Accident Prediction Model Based on Traffic and Geometric Design Characteristics

Mahmoud Saffarzadeh[1] Maghsoud Pooryari[2]
1.Associate Professor, Tarbiat Modares University, Tehran, Iran
E-mail: saffar_m@modares.ac.ir
2.Research Associate, Transportation Research Institute, Tehran-Iran
E-mail:mpooryari@yahoo.com

Abstract:*This paper specifies the relationship among various factors contributing to road accidents including geometrical design characteristics, environmental and traffic specifications, by multiple regression analysis. The main objective of this paper is identification of problems associated with the safety issue of road networks by application of accident prediction models. Data from previous accidents were used to develop the models.*
Results of this study showed that the rate of road accidents is to a large extent dependent on the rate of traffic volume. Type of road and land-use are other important factors influencing the number and intensity of accidents. The mountainous roads in this respect require special attention regarding their safety factors. The quantitative rate of road safety upgrading has also been specified by adding traffic lanes in road networks.

Keywords: *Accident Prevention, Highways, Traffic, Safety*

1- Introduction

Accident prevention is the most effective way to reduce the rate of fatalities and increase the safety of roads. In this relation, the prediction models are very good tools in analysis of accidents. Due to the high number and complexity of road accidents, identification of accident prone areas and suitable alternatives for their prevention is quite difficult.
More than 1.17 million persons are killed in accidents every year, 70% of whom are residents of developing countries. In road accidents more than 10 Million get injured, resulting in various handicaps. If appropriate measures are not taken, it is predicted that within the next 10 years, about 6 million people will die and 60 million get injured.
Studies of the World Health Organization show that in 1990 the road accidents occupied the 9th rank in the health problems list and it is predicted that till 2020, the road accidents will reach the 3rd rank in the list of death and handicap causes in human societies.
As some of the influencing factors of road accidents and their interaction are qualitative and random in their nature, (like the behavior of drivers and the climatologic conditions), the actual relationship between accidents and their factors are experimental and probable. For instance, a road that is considered as safe, appropriate and ideal from designing point of view can be unsafe and dangerous when used by careless drivers.
A major reason that makes the development of scientific prediction models difficult is the data collection of accidents and their inappropriate classification. It is tried in this research to develop a model for the purpose of accident prevention in intercity roads so that it can be used as a tool for measurement of road safety [1].

2- Research Background

Miaou et al. (1992) used Poisson regression on traffic data from 8779 miles of roadway from the Highway Safety Information System (HSIS) to establish quantitative relationships between truck accident rates and highway geometric characteristics. Their results indicate that surrogate measures for mean absolute curvature (for horizontal alignment) and mean absolute grade (for vertical alignment) are the most important variables for accident rate estimation [2].

In a study of approximately seven thousand miles of roadway lengths in Utah, Mohamedshah et al. (1993) used linear regression to predict truck accident involvement rate per mile per year, based on Average Annual Daily Traffic (AADT) and truck AADT per lane, shoulder width, horizontal curvature, and vertical gradient. The results suggest that truck involvement rate increases with AADT and truck AADT, degree of curvature and gradient [3]. Hadi et al. (1993), using data from the Florida Department of Transportation's Roadway Characteristics Inventory (RCI) system, estimated negative binomial (NB) regression for accident rates on various types of rural and urban highways with different traffic levels. Their results suggest that higher AADT levels and the presence of intersections are associated with higher crash frequency, while wider lanes and shoulders are effective in reducing crash rates. In that paper, the authors also provide an extensive review of earlier findings relating accident rates and geometric characteristics [4].

More recently, Ivan and O'Mara (1997), using regression on 1991-1993 data from the traffic Accident Surveillance Report of Connecticut found that annual average daily traffic was a critical accident prediction variable, while geometric design variables and speed differential measures were not found to be effective predictors of accident rates [5]. Karlaftis and Tarko (1998), based on a county accident data set from Indiana, estimated macroscopic accident models that attempt to explicitly control cross-section heterogeneity in NB regression that may otherwise seriously bias the resulting estimates and invalidate statistical tests. Data collected from the states of Minnesota and Washington on rural two-lane highways, estimated accident models for segments and three-legged and four legged intersections stop-controlled on the minor legs. Independent variables for their models included traffic, horizontal and vertical alignments, lane and shoulder widths, roadside hazard rating, channelization, and number of driveways. Results imply that segment accidents depend significantly on most of the roadway variables collected, while intersection accidents depend primarily on traffic [6].

