* 40 (3), 964-975.  
15
- 16 Miaou, S.-P., 1994. The Relationship between Truck Accidents and Geometric Design of Road  
17 Sections: Poisson versus Negative Binomial Regressions. *Accident Analysis and Prevention* 26  
18 (4), 471-482.  
19
- 20 Miaou, S.-P., Lum, H., 1993. Modeling Vehicle Accidents and Highway Geometric Design  
21 Relationships. *Accident Analysis and Prevention*, 25 (6), 689-709.  
22
- 23 Miranda-Moreno, L., Heydari, S., Lord, D., Fu, L., 2013. Bayesian Road Safety Analysis:  
24 Incorporation of Past Evidence and Effect of Hyper-Prior Choice. *Journal of Safety Research* 46,  
25 31-40.  
26
- 27 Mitra, S., Washington, S., 2007. On the Nature of Over-Dispersion in Motor Vehicle Crash  
28 Prediction Models. *Accident Analysis and Prevention* 39 (3), 459-468.  
29
- 30 Natarajan, R., McCulloch, C., 1998. Gibbs Sampling with Diffuse Proper Priors: a Valid  
31 Approach to Data-Driven Inference? *Journal of Computational and Graphical Statistics* 7 (3),  
32 267-277.  
33
- 34 R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for  
35 Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, 2013. <http://www.r-project.org/>.  
36 Accessed Jan. 20, 2016.  
37
- 38 Sawalha, Z., Sayed, T., 2006. Traffic Accident Modeling: Some Statistical Issues. *Canadian  
39 Journal of Civil Engineering* 33 (9), 1115-1124.  
40
- 41 Spiegelhalter, D., Best, N., Carlin, B., van der Linde, A., 2002. Bayesian Measures of Model  
42 Complexity and Fit. *Journal of the Royal Statistical Society: Series B* 64 (4), 583-639.



- 1  
2 Sturtz, S., Ligges, U., Gelman, A., 2005. R2WinBUGS: a Package for Running WinBUGS from  
3 R. *Journal of Statistical Software* 12 (3), 1-16.  
4  
5 Taylor, M., Baruya, A., Kennedy, J., 2002. The Relationship between Speed and Accidents on  
6 Rural Single-Carriageway Roads. Publication TRL511. Transport Research Laboratory, United  
7 Kingdom.  
8  
9 Washington, S., Oh, J., 2006. Bayesian Methodology Incorporating Expert Judgment for  
10 Ranking Countermeasure Effectiveness Under Uncertainty: Example Applied to at Grade  
11 Railroad Crossings in Korea. *Accident Analysis and Prevention* 38, 234–247.  
12  
13 Xie, F., Gladhill, K., Dixon, K., Monsere, C., 2011. Calibration of Highway Safety Manual  
14 Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the*  
15 *Transportation Research Board* 2241, 19-28.  
16  
17 Yu, R., Abdel-Aty, M., 2013. Investigating Different Approaches to Develop Informative Priors  
18 in Hierarchical Bayesian Safety Performance Functions. *Accident Analysis and Prevention* 56,  
19 51-58.