Eur. Transp. Res. Rev. (2012) 4:167–174
DOI 10.1007/s12544-012-0072-y

ORIGINAL PAPER

Sensitivity of a real-time freeway crash prediction model to calibration optimality

Amir Samimi · Bruce Hellinga

Received: 26 October 2010 / Accepted: 4 January 2012 / Published online: 18 January 2012

Abstract
Background Real-time crash prediction models are often structured as general log-linear categorical models which must be calibrated using an extensive database. However, there is no method to optimally select the number of categories and the values that define the boundaries between categories when representing continuous measures as categorical variables within the log-linear model. This raises the question of how important the calibration is to the safety impacts estimated when using the crash prediction model. In this paper, we examined the impact that the process used to calibrate the crash prediction model has on estimates of safety impacts of a variable speed limit system.
Methods Two calibration methods were compared, namely a heuristic ad hoc method and a nearoptimal method. Both methods were applied to calibrate a crash prediction model using the same set of data from an urban freeway in Ontario, Canada. The calibrated crash prediction models are used to evaluate the safety benefits of a candidate variable speed limit system under three different traffic demand levels (Peak, Near-Peak, and Off-Peak).
Concluding remarks It was found that safety improvements estimated by the two calibrated crash prediction models are within approximately 13% of each other for the Peak and Near-Peak scenarios, but differ by a larger amount for the Off-Peak scenario. However, despite these differences in the estimated magnitude of the safety impacts, the sign of the impact (i.e. increase versus decrease in safety) were consistent irrespective of the calibration method used. The results suggested that the safety impacts provided by the crash prediction model are robust in that they are relatively insensitive to the optimality of the calibration.

Keywords Crash prediction models · Sensitivity to calibration · Log-linear categorical estimation

A. Samimi (✉)
Department of Civil Engineering, Sharif University of Technology,
Azadi Avenue,
Tehran, Iran
e-mail: asamimi@sharif.edu

B. Hellinga
Department of Civil and Environmental Engineering,
University of Waterloo,
200 University Ave. West,
Waterloo, ON N2L 3GI, Canada
e-mail: bhellinga@uwaterloo.ca

1 Introduction

Traditional safety analysis focuses on determining locations that have experienced higher than expected number of collisions and then identifying actions, typically geometric improvements that can improve safety. More recently, attention has been given to identifying, in real-time, traffic conditions that are associated with high crash prediction, and when such conditions are observed, to intervene using various freeway traffic management options such as ramp metering and variable speed limits [8, 9]. These models are often referred to as "real-time crash prediction models".

Real-time crash prediction models can be used in combination with micro-level traffic simulation models to estimate the safety impacts of proposed traffic control strategies or geometric changes [1]. However, the real-time crash prediction models, which are often structured as general log-linear categorical models, must be calibrated using an extensive database containing roadway geometric characteristics, weather conditions, crash records, and real-time traffic conditions (typically determined from loop detector data). This makes the development of a crash prediction model costly and unaffordable in cases of insufficient accident statistics [15]. Furthermore, different approaches to model calibration may result in different values for model parameters. The determination of model adequacy is made on the basis of goodness of fit statistics; however, several models all with different parameter values may pass these statistical tests. The important practical question is what impact do these different parameter values have on the safety evaluations that are conducted using the crash prediction models? If the differences are small, the method used to calibrate the crash prediction model does not need to be overly accurate and could be performed with considerably less resources. However, if the differences are large, then much more care must be applied in selecting and employing a calibration method, as the final safety evaluation results will be highly dependent on the calibration process. Given the effort and cost required to calibrate crash prediction models, transferring model parameters that are not very sensitive to the local data and are unlikely to vary among jurisdictions with fairly similar conditions seems reasonable [15]. In this paper we determine the sensitivity of safety impacts to the safety model calibration. Specifically, we use two methods to calibrate a log-linear crash prediction model to a single set of data from a section of an urban freeway near Toronto, Ontario, Canada. These two crash prediction models are then used in combination with a simulation model to evaluate the safety impacts of a candidate variable speed limit system. Differences in the safety impact estimates are examined to provide insights into the sensitivity of the safety impacts to the crash prediction model calibration methods.