Vogt and Bared (1998) used the following model for road sections containing curves in which the basic parameters are known. The calibrated models are as follows:

$$\text{Crashes/year} = (L)\,(AADT)^{b}\,EXP\,[a + hW + g\,(TER) + p(H)] \tag{1}$$

where:

L= Section length in kilometers,

W= Total surface width (lanes and shoulders) in meters,

H= Weight of average curvature value,

TER= Terrain classification, a categorical variable that takes on the value of 1 if the section is in flat terrain or of 2 if it is in rolling terrain,

AADT = Average Annual Daily Traffic, and

b,a,h,g,p = Constant coefficients.

These are the values as coded in the original data, so it was decided to estimate the model coefficients for categories so defined. The model estimates would have been the same using a more conventional procedure, of using a dummy variable taking on a value of 0 for flat terrain or 1 for rolling terrain. As mentioned above, H is a weighted average curvature, and is defined as:

$$H = \sum WH_i (1000 / R_i) \tag{2}$$

where WH_i, is the length of horizontal curve i lying in the segment divided by the segment length, and R_i is the radius in meters for horizontal curve i [7].

A model developed by Harwood, et al.(2000) for predicting the number of crashes on two-lane highway segments uses the following independent variables as:

- Average Daily Traffic,
- Lane and shoulder widths,
- Driveway density,
- Roadside hazard rating,
- Horizontal curve lengths and radius, and
- Vertical curve lengths

The mathematical formulation of these models usually is as follows [8]:

$$\textit{Number of crashes} = \textit{Exposure Measure} \cdot \exp(C0 + C1X1 + C2X2 + \ldots) \quad (3)$$

The same researchers, Harwood, et al., also developed a crash prediction model for intersections on highways. The independent variables used in this model include the following:

- Average Daily Traffic,
- Roadside hazard rating,
- Exclusive right turn lane,
- Angle of intersection,
- Driveways on approaches,
- Protected phases of traffic signals,
- Left turn percentages,
- Grade, and
- Percent of trucks.

The independent variable used by Persuad model was just Average Daily Traffic (ADT). The model formulation is as follows [9]:

$$\textit{Number of crashes on a road section} = \textit{Section Length} \cdot a1\,(ADT)^{b1} \quad (4)$$

The coefficient 'a1' and the exponent 'b1' of the model depend on the class of road. This model applicable for long-range planning, since future ADT values for individual sections/links of future network are predicted with the help of travel forecasting models, and the classes of different roads.

3- Conceptual Formation of Model

The accident model in this research was formed in a way that based on conducted studies, factors affecting road accident and its relevant parameters in the number of accidents were identified. Then information including geometric characteristics of road, traffic flow and volume, road environment and accident reports for routes under study were collected. To process data and determine the correlation between different kinds of accidents and effective factors using the regression technique, different forms of functions for accident prediction were analyzed. Then with the evaluation of quantities and experiments used in the assessment of model, reliability of the whole model and parameters (used as independent variables) was performed and at the end an appropriate model was recommended.
One of the prominent characteristics of this model is the relation between factors affecting accidents and the severity of accidents.

4- Selection of Effective Variables in Model

The effective variables selected for model are as follows:

- Average Daily Traffic (ADT)

This is a basic variable which encompasses the role of human behavior, vehicle and combined effect of these two. It is obvious that no matter how dangerous the road may be, there will be no accidents in it unless there is traffic.

- Road Surface Width (RSW)

This variable expresses width of the road (total sum of both traffic directions) in chosen sections and its unit is in meters. In roads with divided lanes, width of the road in only one direction is taken into account.

- Shoulder Width (SW)

This variable, takes into account shoulder width in different sections of the road and its unit is in meter. This includes paved and unpaved surface and in case of divided lanes, width of the right shoulder is taken into account.

- Percentage of Commercial Vehicles (PCV)

Considering the unique movement of commercial vehicles, this variable is considered in its involvement in road accidents.

- Access per Kilometer (APK)

This variable includes cross roads and ramps in sections under study and number of access to commercial centers, offices, complexes and etc… For this reason, the number of cross roads, exit and entrance ramps are taken into account.

- Terrain Conditions (TC)

This variable is used to consider geographical condition of the area where the road passes through. Due to the unsafe conditions of, mountainous roads and because of restrictions imposed by geometric criteria, this variable is entered into the equation separately. In this regard, using dual variables one and zero for mountainous and other roads are assigned as values.

