Collision Prediction for Two Lane Rural Roads Using IHSDM: A Canadian Experience

Matthew Marleau, Eric Hildebrand
University of New Brunswick Transportation Group

Abstract

A study at the University of New Brunswick was performed on the transferability of the collision prediction capabilities of Interactive Highway Safety Design Model (IHSDM) within a rural Canadian context. Collision prediction models are generally created with data that intrinsically reflect geographic, environmental, and operational characteristics unique to a particular region. The transferability of a model to a different area can sometimes be problematic. The IHSDM model has been developed with two levels of calibration that are intended to facilitate its widespread application.

This study evaluated both the 2008 and 2009 (beta) versions of the collision prediction models for two lane rural roads embedded in IHSDM. A sample of seventy-five rural two-lane road segments from the province of New Brunswick was analyzed to evaluate the model's performance. The analysis compared the predicted collision frequencies with the empirical collision data provided by the New Brunswick Department of Transportation. Results of the study have shown that overall the model performs poorly when compared with the average number of observed collisions. Predicted collision frequencies for all test sites combined were over-estimated by 38 to 78 percent (depending on the level of calibration employed). Goodness-of-fit testing including mean absolute deviation, mean prediction bias, and R^2 showed that the model does not perform well within a Canadian context and that the calibration methods need refining. Model fits described by coefficient of determination, R^2, ranged from only 0.001 to 0.255.

Résumé

Une étude a été réalisée à l'Université du Nouveau-Brunswick sur la portabilité du potentiel de prédiction des collisions du modèle interactif de conception de la sécurité routière (IHSDM) dans un contexte rural canadien. Les modèles de prédiction des collisions sont généralement créés à l'aide de données qui reflètent intrinsèquement les caractéristiques géographiques, environnementales et opérationnelles propres à une région donnée. La portabilité d'un modèle à une région différente peut parfois poser des problèmes. Le modèle IHSDM a été conçu avec deux niveaux d'étalonnage dont le but est de faciliter la généralisation de son application.

Cette étude a évalué à la fois les versions 2008 et 2009 (bêta) des modèles de prédiction des collisions sur les chemins ruraux à deux voies enchâssés dans l'IHSDM. On a analysé un échantillon de 75 tronçons de chemins ruraux à deux voies dans la province du Nouveau-Brunswick pour évaluer l'efficacité du modèle. L'analyse a comparé les fréquences prévues des collisions aux données empiriques sur les collisions fournies par le ministère des Transports du Nouveau-Brunswick. Les résultats de l'étude ont démontré que, dans l'ensemble, le modèle a un comportement plutôt médiocre par rapport au nombre moyen de collisions constatées. Les fréquences prévues des collisions à tous les sites expérimentaux confondus ont été surestimées de 38 % à 78 % (selon le niveau d'étalonnage utilisé). Les tests de validité de l'ajustement, notamment l'écart absolu moyen, le facteur moyen de correction de la prévision et R^2 ont révélé que le modèle se comporte mal dans le contexte canadien et que les méthodes d'étalonnage doivent être fignolées. Les ajustements du modèle décrits par un coefficient de détermination R^2 variaient d'à peine 0,001 à 0,255.

INTRODUCTION

Collision prediction modeling has many uses in road safety engineering and planning including black spot identification, road safety audits, and benefit/cost analysis of road improvements. In recent years, collision prediction models have been developed using data from many different geographic areas. These models are typically designed to be applied in specific areas and problems arise when attempts are made to use these models in different regions. This is partly due to the fact that collisions are random events which can have many regional specific factors influencing them.

With the development of the Interactive Highway Safety Design Model (IHSDM) by the Federal Highway Administration (FHWA), a standardized approach to safety analysis including collision prediction has been sought [1]. Since IHSDM was originally released, several upgraded versions have been developed with each new version generally introducing a new element or an improvement to other principal components. The latest full version of IHSDM was released in October 2008. The initial focus of the IHSDM was on rural two lane roads; however, a beta version released by the FHWA in November 2009 includes capabilities to predict collisions on urban, suburban, and multilane facilities [2].