2 Methodology

The crash prediction model employed in this study, originally developed by Lee et al. [10–12], calculates crash frequency as a function of traffic conditions, external control factors and exposure in a log-linear modeling form. Also, it can incorporate a value of exposure that is associated with each traffic condition and external control factors. Equation 1 shows the log-linear function developed by Lee et al. to calculate crash potential.

$$\frac{F}{EXP^{\beta}} = \exp\left(\theta + \lambda_{CV(i)} + \lambda_{Q(j)} + \lambda_{COVV(k)} + \lambda_{R(l)} + \lambda_{P(m)}\right) \tag{1}$$

Where,

F: expected number of crashes
EXP: exposure (veh-km)
β: parameter for exposure
θ: constant
$\lambda_{CVS(i)}$: effect of the crash precursor variable CVS having i levels
$\lambda_{Q(i)}$: effect of the crash precursor variable Q having j levels
$\lambda_{COVV(i)}$: effect of the crash precursor variable COVV having k levels
$\lambda_{R(l)}$: effect of road geometry (control factor) having l levels
$\lambda_{P(m)}$: effect of time of day (control factor) having m levels.

The measures of traffic conditions, termed crash precursors, represent the traffic flow conditions a short time prior to a crash occurrence. Abdel-Aty and Pande [2] and Golob et al. [6] identified specific traffic conditions that increase crash propensity. Lee et al. [11–13] also explored a number of traffic flow characteristics and found three of them to be the most statistically powerful in predicting crash occurrence: 1) temporal variation of speed at a fixed location; 2) longitudinal variation in speed; and 3) lane changing behavior. These three crash precursors can be calculated on the basis of volume, speed, and occupancy as measured by dual loop detectors. The precursors are continuous variables; however they are captured within the log-linear crash prediction model as categorical variables (Eq. 1).

The temporal variation of speed at a fixed location represents the stability of speeds between vehicles in a traffic stream. Low variation of speed is an indication of smooth traffic flow in which vehicles are traveling at nearly constant speeds. An increase in this speed variation indicates more variability in the speed choice among drivers. This in turn requires drivers to adjust speeds more frequently, leading to the deterioration of flow stability and a higher risk of driver error and an impending crash situation. The temporal variation of speed is measured by the coefficient of variation of speed (precursor CVS), calculated at the nearest detector station upstream of a crash location. CVS is a measure of dispersion which normalizes the standard deviation and is explained in more detail by Lee et al. [13].

The spatial (longitudinal) variation of speed along road sections measures the difference in average travel speeds between two consecutive loop detector stations. A small spatial variation indicates near constant speeds. However, a large spatial variation in speed indicates traffic will experience an abrupt change in travel speed, requiring either sudden acceleration or deceleration. A state of sudden deceleration may increase the likelihood of rear-end collisions and often occurs as a traffic queue is formed during recurrent or non-recurrent congestion. A detailed description of the method for calculating the spatial variation of speed, represented by precursor Q, is provided by Allaby [3].

Lane changing behavior, an indication of turbulence in the traffic stream, is estimated by the average covariance of volume difference between upstream and downstream locations across adjacent lanes (precursor COVV). The covariance of volume captures the correlation of traffic volume changes between two lanes (i.e. traffic moving from lane 1 to lane 2 creates a volume reduction on lane 1 and subsequently a volume increase on lane 2). Thus, COVV is a surrogate measure of lane changing activity. High levels of COVV indicate frequent lane changing that may contribute to an increased likelihood of a crash occurrence. The mathematical formulation for estimating COVV is expressed by Lee (2004).

