Developing Updatable Crash Prediction Model for Network Screening

Case Study of Czech Two-Lane Rural Road Segments

Jiří Ambros, Veronika Valentová, and Jiří Sedoník

This paper focuses on application of crash prediction models in network screening. The two main questions were (*a*) What variables should be involved in the model? and (*b*) What length should the modeled period be? Answers to these questions should provide guidelines for developing an updatable crash prediction model (i.e., a model that is both reliable and simple so that its updating for periodical network screening is not highly demanding). Data on approximately 1,000 km (600 mi) of a two-lane rural-road network from South Moravia, Czech Republic, were used. On the basis of 8 years of annual crash frequencies, together with exposure and geometrical variables, several variants of prediction models were developed. To study the quality of the models, a series of consistency tests was applied relative to comparison of the models themselves as well as to their diagnostic performance. As a result, simple crash prediction models (that included traffic volume, segment length, and curvature change rate) were found sufficient for network screening. If one supposes that length and curvature are unlikely to change often, only traffic volume data need to be periodically updated. Consistency analyses indicate that this period should be 4 years. Under these conditions, models are being applied in the studied region. Further planned activities include extensions to intersections and also to other Czech regions.

Network screening [also known as "identification of hazardous road locations" and "hot-spot identification" (HSID)] is defined as "the process by which a road network is screened to identify sites that require safety investigation" (*1*). As the first step in a process of road network safety management, it is one of the most frequent tasks that road agencies conduct (*1–3*). According to recommended practices, network screening employs crash prediction models [also known as "safety performance functions" (SPFs)] and the empirical Bayes (EB) approach (*4, 5*); in the end, a list is produced enabling the ranking of the locations from most likely to least likely to realize a reduction in crash frequency with implementation of countermeasures (*3*).

Although the mentioned process has been known for several decades, it involves several decisions that have rarely been described in the literature in sufficient detail. The main questions addressed in this paper are the following:

1. What variables should be involved in the model? The models are of two basic types. One model is multivariate. The other model is simple (*1, 6*). The difference between these types is in the independent variables (covariates): while simple models involve risk exposure only (i.e., traffic volume and segment length), multivariate models also use further variables, usually geometric characteristics. Although increasing model complexity may provide better insight into modeled safety performance, additional predictors have often been nonbeneficial. For example, studies of UK single carriageways (undivided two-lane roads) from the 1990s found that only a small fraction of explanatory variables significantly improved the model fit (*7*); a recent follow-up study confirmed that the best-fitting models did not include any of geometric features (*8*). Similar conclusions were reached by Finnish researchers in the 1990s (*9*) and dictated the use of simple models since then (*10*). In a recent U.S. study, only a few variables were found to explain most of the variation in the crash data (*11*). In addition, a previous Czech study showed that simple models may be sufficient for network screening (however only one period, with an arbitrarily set length, was used) (*12*).
2. What length should the modeled period be? Between 1 and 5 years is usually recommended for network screening, with 3 years being the most frequent (*4*). Three years is a compromise between the needs for quick detection and for accumulating sufficient crash numbers to permit analysis (*13*). In contrast, longer periods may cause problems in the stability of conditions, which may no longer reflect the current traffic situation (*14*). Probably because of these issues, no specific guidelines for the choice of period length are usually provided; one exception was the simulation study of Cheng and Washington, which concluded that network screening accuracy shows little gain when a period longer than 6 years is used (*15*).

These two questions are interrelated and critical in the process of developing an updatable crash prediction model. Such an updatable model should be sufficiently reliable (i.e., describing safety performance of a modeled data set) while also sufficiently simple and parsimonious so that its updating (in a specific period to be estimated) is not highly demanding.

This case study presents the development of an updatable crash prediction model for network screening of almost 1,000 km (600 mi) of two-lane rural road segments in the region of South Moravia, Czech Republic. It is meant as the first application of such study under Czech conditions, aiming to prove the approach's feasibility as well as its practical applicability for the needs of a regional road agency. On the basis of 8 years of annual crash frequencies, together with exposure and geometrical variables (described in the section on data preparation), several variants of prediction models will be developed. To study the quality of the models, a series of tests are applied (with later sections describing the comparison of the models and

Transport Research Centre, v.v.i., Líšeňská 33a, 636 00 Brno, Czech Republic. Corresponding author: J. Ambros, jiri.ambros@cdv.cz.

Transportation Research Record: Journal of the Transportation Research Board, No. 2583, Transportation Research Board, Washington, D.C., 2016, pp. 1–7.
DOI: 10.3141/2583-01

then model performance). The discussion and conclusions answer the two questions posed here.