Since the release of IHSDM, few studies have been performed on the transferability of the model. This report synthesizes a study performed at the University of New Brunswick to test the transferability of the collision prediction module within IHSDM, as well as the level of calibration required to produce optimal results. If this design suite is to become widely adopted in the United States, it is important to understand its limitations and whether it can accurately be employed in the Canadian context.

IHSDM BACKGROUND

In an attempt to improve the level of safety included in road design and planning, the FHWA has been developing a suite of tools collectively known as the IHSDM since the mid-1990s. It is widely expected that these tools will establish a new standard for highway design in the United States. The IHSDM software was initially developed when a deficiency was recognized in checking road compliance with federal, state, and local policies. A need was also recognized to determine the road users' comprehension of road designs in their driving practices [1]. The current full release of the IHSDM software suite contains six analysis modules. The components are [2]:

- Collision prediction
- Driver/vehicle
- Policy review
- Design consistency
- Traffic analysis
- Intersection review

The collision prediction model embedded in the October 2008 version was created in the late 1990's by Harwood *et al.* [3]. This collision prediction model was only developed to prediction collisions on two-way, two-lane rural highways. The model was created by using negative binomial regression using data taken from Minnesota and Washington. The model applies the following basic structure:

$$\mathbf{N_{rs} = N_{br} (AMF_1\, AMF_2 \ldots\ldots AMF_N)} \qquad (1)$$

Where N_{rs} is the total number of predicted collisions for a given road segment. Nbr represents the number of collisions under nominal conditions, or a base condition. This is calculated by using several factors including lane width, shoulder width, segment length, average daily traffic, etc. Each AMF represents a multiplication factor which positively or negatively affects the number of collisions at a location. The ideal AMF factor is set to 1.0, where no change will occur to the model.

In order to facilitate transferability of the collision prediction model, a calibration method was derived by Harwood *et al.* The calibration model contains two levels for which calibration can be performed, named "level 1" and "level 2." Both of these levels of calibration require a minimum set of road segment specific data including the total length of two-way, two-lane highway, accident data, alignment and grade data, and require this information to be sorted into several traffic volume groups. For a level 2 analysis, the same data is required but it is further segregated into sub-groups by shoulder and lane widths [3].

The beta version of a newer IHSDM collision prediction model was released in November 2009. This new model is set to conform with Part C of the upcoming Highway Safety Manual [2], which has been scheduled for release sometime later in 2010 [4]. Since the Highway Safety Manual has not been released at the time of publication, it is unknown what level of calibration will be included, or whether there is a major fundamental difference beyond the 2008 version. An important item of interest with the 2009 version of IHSDM is the inclusion of urban and suburban roads, as well as the inclusion of multi-lane facilities in rural areas.

PREVIOUS STUDIES

Since the initial development of IHSDM there have been few studies on the accuracy and transferability of the model. Several projects have been undertaken using the suite of software to evaluate the potential for collision reductions in realignment projects, but no retrospective reports on the accuracy of the model have been published [2].

Saito and Chuo [1] performed a study in 2008 of the viability of using IHSDM in road safety audits. Their study focused on three road segments, the US-10, SR-150, and US-6. The study utilized the built-in Empirical-Bayes (EB) modification utility by using previous collision history on the three highway segments. The study by Saito and Chuo showed a large variation between the actual collision data and the estimated collision rates.

A study by Donnell *et al.* [5] tested the IHSDM collision prediction model on two highway segments in the state of Pennsylvania. The Donnell *et al*. study tested IHSDM using the level 1 and level 2 calibration techniques over three geographic areas: county, district and state. The study, along with the Saito and Chuo study found there was a large variation between the actual and estimated collision data.

A third study was performed at the University of New Brunswick in 2006. The study compared three collision prediction models to historical data for the province. The three models tested included IHSDM, the Transportation Association of Canada model, and MicroBENCOST. The study found that of the three models, IHSDM produced estimates that more closely represented observed values, however, it was noted that the prediction error was approximately 46 percent for the total number of collisions [6].