Other studies have also found that measures of turbulence (particularly variance of speed) within the traffic stream can be linked to crash likelihood [5, 7, 14]. External factors including road geometry, time of day and environmental conditions, was also included in the model, since these factors alone can affect driver behavior. Lee et al. [10–13] found that freeway segments with merging or diverging traffic contribute more to crash potential than straight freeway segments with no changes in lane configuration. Time of day refers to peak and off-peak periods. Typically, traffic volumes and congestion are higher during peak periods and drivers, particularly commuters, may react more aggressively to maintain their schedules. These factors are likely to increase the likelihood of a crash occurrence during the peak periods. Lastly, environmental conditions include such factors as local weather, road surface quality, and lighting. Due to the limited amount of available environmental data, the effect of these factors can be difficult to capture.

In Eq. 1, exposure forms a relationship between the frequency of traffic and environmental events and the associated crash frequency. For example, consider two traffic scenarios. The first scenario arises 20% of the time and experiences 20% of all crashes. The second scenario also experiences 20% of all crashes, but arises only 5% of the time. Although the crash frequencies are identical, the crash rate is clearly higher for the second scenario. Lee et al. [11–13] expressed exposure as the number of vehicle kilometers exposed to each combination of traffic characteristics and external control factors. In other words, the average vehicle-kilometers present over a road section are multiplied by the probability of a certain time of day (peak or off-peak) at a certain type of road geometry (merge/diverge or straight) under a certain range of crash precursor values.

Since the crash prediction model is log-linear, crash precursors are categorized. The probabilities of the precursors in the expression for exposure depend on the selection of categories. A category spanning a larger range of crash precursor values results in a higher probability for exposure than a category spanning a small range of values.

3 Data

A segment of the Queen Elizabeth Way (QEW) in Mississauga, Ontario was chosen to calibrate the crash prediction model for this study. The QEW is a multilane freeway located in south western Ontario, Canada. The freeway begins near the Canadian/American border at Fort Eerie and, following the coastline of Lake Ontario, passes through several urban centers such as Niagara Falls, St. Catharines, Hamilton, Burlington, Mississauga, and finally into Toronto. The QEW near Toronto services a large volume of commuter traffic in the morning and evening peak periods, resulting in heavy congestion and a high frequency of crashes.

The segment used for calibration was a 13 km section between Royal Windsor Dr. and Highway 427 including both directions of travel. This freeway segment features a posted speed limit of 100 km/hr, has 3 to 4 mainline lanes, and experiences a directional AADT of about 70,000 vehicles. The section is instrumented with dual loop detector stations in each mainline lane spaced at approximately 600 m. The study segment contains 26 loop detector stations in each travel direction. Every 20 s, speed, volume, and occupancy are recorded for all mainline stations.

Detector data and Freeway Traffic Management System (FTMS) incident logs were obtained from the Ministry of Transportation of Ontario (MTO). Crash records were compiled for the period of January 1998 through February 2003. The FTMS incident logs provided several pieces of information on every incident detected on the highway, including date and time incident was reported, identity of upstream loop detector station, and type of incident.

Using the information provided, the FTMS incident logs were filtered to form a crash database with records appropriate for the crash model calibration. Loop detector data were obtained for the upstream and downstream location of each crash, for a 30 min time period before and after the reported time of the crash. The time of the crash was estimated on the basis of the traffic stream speed measured by the upstream and downstream loop detectors. The time of crash was determined as the time when the speed at the upstream detector station abruptly decreased due to the passage of a compression shockwave. This process resulted in the development of a crash database containing 299 crashes. For each of the crashes in the crash database two control factor conditions (time of day and geometric configuration) were also recorded. A detailed description of the calibration data is provided by Allaby [3].

4 Model calibration

Calibration of the crash prediction model consists of the following 6 steps:

1. Selection of Model Calibration Site.
2. Development of Crash Database.

3. Identification of External Control Factors.
4. Calculation of Crash Precursors.
5. Categorization (Selection of Boundary Values).
6. Log-Linear Analysis.

This paper focuses on the impact of the fifth step on the model fit and on safety estimates obtained from the model. In particular, we compare two categorization methods, namely an ad hoc method, and a near-optimal method. The ad hoc method was introduced by Lee et al. [11, 12] and used by Allaby [3].