DATA PREPARATION

Segmentation

The studied network consists of road sections (excluding intersections) of undivided two-lane paved rural roads. Their length [995 km (618 mi)] was divided into homogeneous segments with respect to following variables: annual average daily traffic (AADT), speed limit reduction, road category, number of lanes, and paved shoulder. A change of any of these variables marked the end of a segment and the beginning of another one. Figure 1 illustrates the principle of this segmentation, following the work of Cafiso et al. (*16*).

To obtain segment lengths that would be practical for follow-up safety inspections, segments longer than 500 m (1,640 ft) were divided into 250-m (820-ft) parts. After this process, most of the segments (78%) were 250 m long; 17% were longer, and 5% were shorter. The number of segments was 3,764.

Variables

The segments were assigned specific values of response variable (crash frequency) and explanatory variables (exposure data, road and traffic characteristics, context and environment variables) that represent safety-related features. Reported crash frequency data for the period from 2007 to 2014 were obtained from the Czech traffic police. They included all injury crashes (i.e., those with slight, severe, or fatal personal consequences). After exclusion of intersection-related crashes, 2,219 crashes remained, with a frequency between 0 and 18 within a segment. Further explanatory variables were added:

- Data on AADT and on percentage of heavy-goods vehicles were used to represent crash exposure. These data were acquired from the Czech Road and Motorway Directorate and were based on the results of the national traffic census of 2010.
- The segment length described earlier was used as the second risk exposure variable.
- Road and traffic characteristics data were obtained from a Czech Road and Motorway Directorate database that reflected conditions in July 2010. Density of intersections with minor rural roads and density of roadside facilities were computed as frequencies divided by segment lengths. Some variables were not well represented across their range of values; for example, for the number of lanes, 95% of cases were two-lane segments, and for speed limit reduction, 99% of cases were without reduction [i.e., 90-km/h (56-mph) speed limit applied]. These two variables were thus removed.
- Another variable was related to the quality of the road pavement, which is defined according to collected data about cracks and potholes in five classes from 1 (excellent) to 5 (wrecking). The data used were collected in 2011 by the Czech company PavEx Consulting and listed sections classified according to the worst quality level present (*17*); however, only the list containing sections with Levels 4 (unsatisfactory) and 5 (wrecking) was available to the authors. Therefore, for further analyses, each segment was assigned the ratio of length with the pavement quality of Level 4 or 5; the value was