MODEL EVALUATION

In order to evaluate the IHSDM collision prediction model, 75 random test segments across the province of New Brunswick were selected. Criteria for these test segments were taken and modified from research published by Ye [6] in which each test segment was required to be over 2.0 km in length, and more than 80 m from any intersection in order to reduce the number of intersection related crash data from being included.

Since there were two versions of the IHSDM collision prediction model available for use, both were tested. The 2008 version of IHSDM, the latest full release version, was tested at three application levels. The tests included evaluating the model in a base scenario over five years using an un-calibrated model and comparing the estimated collision rates to the actual observed collision rates for the test segments. The second and third evaluations of the 2008 model used the level 1 and level 2 calibration methods and compared their respective estimated collision rates to empirical crash data. The 2009 beta version of the IHSDM was tested in an un-calibrated state only as the documentation on the calibration process has not been made available at the time of study.

The model was not evaluated the accuracy of the severity of collisions predicted. This is because the IHSDM model uses set values for each type of collision, such as property damage only, injury, and fatality. It is possible to set these values to reflect the conditions for the geographic area including the configuration of collisions that occurred. The calibration for collision configuration does not affect the overall number of collisions predicted as each modification factor for the type of collision sums to 100% of the total number of collisions.

Both versions of IHSDM have a built-in Empirical-Bayes calibration method. This method utilizes previous collision history and will have the predicted number of collisions conform more to the historical values. This method was not tested in this study as the effects of the Empirical-Bayes method will sway the predicted number of collisions closer to the actual historical values and would undermine the test of goodness-of-fit of the model under actual conditions. The intended application of this method is more related to estimating the impacts of design modifications to an existing road which is outside the objective of this study.

Data Collection and Model Calibration

Test segment data were acquired from the New Brunswick Department of Transportation and included basic characteristics such as average annual daily traffic, speed limits, road and shoulder widths, degrees of curvature, grade severities, intersection locations, and collision history. Once the data were collected, the information was entered into the IHSDM Highway Editor in order to model the highway segment within the software.

In order to perform the level 1 and level 2 calibrations, more road segment descriptive data were required for each test segment. All of the data collected for the calibration methods were placed into different annual daily traffic (ADT) groups as well as subdivided by lane and shoulder widths. The data collected for the level 1 and level 2 calibrations included the following:

- Total length of road
- Total length of road on grade and on a curve
- Average grade for those sections on grade
- Average curvature for those sections of curves
- Total collisions

- Shoulder and road widths
- Percentage of segment length that is flat, rolling, or mountainous

The calibration method for IHSDM consisted of using the Administration Tool provided with the software. The tool creates a spreadsheet where the collected data for the road network is compiled. Once the data entry is complete, the spreadsheet calculated a calibration ratio (C_r). Depending on the various geometry and collision history the value will be either above or below zero. The calibration ratio is then entered into a configuration file and used in the analysis of the section. In the case of this project, the C_r values were 1.291 and 1.283 for the level 1 and level 2 calibrations, respectively, which basically means that the base model was estimated to under-predict the total collisions by 28-29% given the site-specific characteristics included in the calibration algorithms.

Data Analysis

The model evaluation techniques used in this research consisted of testing the goodness-of-fit of the estimated collision rates from IHSDM to the historical data provided by the New Brunswick Department of Transportation. Three tests of goodness-of-fit were undertaken during this project including the mean prediction bias, the mean absolute deviation, and linear regression.

The mean prediction bias (MPB) test provides a measure of the goodness-of-fit of the model by comparing the overall difference between the test data and the historical data, as well as showing what direction the output is from the historical data. A low MPB value indicates the model performs well in comparison to the historical data, where a high MPB value indicates poor conformance. Positive MPB rates show the model over-predicts the number of collisions, while negative MPB rates show the model under-predicts. MPB is calculated using the following formula:

$$\mathbf{MPB} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - Y_i)}{n} \qquad (2)$$

Where $\hat{Y}_i$ is the predicted value, Y_i is the actual value and n is the number of samples [7].