Model categorization, the fifth step in the safety model calibration process, is one of the most important steps. The control factors are already categorical (e.g. merge/diverge geometry versus straight sections). However, the precursors are continuous values and must be categorized. Categorization requires two decisions, namely (1) the number of categories into which the precursors will be discretized and (2) specification of the precursor values that defines the boundary between two categories. If three categories are chosen, then two boundary values must be specified.

After all precursor data were transformed into categorical references and values of exposure were calculated, log-linear analysis (Step 6) was performed using SPSS Version 13.0 [16] to calibrate the crash prediction model. This procedure analyzes the frequency of samples in each cell of the contingency table to yield maximum likelihood estimates of the expected frequency of crashes under each possible condition. The analysis used an iterative fitting process until the difference between the current and previous estimates converged to 0.001. Because crashes are considered to be random events, the crash frequency was assumed to follow a Poisson distribution in the fitting process.

The performance of each categorization case was measured in terms of 1) overall model fit; 2) the statistical significance of individual coefficients; and 3) the consistency of coefficients with the order of levels of precursors (i.e. it is expected that "high" levels of precursors contribute more to crash risk than "low" level precursors). The overall model fit was measured by a log-likelihood ratio $\chi 2$ test. This test measures the differences between the observed crash frequencies and expected crash frequencies for any combination of crash precursor categories and control factors. A low $\chi 2$ and a high p-value indicate that the distribution of the expected crash frequencies is not significantly different from the distribution of the observed crash frequencies, meaning that the model fits well.

4.1 Ad hoc categorization method

This heuristic method was initially introduced by Lee et al. [11, 12] and was improved by Allaby [3] to calibrate a crash prediction model. He selected a number of categories for each of the three pre-cursors and then selected boundary values for each precursor. The number of categories is practically limited to between 2 and 4 as the total number of cells in the contingency table must be less than the number of crashes in the database. However, any set of boundary values is possible. Allaby constrained the number of options by assuming a proportion of the observed values would fall into each precursor category. For example, if 4 categories are assumed for precursor CVS, then one set of boundary values can be determined by assuming 20% of CVS observations in the first category, 30% in the second, 30% in the third, and 20% in the last category. Allaby evaluated a relatively small number (approximately 30) of categorizations and selected the one which provided the best model fit.

4.2 Near-optimal categorization method

One criticism of the method adopted by Allaby [3] is that there is no certainty that the selected model is the best. For example, a better model may have been obtained if some other categorization had been attempted. To address this concern, we developed software to automate the model calibration process (Steps 5 and 6) permitting the evaluation of a very large number (almost 5,000) of categorizations. The statistical fit of the model for each categorization was maintained in a separate database. Model suitability criteria, similar to those used by Allaby, were automatically applied to the database to identify the near-optimal model categorization.

4.3 Categorization results

Both methods were applied to the same set of field data to obtain a calibrated crash prediction model. The model characteristics are provided in Table 1. For the ad hoc method, the log-linear analysis resulted in a p-value close to 1.0 and a chi-squared likelihood ratio of 112.18 with 180 degrees of freedom. For the near-optimal method, however, the log-linear analysis resulted in a p-value of 1.0 and a chi-squared likelihood ratio of 49.47 with degrees of freedom of 98. The parameter estimates and statistical significance resulting from the two methods are provided in Table 2. The constant term in the model provides the crash frequency for base case factors for given values of exposure. Parameters for which estimates are negative indicate a declining contribution to crash risk, whereas positive estimates indicate an increasing contribution to crash potential.

5 Analysis of results

The two calibrated crash prediction models described in the previous section were used to estimate the safety impacts of a candidate variable speed limit control strategy. The focus on this application is the degree of consistency of the safety

Table 1 Comparing ad hoc and near-optimal methods

Method	Crash Precursor	Categorization[a]	Boundary Values (B1, B2, B3)
Ad Hoc	CVS	20/30/30/20	(0.062,0.089,0.139)
	Q	20/30/30/20	(−9.19,0.09,8.77)
	COVV	40/40/20	(1.49,3.44,–)
Near-Optimal	CVS	37/33/30	(0.079,0.126,–)
	Q	43/37/20	(−1.433,10.05,–)
	COVV	23/34/43	(0.815,1.833,–)

[a] Numbers in this column represent the percentage share of each category. For instance 40/40/20 means 40% of the observations for the precursor fall into the first category, 40% fall into in the second category and 20% fall into the third category

impacts obtained from the two calibrated safety models rather than on the variable speed limit sign system itself. This is not to investigate which model has a better fit, rather to evaluate sensitivity of model calibration on the final safety benefits of a candidate variable speed limit system.