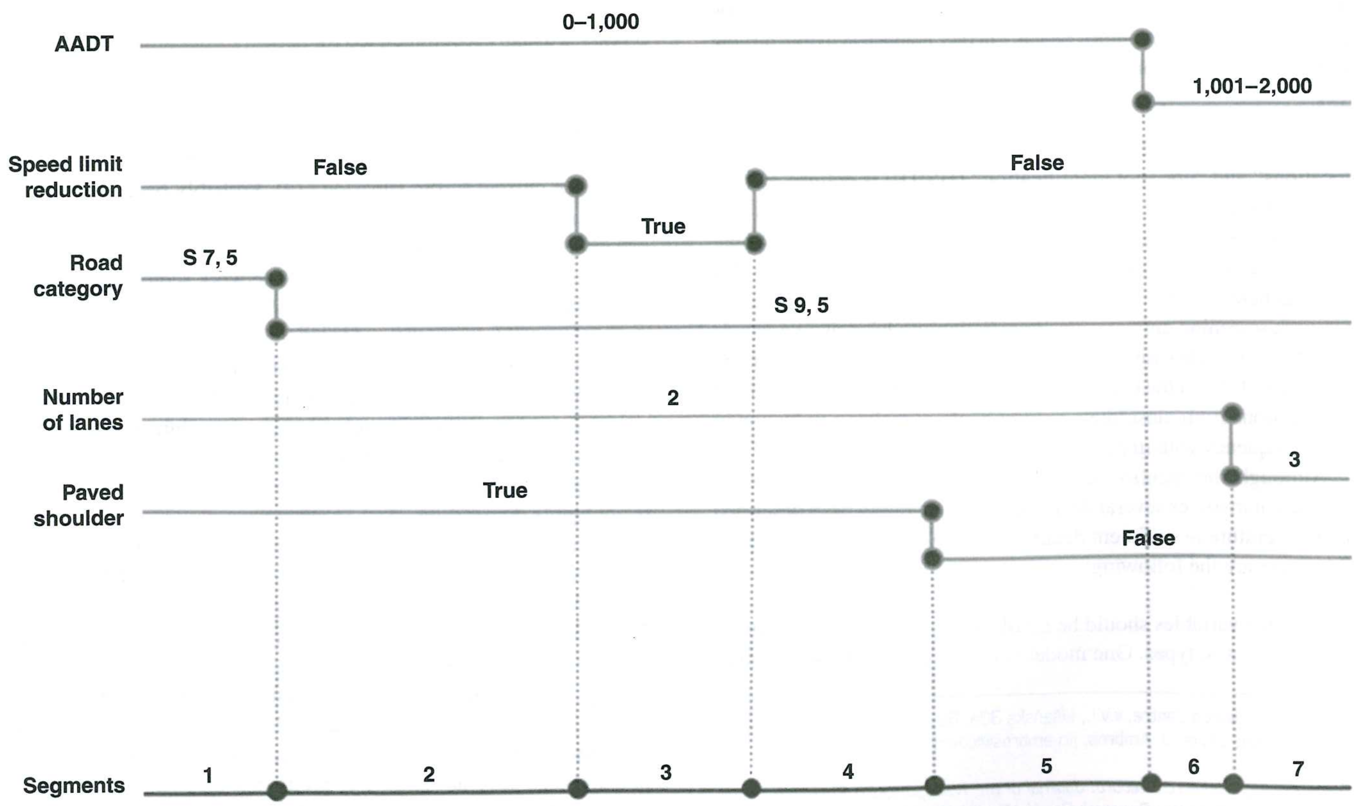

FIGURE 1 Principle of division into homogeneous segments.

between 0 and 1 (0 being quality better than Level 4 or 5; 1 being total length of these quality levels).

- Some context and environmental variables were also acquired from the Czech Road and Motorway Directorate database. Data for other ones were also collected; these were average curvature change rate (CCR), a traditional alignment consistency indicator, and forest environment, which is linked to wet surfaces or game crashes (*18*). CCR was computed as sums of angles between vertices divided by the sum of their lengths (*19*). Information about continuous forest around the road segment was collected manually from online maps that use data from the Czech Environmental Information Agency.

All variables and their characteristics are summarized in Table 1.

Modeling

Before the modeling, intercorrelation of variables was investigated. The highest significant correlation was found between road category and paved shoulder (almost .67). Because the difference in lane width for each road category was minimal (Category 7.5 is narrower than Categories 9.5 and 11.5 by 0.25 m), the connection may therefore be caused by the relation between road category and paved shoulder, which is governed by Czech standards: wider roads tend to be equipped with a paved shoulder. Because the amount of intercorrelation was considerable and given that road category is a derived variable, it was removed from the data set.

Negative binomial regression modeling was used to develop the models [details are available elsewhere, e.g., Hauer (*20*)], with AADT modeled in function form $AADT^{\beta_1} \cdot \exp(\beta_2 \cdot AADT)$ [following Hauer et al. (*21*)]. The required prediction model form P_i was then as follows:

$$P_i = \exp(\beta_0) \cdot AADT_i^{\beta_1} \cdot \exp\left(\sum_{i=2}^{n} \beta_i x_i\right) \quad (1)$$

where β_i are coefficients to be estimated in modeling and x_i are explanatory variables. For estimation, the procedure GENLIN from the Statistical Package for the Social Sciences was used.

COMPARISON OF MODELS

As indicated earlier, the number of explanatory variables determines whether the model type is simple or multivariate. To study the suitable type within the available time frame (8 years), the following investigation scheme was planned:

- A 3-year period is most often used in the literature; thus, it was chosen as a minimum period. An 8-year time frame includes six variants of 3-year periods.
- Five years are usually seen as the maximum; to check performance above this level, a 6-year period was chosen as the maximum. An 8-year time frame has three variants of 6-year periods.
- Periods of 4 years (five variants) and 5 years (four variants) were chosen in a similar way, by using overlapping variants. For an illustration, see Figure 2.

Crash prediction models were developed for all 18 variants, in a backward elimination manner, with only the variables with at least 95% statistical significance being kept in the model.

Variants of 5-year models could not be built; the error (caused by singularity or nonpositive definiteness of the Hessian matrix) is a "frustrating but common occurrence in applied quantitative research" (*22*). Therefore, only model variants for 3-, 4-, and 6-year periods were developed. To illustrate the effect of variable choice on the least common denominator, Figure 3 shows the resulting significance of explanatory variables for 3-year variants (as described in Figure 2). Gray cells in Figure 3 indicate significance, and white ones indicate lack of it. For all the variants, three variables (in bold) were always statistically significant: the logarithm of AADT, the length, and CCR.

Two model types are distinguished here:

- The ones to be called simple, including the three mentioned variables (log AADT, length, CCR) and
- The ones to be called multivariate, which also include other variables.