The mean absolute deviation (MAD) provides a similar goodness-of-fit comparison as the MPB test does; however, the MAD model gives the average difference in prediction of the model in an absolute format, meaning negative and positive differences in prediction will not cancel each other out. Similar to the MPB, values closer to 0 show that the model performs well when compared to historical data. MAD is calculated using the following formula:

$$\mathbf{MAD} = \frac{\sum_{i=1}^{n}|\hat{Y}_i - Y_i|}{n} \qquad (3)$$

Where $\hat{Y}_i$ is the predicted value, Y_i is the actual value and n is the number of samples [7].

Linear regression can be used to show whether or not there is a direct linear relationship between the model output and the observed collision data when plotted against one another. The model takes a general form of:

$$\mathbf{Y = a + b*X} \tag{4}$$

Where Y is the predicted number of collisions, X is the historical number of collisions, “a” is the y-intercept and “b” is the slope of the linear line. Generally, the model will want to have an intercept close to 0 and a slope close to 1 in order to show a strong relationship between the model output and empirical data. A test of the strength of the linear relationship between the two sets of data is reflected by the R^2 coefficient. The value of R^2 is always between 0 and 1. A value of R^2 that is closer to 0 shows that there is very little or no linear relationship between the two variables and the model is not a good fit [6].

All of the IHSDM output data were compared against the observed average number of collisions per year, and the average number of collisions per million vehicle-kilometres. These data were delineated into eight sub-groups which were selected based on the road classes and sub-groups required to perform the level 1 and level 2 calibrations provided by Harwood *et al.* [3]. The sub-groups used to perform the data analysis were as follows:

- All test sections
- Arterial roads
- Collector roads
- Average daily traffic (ADT) under 1,000
- Average daily traffic (ADT) between 1,001 and 3,000
- Average daily traffic (ADT) between 3,001 and 5,000
- Average daily traffic (ADT) between 5,001 and 10,000
- Average daily traffic (ADT) over 10,0001

Table 1 gives a breakdown of how many test segments are located within each sub-group. The sub-group with average daily traffic over 10,001 has a very small sample size. This is due in part to the low volume of many of the roads in rural New Brunswick and the data requirements for road segment length between intersections.

Table 1 – Road Segment Sample Sizes

Arterial Roads	Collector Roads	ADT Volume Bins				
		Under 1,000	1,001 – 3,000	3,001 – 5,000	5,001 – 10,000	Over 10,001
24	51	16	24	21	12	2

RESULTS

All Test Segments Combined

The data in Tables 2 and 3 synthesize the results for the model results for all 75 road test segments. The data in Table 2 represent the results for collisions per year while the data in Table 3 represents the results for collisions per million vehicle-kilometres (mvkm).

Table 2 - All Test Segments Results (collisions/year)

	Actual	IHSDM Model Results			
		2008 No Calibration	2008 (Level 1)	2008 (Level 2)	2009 (beta)
Avg. of all segments	1.279	1.760	2.276	2.255	2.089
MPB	-	0.481	0.997	0.976	0.810
MAD	-	0.843	1.183	1.167	1.034
R^2	-	0.184	0.184	0.184	0.177

Table 3 - All Test Segments Results (collisions/mvkm)

	Actual	IHSDM Model Results			
		2008 No Calibration	2008 (Level 1)	2008 (Level 2)	2009 (beta)
Avg. of all segments	0.503	0.587	0.759	0.753	0.691
MPB	-	0.085	0.257	0.250	0.189
MAD	-	0.282	0.379	0.375	0.330
R^2	-	0.001	0.001	0.001	0.005

The results in Tables 2 and 3 show that, on average, the 2008 model with no calibration is the most accurate in predicting the number of collisions in terms of both collisions per year and collisions/mvkm. The difference between the averages for collisions per year and collisions/mvkm are 38% and 17%, respectively. It is noteworthy that the calibration (levels 1 and 2) exercise actually adjusted the base model output in the wrong direction. According to the calibration algorithms, the values for variables which further describe test segment geometry and characteristics result in correction factors that expect the base model to be under-estimating collisions. This was found to be in error since the model always over-predicted actual collision experience. It is also of note that the 2009 beta model generates results that are worse than the earlier 2008 base model.