5.1 Variable speed limit sign (VSLS) systems

Variable Speed Limit Sign (VSLS) systems consist of dynamic message signs (DMS) deployed along a roadway and connected via a communication system to a traffic management centre. The VSLS are used to display a regulatory or advisory speed limit. Unlike typical static speed signs, the VSLS system enables transportation system managers to dynamically post a speed limit that is appropriate for current traffic, weather, or other conditions. VSLS are thought to improve safety and reduce driver stress while improving traffic flow and travel times. However, few evaluation studies have been conducted to establish the safety impacts associated with VSLS. Allaby et al. [4] combined a crash prediction model (i.e. ad hoc categorization) with a micro-level traffic simulation model (PARAMICS) to estimate the safety and system delay impacts of a candidate VSLS system.

An 8 km section of the eastbound Queen Elizabeth Way (QEW) located near Toronto, Canada was selected as the test network. The QEW services a large volume of commuter traffic in the morning and evening peak periods, resulting in heavy congestion and a high frequency of crashes. The study segment features a posted speed limit of 100 km/hr, has three mainline lanes, contains four interchanges, and experiences a directional AADT of about 70,000 vehicles. The section is instrumented with dual loop detector stations in each mainline lane spaced at approximately 600 m and single loop stations on entrance and exit ramps. Every 20 s, speed, volume, and occupancy are recorded for all mainline stations, whereas volume is recorded for all ramp stations. During the morning peak period this freeway section experiences high level of recurrent congestion. This congestion is mainly caused by a bottleneck created at the most downstream interchange. At this location, a high volume of traffic (~1000 veh/hr) entering the already congested mainline results in reduced freeway speeds, queues, and an upstream moving shockwave that penetrates much of the section. Freeway speeds through the bottleneck during this period typically range from 30 km to 50 km, but at times traffic is observed to be at a standstill.

Table 2 Parameter estimates for heuristic ad hoc and near-optimal methods

Parameter	Category Level	Ad hoc		Near-optimal	
		Estimate	Z-Value	Estimate	Z-Value
θ	N/A	1.518	9.78	1.929	4.81
λ_{COVV}	[1] Low	−1.300	−5.32	−2.132	−3.13
	[2] Intermediate	−0.884	−3.70	−2.107	−5.20
	[3] High	0	–	0	–
λ_{CVS}	[1] Low	−0.914	−5.17	−1.577	−6.88
	[2] Intermediate A	−1.735	−8.11	−1.203	−7.40
	[3] Intermediate B	−1.496	−7.32	NA	NA
	[4] High	0	–	0	–
λ_Q	[1] High Acceleration	−0.875	−4.89	−2.452	−4.14
	[2] Moderate Acceleration	−1.738	−8.12	−2.107	−4.74
	[3] Moderate Deceleration	−1.508	−7.44	NA	NA
	[4] High Deceleration	0	–	0	–
λ_R	Straight	−0.530	−4.40	−0.618	−5.03
	Merge Diverge	0	–	0	–
λ_P	Off-Peak	−1.254	−8.16	−1.544	−6.41
	Peak	0	–	0	–
β	N/A	0.084	7.22	0.049	3.78

The simulation model was calibrated to observed network conditions for the morning peak period. Simulation parameters were adjusted until the speed profiles adequately matched the observed profiles. The VSLS system infrastructure was represented within PARAMICS by 13 variable speed limit signs placed throughout the network. Each VSLS was placed next to a loop detector, spaced at approximately 500 m to 600 m. Since PARAMICS assigns speed limits by link, the mainline was coded as a series of links corresponding to each detector-VSLS pair. Each link/detector/VSLS set acted as its own entity—the detector gathered information about traffic conditions, the appropriate "condition based" speed was assigned to the link, and the VSLS displayed the current speed limit for the benefit of the user/observer. Based on traffic data received every 20 s from a loop detector, a control algorithm determined the appropriate speed limit to be displayed at the respective VSLS. A candidate VSLS system control strategy was developed and is described elsewhere [4].