To compare the performance (goodness of fit) of simple and multivariate models, various indicators may be used. For example, Oh et al. used five measures to assess external validity (Pearson correlation coefficient between observed and predicted crash frequencies, mean

TABLE 1 Overview of Data with Description and Descriptive Statistics of Variables

Data Type	Variable	Description	Data Type and Unit[a]	Descriptive Statistics[b]
Crash data	R	8-year frequency of reported injury crashes	Count	0/18/0.59/1.21
Exposure data	AADT	Annual average daily traffic	Continuous (vehicles per day)	91/18,498/2,459.37/2,229.86
	HGV	Heavy goods vehicle percentage	Continuous	0.06/0.50/0.18/0.06
	LEN	Segment length	Continuous (m)	51.00/499.88/264.29/64.03
Road and traffic data	CAT	Road category	Binary (0 = 7.5 m; 1 = 9.5 or 11.5 m wide)	0: 3,156; 1: 608
	SH	Paved shoulder	Binary (0 = false; 1 = true)	0: 3,333; 1: 431
	PAV	Pavement quality	Continuous	0/1/0.49/0.49
Context and environment data	CCR	Average curvature change rate	Continuous (gon per km)	0.0/1,498.18/98.03/134.27
	INT	Density of intersections with minor rural roads	Continuous (number per km)	0.0/16.90/1.16/2.40
	FAC	Density of roadside facilities	Continuous (number per km)	0.00/52.00/2.58/5.76
	FOR	Forest environment	Binary (0 = false; 1 = true)	0: 2,977; 1: 787

[a]1 m = 3.3 ft; 1 km = 0.6 mi.
[b]Minimum/maximum/mean/standard deviation/or frequencies.

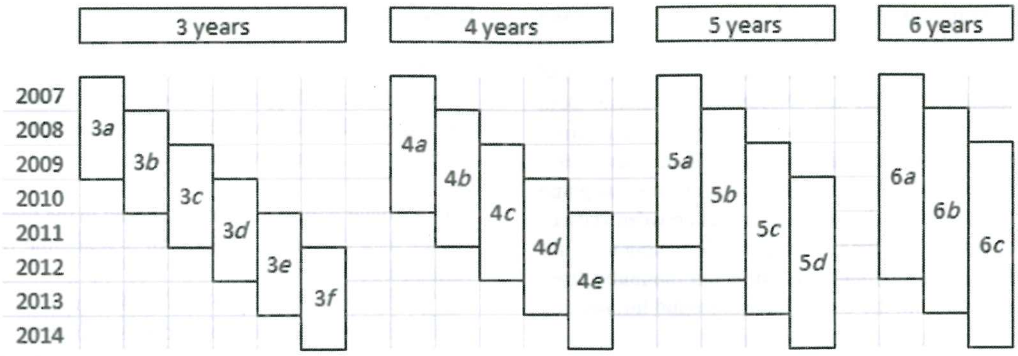

FIGURE 2 Scheme of investigated variants of periods.

prediction bias, mean absolute deviation, mean squared prediction error, and mean squared error), while noting that they all should be considered jointly (*23*). For the sake of brevity, a single indicator was used here: the proportion of systematic variation in the original crash data set explained by the model (% SV) (*24–26*). The indicator % SV was computed for all variants of 3-, 4-, and 6-year models. The values were averaged for individual periods. The results are shown in Figure 4, in gray for multivariate models and in white for simple models.

Figure 4 makes clear that results are all near 60% and relatively similar for both model types, with differences below 3%. These results show that the explanatory variables that are missing from the simple models provide only minor improvement of model performance.

COMPARISON OF MODEL PERFORMANCE

As stated at the outset, this study focused on network screening and was based on the EB method with crash prediction models. In principle, crash prediction models were used to obtain predicted crash frequency (P) for each segment (i). The EB estimate of expected crash frequency was then calculated by using predicted crash frequency, reported crash frequency, and a length-dependent overdispersion parameter (*5, 27*). Finally, potential for safety improvement (PSI) was obtained as a difference between predicted crash frequency and the EB estimate (*28*).

$$\text{EB}_i = w_i \cdot P_i + (1 - w_i) \cdot R_i \tag{2}$$

$$w_i = \frac{k_i}{k_i + P_i} \tag{3}$$

$$k_i = k \cdot L_i \tag{4}$$

$$\text{PSI}_i = \text{EB}_i - P_i \tag{5}$$

where

w_i = weight,
R_i = reported crash frequency,
k_i = overdispersion parameter, and
L_i = segment length.

Values of PSI were used for network screening. After these values were ranked in descending order, two lists of segment numbers were developed. Following previous studies, 1%, 3%, and 5% upper tails (top parts of the distribution) were further used to investigate differences (1% = 38 segments, 3% = 113 segments, 5% = 188 segments).