When the data are compared with MPB and MAD statistics along with the R^2 value, it was found that the models do not accurately predict the total number of collisions. This was reflected most with the MAD values which range from 0.843 to 1.183 for annual collision frequencies. When evaluating the data based on the number of collisions/mvkm travelled, the MPB values are closest to 0 for the model when it is applied as un-calibrated. The R^2 values ranging from 0.001 to 0.184 indicate that the collision prediction models contribute little beyond using a simple across-the-board average,

The data synthesized in Figure 1 represent the estimated collision frequencies plotted against the observed frequencies for the 2008 IHSDM model with no calibration. The top line in the figure represents the 45-degree line that would occur if the model results perfectly reflected the actual collision frequencies. The second line represents a best-fit linear regression line. It is noteworthy that there are several large values predicted by the model compared to the relatively small values representing the actual collision rates (bottom-right corner of plot area).

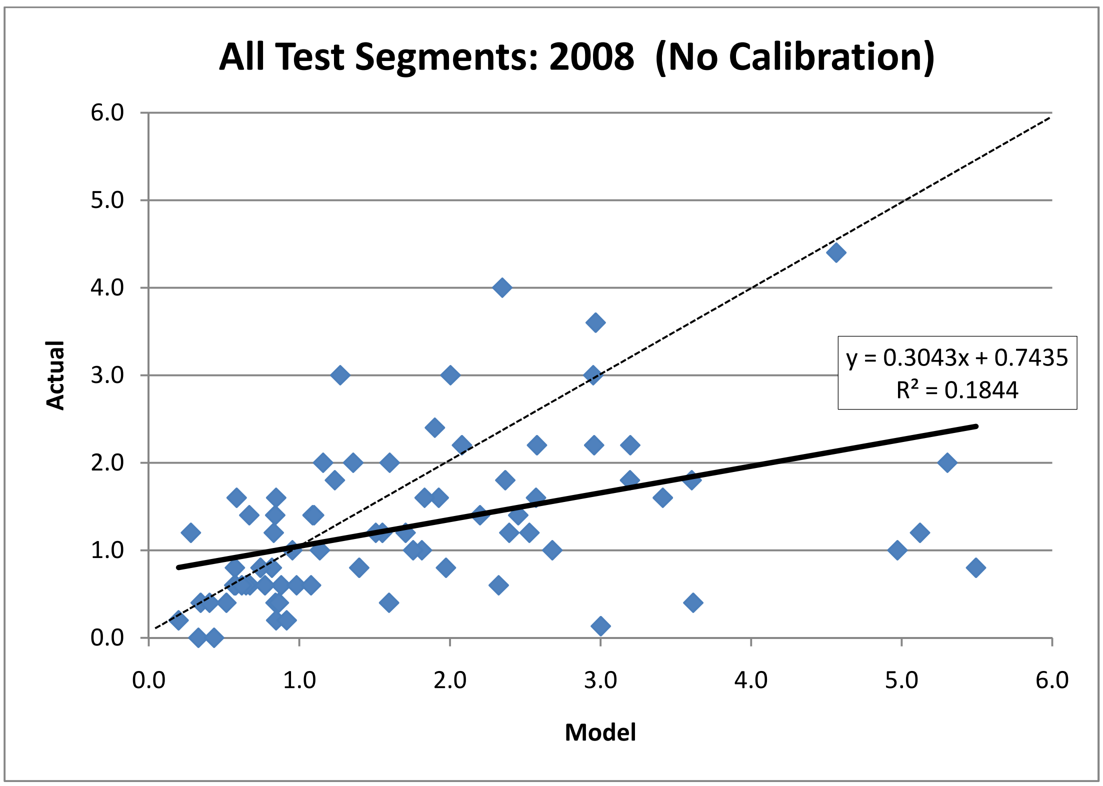

Figure 1: 2008 Model - All Test Segments (collisions/year)