The VSLS impact analyses were performed on three traffic scenarios of varying levels of congestion: heavy, moderate, and light. These scenarios were termed Peak, Near-Peak, and Off-Peak, respectively. The validated simulation model from the observed morning peak period conditions represented the Peak traffic scenario. The Near-Peak and Off-Peak scenarios were represented by approximately 90% and 75%, respectively, of the peak volumes. These scenarios were not calibrated for existing conditions as their purpose was to investigate and understand the varying reaction of the VSLS system to changes in congestion, rather than to replicate real traffic conditions. The VSLS impact was quantified in terms of the relative changes in safety (crash risk) and vehicle travel times before and after the implementation of the VSLS control strategy.

5.2 Estimates of safety impacts

In this study, the safety impact of VSLS was measured by calculating the relative change in crash risk from the non-VSLS case to the VSLS case. Ten simulation runs were performed for the non-VSLS case and ten for the VSLS case. The same set of ten seed values was used for the VSLS and non-VSLS runs. For each simulation run, at each station, a value of crash potential (CP) was calculated from crash precursor values on 20-second intervals. Then, average values of station crash potential (SCP) were obtained for each run over the simulation period using Eq. 2.

$$SCP_i = \frac{1}{n}\sum_{j=1}^{n} CP_{ij} \tag{2}$$

Where,

SCP_i: Station Crash Potential for Station i (crashes/million veh-km)

CP_{ij}: Crash Potential for Station i at 20-second interval j (crashes/million veh-km)

n: Number of 20-second intervals in simulation period (720 for 4-hour period)

Since the non-VSLS and VSLS cases differed only by the introduction of the VSLS system, the SCP values could be paired by simulation run. A paired 2-tailed student *t*-test was used to test for the significance of the change in SCP (or VSLS impact) at the 95% level of confidence. If the difference was found to be significant, the relative safety benefit (RSB) was calculated using Eq. 3. A positive relative safety benefit represented a decrease in crash potential.

$$RSB_i = \left(\frac{ASCP_i(non-VSLS) - ASCP_i(VSLS)}{ASCP_i(non-VSLS)}\right) \times 100 \tag{3}$$

Where,

RSB_i: Relative Safety Benefit at Station i (%)

$ASCP_i$: Average Station Crash Potential (average of SCP over x simulation runs) at Station i (crashes/million veh-km).

For example, the average relative safety benefit associated with the VSLS as estimated using near-optimal categorization method for the Near-Peak scenario is 17% (Table 3). However, in this paper we are specifically interested in a comparison of the average relative safety benefit obtained from the two categorization methods (Table 4).

For the Peak scenario, the use of near-optimal categorization method suggested an average relative safety benefit (i.e. improvement in safety) of 44.3%. The use of ad hoc categorization method, with all other aspects of the analysis unchanged, estimates an average relative safety benefit of 40.1%. These estimates are relatively similar (with the absolute difference being only approximately 10.1% of the average of the relative safety benefits associated with different methods).

For the Near-Peak scenario, the average relative safety benefits obtained from the two categorization methods are again similar. The ad hoc method results in an estimate of 19.9% improvement in safety and near-optimal method results in an improvement of 17.4%. The absolute difference between these two estimates is approximately 13% of the average of the estimates provided by the two categorization methods. However, for this scenario near-optimal method provides a lower estimate of the safety benefits than the other method. In the Peak scenario, however, near-optimal categorization method provides a higher estimate of the safety benefits.