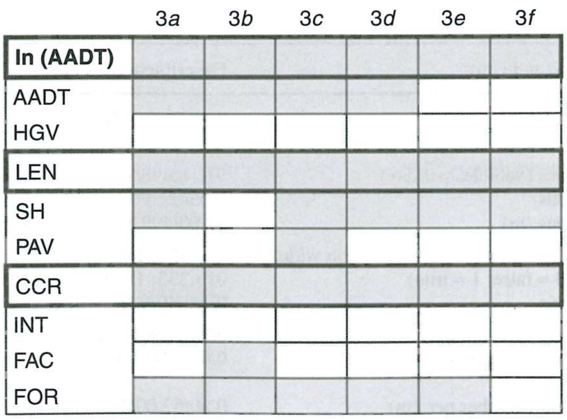

FIGURE 3 Overview of achieved statistical significance of explanatory variables in 3-year models (gray cells = significant, white cells = nonsignificant).

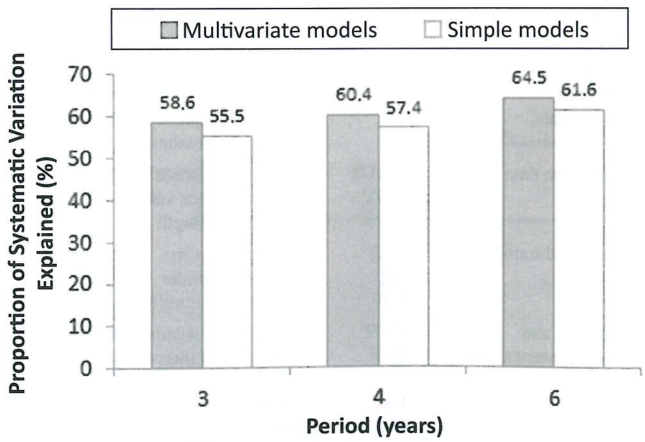

FIGURE 4 Comparison of average proportion of systematic variation explained for multivariate and simple models.

TABLE 2 Summary and Description of Applied Consistency Tests

Test	Test Premise	Test Criterion
Site consistency (Test 1)	A segment identified as risky during time period i should also be identified as risky in time period $i + 1$.	The method that identifies segments in time period i, with the highest crash frequency in a time period $i + 1$, is the most consistent method.
Method consistency (Test 2)	A list of segments that were identified as risky during time period i should be similar to the list of segments will be identified as risky in time period $i + 1$.	The method that identifies segments in time period i, which has the largest overlap with the list of segments identified in time period $i + 1$, is the most consistent method.
Total rank differences (Test 3)	The segments in the lists, produced in Test 2, should have similar rankings between time periods i and $i + 1$.	The method with the smallest sum of total rank differences between the lists from time periods i and $i + 1$ is the most consistent method.
Epidemiological diagnostic (Test 4)	The method should identify as many of the truly risky segments as possible (sensitivity), and as few of the truly nonrisky segments as possible (specificity).	The method with the greatest sum of sensitivity and specificity is the most consistent method.

The objective here was to assess the implications of the choice of different modeling periods on the results of network screening (i.e., ranked lists of segment identification numbers). Assessment may be done in terms of consistency. In the literature, various consistency tests have been used. For example, Miranda-Moreno et al. applied percentage deviation and Spearman correlation coefficient to compare the performance of two ranking criteria (*29*). Cheng and Washington, in their study cited in (*32*), used false positives, false negatives, and the effects of crash history duration to compare three HSID methods. To evaluate the diagnostic performance of five HSID techniques, Elvik used two epidemiological criteria (sensitivity and specificity) (*30*). Montella employed four consistency tests (site consistency, method consistency, total rank differences, total score) to assess the performance of seven HSID methods (*31*).

This literature review shows that consistency criteria have been used primarily for comparison of different screening methods. In the current study, that same principle was adapted to compare the effect of using screening that was based on models from different times. Therefore, the original meaning of "identification with different methods" is applied in the meaning of "identification in different periods" (all with three levels of upper tails). Four criteria were selected from the mentioned studies. Their descriptions and definitions are summarized in Table 2; further details may be found in Elvik (*30*), Montella (*31*), and Cheng and Washington (*32*). In Test 4, segments identified as hazardous in the maximum (8-year) period were considered positives.

The tests were applied on all 28 possible variation pairs of lists of segment numbers (15 pairs for 3-year models, 10 pairs for 4-year models, and three pairs for 6-year models), for three upper tails (1%, 3%, and 5%). Figures 5 through 8 show the test results, averaged for specific combinations of upper tail and period.