Arterial Roads

The data in Tables 4 and 5 synthesize the average, MPB, MAD and R^2 values for the 24 arterial road segments that were tested. The ADTs for these roads ranged from under 1,000 vehicles per day to over 10,000 vehicles per day.

Table 4 - Arterial Roads (collisions/year)

	Actual	IHSDM Model Results			
		2008 No Calibration	2008 (Level 1)	2008 (Level 2)	2009 (beta)
Avg. of all arterials	1.481	1.871	2.417	2.391	2.192
MPB	-	0.391	0.937	0.910	0.711
MAD	-	0.887	1.144	1.125	0.979
R^2	-	0.010	0.010	0.010	0.015

Table 5 - Arterial Roads (collisions/mvkm)

	Actual	IHSDM Model Results			
		2008 No Calibration	2008 (Level 1)	2008 (Level 2)	2009 (beta)
Avg. of all arterials	0.466	0.491	0.633	0.627	0.498
MPB	-	0.025	0.167	0.161	0.032
MAD	-	0.210	0.249	0.245	0.210
R^2	-	0.024	0.015	0.016	0.020

The results show that the arterial road segments exhibit the same behaviour as the data synthesized in Tables 3 and 4. In the case of arterial roads the IHSDM model is shown to over-predict collisions/year by no less than 26 percent (2008 model with no calibration). When expressed as collisions/mvkm, the model over predicts by at least 5 percent. Similar to Tables 3 and 4 the MPB and MAD for the collisions per year are higher than the collisions per million vehicle kilometres with the collisions /mvkm displaying closer matches to the actual collision rates. Again, very weak R^2 values indicate poor model fit.

The data in Figure 2 illustrate the comparison between model output and actual observed collision data for arterial roads only.