For the Off-Peak scenario, the safety impacts obtained from the two categorization methods differ substantially. The ad hoc method suggests a 10.8% decrease in safety, while near-optimal method suggests a decrease of almost 54%. Examination of the results from the near-optimal method

Table 3 Near-peak scenario in near-optimal method

Station	ACP		Impact	Significant at 95%?
	Non-VSLS	VSLS		
40	1.014	2.113	−108%	Y
50	0.218	0.226	−4%	N
60	0.208	0.062	70%	N
70	0.455	0.302	34%	Y
80	0.207	0.133	36%	N
90	0.52	0.429	18%	N
100	0.839	0.663	21%	N
110	0.816	0.643	21%	Y
120	0.573	0.314	45%	Y
130	0.653	0.31	53%	Y
140	0.332	0.131	61%	Y

for the Off-Peak scenario revealed that the average relative safety dis-benefit was computed from only 3 of the 11 stations, as only these three stations had statistically significant results at the 95% level. When more of the stations are included in the calculation (i.e. using a 90% confidence limit) then the average relative safety impact changes to a 28% decrease in safety.

Despite the differences that are observed between the results from the two crash model calibration methods, the two methods do provide consistent estimates of the sign of the safety impact (i.e. increase in safety versus decrease in safety). Furthermore, for the Peak and Near-Peak scenarios, the magnitude of the safety impact estimated by the two calibration methods differs by less than approximately 13%. Given the level of uncertainty associated with other aspects of these types of safety impact studies (e.g. simulation calibration, driver behavior assumptions, etc.) this level of consistency is likely adequate for decision making. Furthermore, the demonstrated level of consistency suggests that the safety model approach demonstrated in this paper is robust in that the final conclusions regarding the expected safety impacts of the candidate VSLS system are not highly sensitive to the degree of optimality of the crash prediction model.

Another interesting insight from the comparison between two calibration methods in Table 2 is the difference between the estimated coefficients. Since the categorization method

Table 4 Comparison of estimated safety impacts

Scenario	Average Safety Benefit		Change
	Ad hoc	Near-optimal	
Peak	40.10%	44.30%	10.10%
Near-Peak	19.90%	17.40%	−13.00%
Off-Peak	−10.80%	−53.90%	133.30%

in the hoc and near-optimal procedures are different, coefficients of the three traffic-related variables may not be compared directly. However, different values for λ_R, λ_P, and β in Table 2 are worth noting. The coefficient of the dummy variable for a straight section is −0.532 in the ad hoc method, while it is −0.618 in the near-optimal method. This coefficient is the increase in the value of the natural log of crash risk (F over EXP^β) when evaluated for a straight versus merge/diverge section. The coefficient of the dummy variable for the off-peak period is −1.254 and −1.544, in the ad hoc and near-optimal methods, respectively. This coefficient has a similar interpretation as well. β is the parameter to which the exposure value is raised. It is around 0.04 in the near-optimal method and 0.08 in the ad hoc method. β is the elasticity of the expected number of crashes with respect to the exposure (σF/F over σEXP/EXP). In other words, the ad hoc model shows a 0.04% increase in the expected number of crashes with 1% growth in the exposure value. This elasticity seems to be sensitive to the calibration accuracy and its volatility should be carefully considered when policies are assessed based on the crash risk models of this type.

6 Conclusions

In this paper we have examined the sensitivity of the estimated safety performance of a variable speed limit system to the method used to calibrate the log-linear crash prediction model.

The log-linear crash prediction model contains precursor variables that reflect various real-time measures of traffic conditions. These precursors are continuous variables that are represented with the log-linear model as categorical variables. Consequently, calibration of the log-linear crash prediction model requires the specification of the number of categories and values defining the boundaries between categories. Different approaches to selecting the number of categories and their associated boundary values can result in different model structures with different coefficient values. Unfortunately, the importance of these differences, in terms of the estimated safety performance of the intervention being examined (a candidate VSLS in this paper), cannot be determined solely on the basis of a comparison of the crash prediction model coefficients. Rather it is necessary to apply the crash prediction model to estimate the safety impacts of a particular intervention (in this case a candidate VSLS system). Comparing the safety impacts from the different candidate safety models provides an indication of the importance of differences in the safety model that arises from the use of different calibration techniques.