Test 2 (method consistency) showed relative similarity between the lists that were based on both the simple and the multivariate models: the agreement was 90% or above for all variants. For clarity, the following test results are reported for simple models only.

On the basis of the test outputs (Figures 5 through 8) and test criteria definitions in Table 2, the following results may be stated:

- In Test 1 (site consistency), optimal (maximum) values are reached with 4-year models (Figure 5).
- In Test 2 (method consistency), 6-year models provide optimal (maximum) values (Figure 6). However, in other studies that used this test to compare different HSID methods, values below 50% were usually reached as maximums [46.9% in Montella (*31*), 47.3% in Cheng and Washington (*32*), and 46.2% in Yu et al. (*33*), with all values being for 5% upper tails]. Should these values present an acceptable threshold, 4-year models would be sufficient because their values are all above 50% overlap.
- In Test 3 (total rank differences), the minimal value is the optimum. According to Figure 7, one may choose between 4- and 6-year periods.

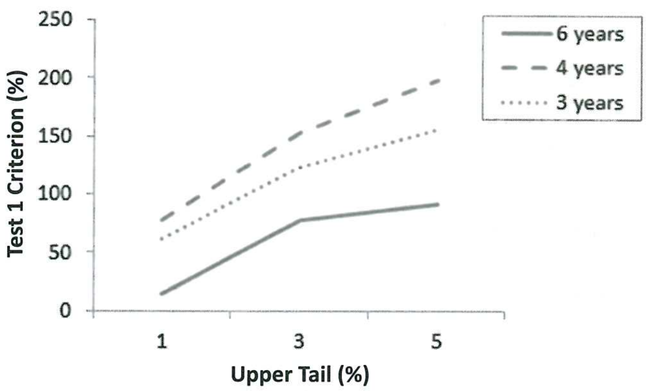

FIGURE 5 Results of Test 1 (site consistency) for different periods and upper tails.

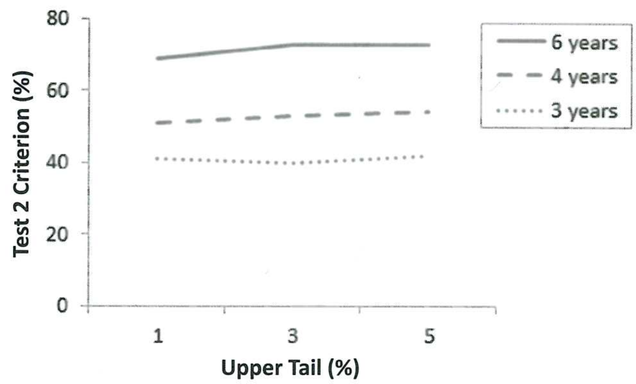

FIGURE 6 Results of Test 2 (method consistency) for different periods and upper tails.

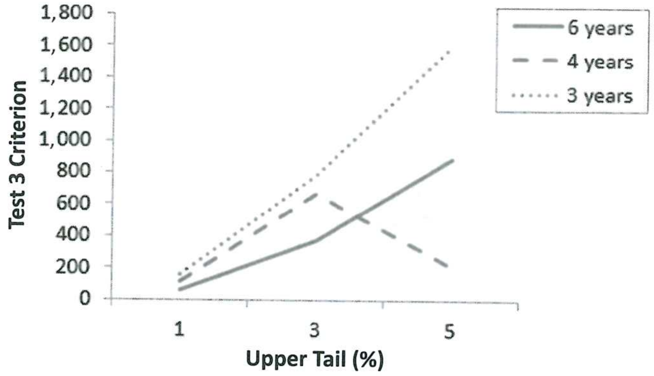

FIGURE 7 Results of Test 3 (total rank differences) for different periods and upper tails.

- In Test 4 (epidemiological diagnostics), optimal (maximum) values are reached with 6-year models. In addition, simple and multivariate models were compared; Figure 8 illustrates that the two model types perform most similarly with 4-year models.

DISCUSSION AND CONCLUSIONS

The case study focuses on application of crash prediction models in network screening (also known as "identification of hazardous road locations" and "hot-spot identification"). The two main questions were (*a*) What variables should be involved in the model? and (*b*) How long should the modeled time period be? Answers to these questions should provide guidelines to developing an updatable crash prediction model (i.e., a model that is both reliable and simple, so that its updating for periodical network screening is not highly demanding).

For choice of model type, simple and multivariate models were distinguished. In relation to the proportion of systematic variation explained, simple alternatives are quite similar to multivariate models (Figure 4). In Test 2 (method consistency), relative similarity between the lists that were based on simple and multivariate models was reached, and the agreement was 90% or above for all variants. In addition, the results of Test 4 (epidemiological diagnostics) illustrate that both model types perform relatively similarly, with the optimum at model variants from the 4-year period.