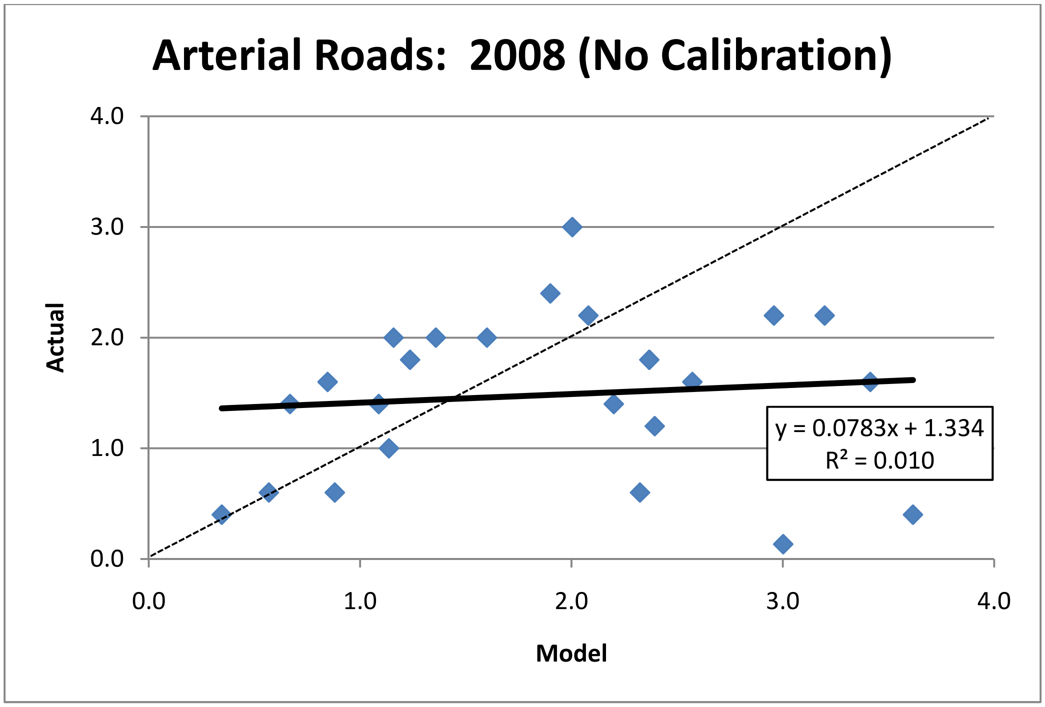

Figure 2: 2008 Model – Arterial Roads (collisions/year)

Collector Roads

The data in Tables 6 and 7 summarize the average, MPB, MAD, and R^2 values for the 51 collector roads that were tested. The ADTs for these roads ranged from under 1,000 vehicles per day to over 10,000 vehicles per day.

Table 6 - Collector Roads (collisions/year)

	Actual	IHSDM Model Results			
		2008 No Calibration	2008 (Level 1)	2008 (Level 2)	2009 (beta)
Avg. of all collectors	1.184	1.708	2.210	2.191	2.041
MPB	-	0.476	0.978	0.960	0.857
MAD	-	0.866	1.238	1.223	1.060
R^2	-	0.254	0.254	0.255	0.238

Table 7 - Collector Roads (collisions/mvkm)

	Actual	IHSDM Model Results			
		2008 No Calibration	2008 (Level 1)	2008 (Level 2)	2009 (beta)
Avg. of all collectors	0.520	0.633	0.819	0.812	0.747
MPB	-	0.113	0.299	0.292	0.227
MAD	-	0.316	0.441	0.436	0.385
R^2	-	0.001	0.001	0.001	0.001

The data in Tables 6 and 7 exhibit the same behaviour as shown in the previous tables, with the 2008 un-calibrated model performing better than the calibrated and 2009 models. In terms of collisions per year, the model gave a 44% difference between the estimated and actual crash rates. Collisions /mvkm were over-estimated by at least 22%. In terms of MPB and MAD the collisions/mvkm again showed the better results. All R^2 values were very weak (0.001 to 0.255) indicating that the model adds little value in estimating collisions.

Figure 3 illustrates the comparison between model output and actual observed collision data for collector roads only.