The results obtained from the combination of the crash prediction model with a micro-level simulation model to assess the safety benefits of a candidate VSLS system show that the estimated safety benefits generally are not highly sensitive to the optimality of the safety model calibration. Most importantly, for all scenarios examined, the sign of the safety impact remained consistent for both calibration methods. The calibration of the log-linear crash prediction model requires the selection of the number of categories and the precursor values that define the boundaries between categories. Given the lack of an optimal mean of determining these, the limited sensitivity of the safety performance increases the level of confidence that the safety benefits estimated from the combination of the simulation and crash prediction models are not a spurious outcome of the method used to calibrate the crash prediction model.

If the safety impacts are not highly sensitive to the optimality of the model calibration to data obtained from a given freeway site, then it may also be true that the safety benefits obtained from a model calibrated to data from a particular freeway site may be relatively similar to the benefits obtained from a model calibrated to data from some other freeway site. Though this study has not directly tested this hypothesis, the results of this study do provide evidence that the safety impacts are not highly sensitive to the safety model calibration. This is encouraging, as it is desirable to be able to apply a calibrated safety model to various freeway sections.

Acknowledgements This work received financial support from the Natural Sciences and Engineering Research Council of Canada. Also, we would like to thank Dr. Chris Lee and Peter Allaby for providing data.

References

1. Abdel-Aty M, Dilmore J, Dhindsa A (2006) Evaluation of variable speed limits for real-time freeway safety improvement. Accid Anal Prev 38(2):335–345
2. Abdel-Aty M, Pande A (2005) Identifying crash propensity using specific traffic speed conditions. J Safety Res 36(1):97–108
3. Allaby P (2006) An evaluation of the safety and operational impacts of a candidate variable speed limit control strategy on an urban freeway. MSc thesis, University of Waterloo, Waterloo, Ontario Canada
4. Allaby P, Hellinga B, Bullock M (2007) Variable speed limits: safety and operational impacts of a candidate control strategy for freeway applications. IEEE Trans Intell Transp Syst 8(4): 671–680
5. Feng C (2001) Synthesis of studies on speed and safety. Presented at the 80th Annual Meeting of Transportation Research Board, Washington, D.C., 2001
6. Golob TF, Recker WW, Alvarez VM (2004) Freeway safety as a function of traffic flow. Accid Anal Prev 36(6):933–946
7. Hughes R, Council F (1999) On Establishing the Relationship(s) between FreewaySafety and Peak Period Operations: Performance Measurement and Methodological Considerations. Presented at the 78th Annual Meeting of Transportation Research Board, Washington, D.C., 1999
8. Kweon Y, Kockelman K (2005) Safety effects of speed limit changes: use of panel models, including speed, use, and design variables. Transp Res Rec 1908:148–158
9. Lee C, Abdel-Aty M (2008) Identifying traffic flow parameters affecting crash occurrence on freeway ramps using two-level nested logit model. Transp Res Rec 2083:145–152
10. Lee C, Hellinga B, Saccomanno F (2004) Assessing safety benefits of variable speed limits. Transp Res Rec 1897:183–190
11. Lee C, Hellinga B, Saccomanno F (2003) Real-time crash prediction model for the application to crash prevention in freeway traffic. Transp Res Rec 1840:67–78
12. Lee C, Hellinga B, Saccomanno F (2003) Proactive freeway crash prevention using real-time traffic control. Can J Civ Eng 30 (6):1034–1041
13. Lee C, Saccomanno F, Hellinga B (2002) Analysis of crash precursors on instrumented freeways. Transp Res Rec 1784:1–8
14. Oh C, Oh J, Ritchie S, Chang M (2001) Real-Time Estimation of Freeway Accident Likelihood. Presented at the 80th Annual Meeting of Transportation Research Board, Washington, D.C., 2001
15. Persaud B, Lord D, Palmisano J (2002) Calibration and transferability of accident prediction models for urban intersections. Transp Res Rec 1784:57–64
16. SPSS Inc. (2004) SPSS for Windows Release 13.0. Chicago, Illinois