In relation to choice of period, several consistency tests were applied. Most of the results indicate that—in relation to site consistency, method consistency, and total rank differences—a 4-year period is acceptable for developing a crash prediction model. Unfortunately, with the exception of Test 2 (with results showing percentages of overlap between ranked lists), relating the test results to other studies proves difficult because of their several differences. For example, with Test 4, different authors considered different periods to represent the "true mean": while an 8-year period (as an available maximum) was used in the presented case study, a shorter period (3 years) was used by the others (*32*).

To sum up, the simple models include explanatory variables of traffic volume (AADT), segment length, and CCR. If one supposes that length and curvature are not likely to change often, only one variable (AADT) needs to be updated. A networkwide AADT update should follow the national traffic census, which takes place every 5 years; in the meantime, AADT growth factors may be used, in overlapping periods of 4 years (*34*). Currently, the period from 2011 through 2014 has been used; as soon as 2015 crash data become available in 2016, a new model for the period from 2012 through 2015 will be calibrated.

Under these conditions, an updatable crash prediction model is currently applied in the studied region. The authors hope the model to be a beneficial tool for road network safety management in the region of South Moravia. Further planned activities include extensions to intersections and to other Czech regions.

ACKNOWLEDGMENTS

The authors appreciate the help of Jan Novák and Robert Zůvala with computations on consistency tests. The study was conducted with the support of the Czech Ministry of Interior's research project Identification and Treatment of High Risk Road Spots and Sections (IDEKO) and the research infrastructure of the Transport R&D Centre.

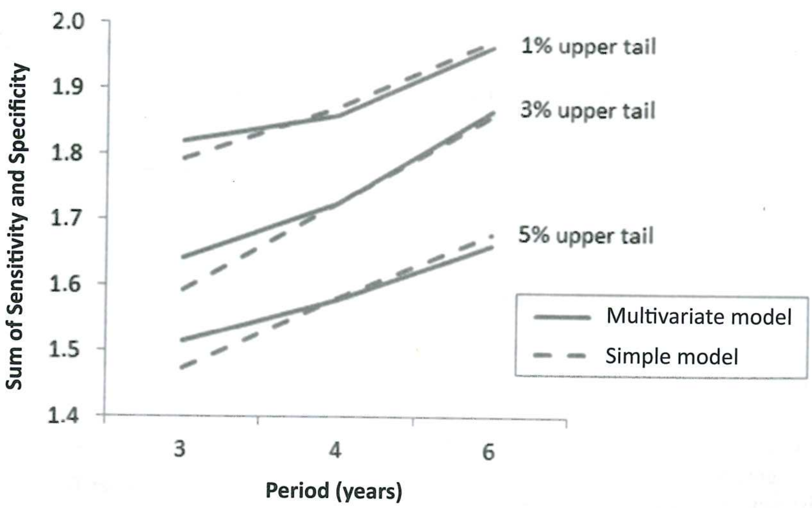

FIGURE 8 Results of Test 4 (epidemiological diagnostics) for different periods and upper tails.