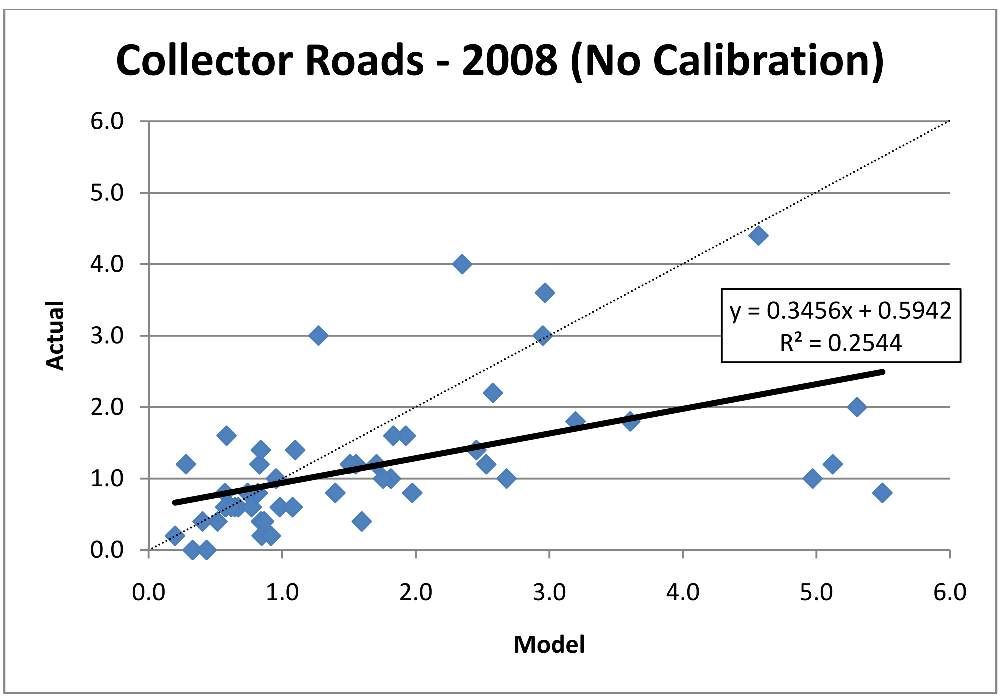

Figure 2: 2008 Model - Collector Roads (collisions/year)

Traffic Volume Sub-groups

In previous sections it was noted that five sub-groups of road segments were delineated for varying levels traffic volumes. Modelling collision frequencies for roads in different volume bins showed no improvement over previous results. Again, model estimates yielded very poor MPB, MAD and R^2 values.

It should be noted that in the data a trend was observed between each traffic volume sub-group in that the maximum differences between the predicted and the actual collision rates would grow larger as the ADT increased. The lowest absolute maximum difference between predicted collisions/year and that observed was 1.016 in the under 1,000 ADT group and the largest absolute maximum difference noted was 5.766 in the over 10,000 ADT group. This was only taken in the average number of collisions per year as the ADT values were normalized when the collision rates per million vehicle kilometres were used.

CONCLUSIONS

The collision prediction models embedded in IHSDM for two-lane rural roads were found to yield very poor results for road segments in New Brunswick. Predicted collision frequencies were over estimated by 38 to 78 percent (for all test segments combined) depending on the level of calibration employed. R^2 values that describe model fit ranged from 0.001 to 0.254 indicating little added value by the model framework. Level 1 and level 2 calibration techniques typically worsen the results over the base model. The beta version of the 2009 collision prediction model produced poorer results than the 2008 model across-the-board. Application of the model to more homogenized sub-groups of road segments (road class and volume bins) did not improve the predictability of the model.

One of the main issues in the calibration of the model was the effort required to collect the necessary data and combining the data into the appropriate sub-groups. The amount of time and resources required to perform a full calibration for the entire province would be great. A secondary issue with the model calibration is the accuracy of the collected data. The data may not be up to date at the time of study and changes are generally made very year. The calibration method was also seen as unsuccessful as the calibration factors are applied across all test sections as one value. With a model that over-predicts and under-predicts fairly equally, very little effect on the final results will occur.

REFERENCES

[1] SAITO, M., Chuo, K., Evaluation of the applicability of the Interactive Highway Safety Design Model to safety audits of two-lane rural highways, *Utah Department of Transportation Report US-2-08,* March 2008.

[2] FEDERAL HIGHWAY ADMINISTRATION, IHSDM website, *www.ihsdm.org*, 2009

[3] HARWOOD, D.W., Council, F.M., Hauer, E., Hughes, W.E., and Vogt, A., Prediction of the expected safety performance of rural two-lane highways. *FHWA Report FHWA-RD-99-207,* 2000.

[4] TRANSPORTATION RESEARCH BOARD. Highway safety manual website, *www.highwaysafetymanual.org,* 2010.

[5] DONNELL, E.T., Gross, F., Stodart, B.P., Opiela, K.S., Appraisal of the interactive highway safety design model's crash prediction and design consistency modules: case studies from Pennsylvania, *Journal of Transportation Engineering*, Vol. 135, No. 2, pp. 62-72, February 2009.

[6] YE, H., Accident prediction for New Brunswick's two-lane rural highways, *Master's Thesis*, University of New Brunswick, Fredericton, New Brunswick, September 2005.

[7] Oh, J., Lyon, C., Washington, S., Persaud, B., and Bared, J. Validation of FHWA crash models for rural intersections. Transportation Research Record, Vol. 1840, pp. 41-49, 2003.