REFERENCES

1. Persaud, B. N. *NCHRP Synthesis of Highway Practice 295: Statistical Methods in Highway Safety Analysis.* TRB, National Research Council, Washington, D.C., 2001.
2. 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.
3. *Highway Safety Manual.* AASHTO, Washington, D.C., 2010.
4. Elvik, R. A Survey of Operational Definitions of Hazardous Road Locations in Some European Countries. *Accident Analysis and Prevention,* Vol. 40, 2008, pp. 1830–1835.
5. Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith. Estimating Safety by the Empirical Bayes Method: A Tutorial. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1784,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 126–131.
6. Hauer, E. *Observational Before–After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety.* Pergamon Press, Oxford, United Kingdom, 1997.
7. Walmsley, D. A., I. Summersgill, and C. Binch. *Accidents on Modern Rural Single-Carriageway Trunk Roads.* Report 336. Transport Research Laboratory, Crowthorne, United Kingdom, 1998.
8. Wood, A. G., L. J. Mountain, R. D. Connors, M. J. Maher, and K. Ropkins. Updating Outdated Predictive Accident Models. *Accident Analysis and Prevention,* Vol. 55, 2013, pp. 54–66.
9. Peltola, H., R. Kulmala, and V.-P. Kallberg. Why Use a Complicated Accident Prediction Model When a Simple One Is Just As Good? Presented at 22nd PTRC Summer Annual Meeting, University of Warwick, United Kingdom, Sept. 12–16, 1994.
10. Peltola, H., R. Rajamäki, and J. Luoma. A Tool for Safety Evaluations of Road Improvements. *Accident Analysis and Prevention,* Vol. 60, 2013, pp. 277–288.
11. Saha, D., P. Alluri, and A. Gan. Prioritizing Highway Safety Manual's Crash Prediction Variables Using Boosted Regression Trees. *Accident Analysis and Prevention,* Vol. 79, 2015, pp. 133–144.
12. Ambros, J., V. Valentová, and Z. Janoška. Investigation of Difference Between Network Screening Results Based on Multivariate and Simple Crash Prediction Models. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.
13. Elvik, R. *Assessment and Applicability of Road Safety Management Evaluation Tools: Current Practice and State-of-the-Art in Europe.* Report 1113/2010. Institute of Transport Economics, Oslo, Norway, 2010.
14. 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.
15. Cheng, W., and S. Washington. Experimental Evaluation of Hotspot Identification Methods. *Accident Analysis and Prevention,* Vol. 37, 2005, pp. 870–881.
16. Cafiso, S., A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud. Development of Comprehensive Accident Models for Two-Lane Rural Highways Using Exposure, Geometry, Consistency and Context Variables. *Accident Analysis and Prevention,* Vol. 42, 2010, pp. 1072–1079.
17. Gregor, J., and L. Mališ. *Stav povrchu vozovek silnic II. a III. třídy v Jihomoravském kraji k 30.4.2011.* PavEx Consulting, s.r.o., Brno, Czech Republic, 2011.
18. Elvik, R., A. Høye, T. Vaa, and M. Sørensen. *The Handbook of Road Safety Measures,* 2nd ed. Emerald, Bingley, United Kingdom, 2009.
19. Dantas, A., M. Fowler, and G. Koorey. Effect of Road Network Bendiness on Traffic Crash Occurrence. Presented at IPENZ Transportation Group Technical Conference, Tauranga, New Zealand, Oct. 7–10, 2007.
20. Hauer, E. Statistical Road Safety Modeling. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 81–87.
21. Hauer, E., F. M. Council, and Y. Mohammedshah. Safety Models for Urban Four-Lane Undivided Road Segments. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1897,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 96–105.
22. Gill, J., and G. King. What to Do When Your Hessian Is Not Invertible: Alternatives to Model Respecification in Nonlinear Estimation. *Sociological Methods and Research,* Vol. 33, 2004, pp. 54–87.
23. Oh, J., C. Lyon, S. Washington, B. Persaud, and J. Bared. Validation of FHWA Crash Models for Rural Intersections: Lessons Learned. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1840,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 41–49.
24. Kulmala, R. *Safety at Rural Three- and Four-Arm Junctions: Development and Application of Accident Prediction Models.* Publication 233. VTT Technical Research Centre of Finland, Espoo, 1995.
25. Vogt, A., and J. G. Bared. *Accident Models for Two-Lane Rural Roads: Segments and Intersections.* FHWA-RD-98-133. FHWA, U.S. Department of Transportation, 1998.
26. Greibe, P. Accident Prediction Models for Urban Roads. *Accident Analysis and Prevention,* Vol. 35, 2003, pp. 273–285.
27. Hauer, E. Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation. *Accident Analysis and Prevention,* Vol. 33, 2001, pp. 799–808.
28. Persaud, B., C. Lyon, and T. Nguyen. Empirical Bayes Procedure for Ranking Sites for Safety Investigation by Potential for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1665,* TRB, National Research Council, Washington, D.C., 1999, pp. 7–12.
29. Miranda-Moreno, L. F., L. Fu, F. F. Saccomanno, and A. Labbe. Alternative Risk Models for Ranking Locations for Safety Improvement. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 1–8.
30. Elvik, R. Comparative Analysis of Techniques for Identifying Locations of Hazardous Roads. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2083,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 72–75.
31. Montella, A. A Comparative Analysis of Hotspot Identification Methods. *Accident Analysis and Prevention,* Vol. 42, 2010, pp. 571–581.
32. 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.
33. Yu, H., P. Liu, J. Chen, and H. Wang. Comparative Analysis of the Spatial Analysis Methods for Hotspot Identification. *Accident Analysis and Prevention,* Vol. 66, 2014, pp. 80–88.
34. Bartoš, L., A. Richtr, J. Martolos, and M. Hála. *Prognóza intenzit automobilové dopravy, II. vydání.* Technical guidelines No. 225. EDIP s.r.o., Plzeň, Czech Republic, 2012.

The Standing Committee on Highway Safety Performance peer-reviewed this paper.