- Roadside Hazard (RH)

Since a considerable amount of accidents results from run-off-road, this variable deals with safety of the roadside. In this research, percentage of accidents involving vehicles crashing another vehicle, fixed object, turning over or run-off-road is considered.

- Area Location Type (ALT)

One of the important issues in road safety is access control. Formation of commercial, residential and agricultural land uses along the highways results in the mobility of non motorized vehicles and pedestrians in the road areas adding to the considerable insecurity of a road. This variable is used as a qualitative one assigning values one or zero considering population density in the area or density of land use.

- Functional Class (FC)

The rate and type of accident in a divided highway is different from that of a one way road. For this reason, two variables one and zero are used for two way and one-way roads, respectively.

- Accident Index (AI)

This variable indicates the accidents occurred in a section of the road. Since accidents in a road are three types, property damage, injury and fatality, research resulted from Transportation Research Laboratories of England [6] regarding developing countries has been used to equate these three types. Equivalent index used in Equation 1 is calculated as follows:

$$EDA = DA + 3\times ID + 12\times FA \qquad (5)$$

where,
EDA= Accident damage Equivalent,
DA= Accident damage,
ID= Injury damage,
FA= Fatal accident, and
Variable AI is also defined as:

$$AI = EDA / L \qquad (6)$$

where:
L is the road section length in kilometers. One of the major advantages of this model is consideration given to severity of accidents. The main shortcoming of the accident

prediction model is the use of different variables such as the number of accidents or rate of accidents and no differentiation is made in the types of accidents.
Hence in this research, this shortcoming is corrected and different kinds of accidents are equated to take into consideration the relevant elements and severity of accidents.

- Metrological factors

Result of different accidents analyses, shows that the number of accidents varies in different times of the year as the result of traffic volume, and any increase in the number of accident is due to the increasing traffic volume. Because of lack of major difference in the number of accidents in bad weather conditions, no variable is introduced for it. Results of the collected data for the model are given in Table 1.

Table 1: Specification of used variables in model analysis (1994-1999)

Section No	Variable									
	AI	RH	ADT	PCV	ALT	TC	APK	FC	RSW	SW
1	13.4	4.8	364.7	6.1	0	0	0.3	0	11	2
2	6.5	10.3	6551	33	0	0	0.1	0	11	3.25
3	4.1	2.9	19190	72	0	0	0.1	1	8.6	4.3
4	5.9	9.4	4364	42	0	0	0.08	0	11	2
5	7.1	8.7	4893	47	0	0	0.08	0	11	2.25
6	9	1.1	39088	30	1	0	0.1	1	22	3
7	8	3.8	3923	41	0	1	0.3	1	7	1.5
8	4.2	3.2	5137	51.4	1	0	0.22	1	7.3	4
9	38	1.4	16680	25	1	0	0.12	1	7.3	3
10	24	3.9	8368	50	0	1	0	1	8.8	2.25
11	0.6	6.2	2557	27	0	1	0	1	7	3
12	17.8	4.8	35724	8.5	0	0	0.3	0	11	2
13	9	10.3	5929	33	0	0	0.1	0	11	3.25
14	6	2.9	11681	80	0	0	0.1	1	8.6	4.3
15	4.9	9.4	5137	53	0	0	0.08	0	11	2
16	4.6	8.7	5015	42	0	0	0.08	0	11	2.25
17	7.6	1.1	22110	56	1	0	0.1	1	22	3
18	9.6	3.8	4086	41	0	1	0.3	1	7	1.5
19	2	3.2	3921	60	1	0	0.22	1	7.3	4
20	36.1	1.4	24630	18	1	0	0.12	1	7.3	3
21	19.4	3.9	8377	50	0	1	0	1	8.8	2.25
22	1.2	6.2	2092	34.4	0	1	0	1	7	2
23	22.2	4.8	44684	6	0	0	0.3	0	11	2
24	12.6	10.3	8655	37	0	0	0.1	0	11	3.25
25	3.9	2.9	16986	88	0	0	0.1	1	8.6	4.3
26	7.6	1.4	21459	50	0	1	0.1	0	22	3
27	5.7	3.2	4382	66	1	0	0.2	1	7.3	4
28	73	1.4	21518	34	1	0	0.1	1	7.3	3
29	49.7	3.9	9712	57	0	1	0	1	8.8	2.25
30	2.3	6.2	2176	41	0	1	0	1	7	2

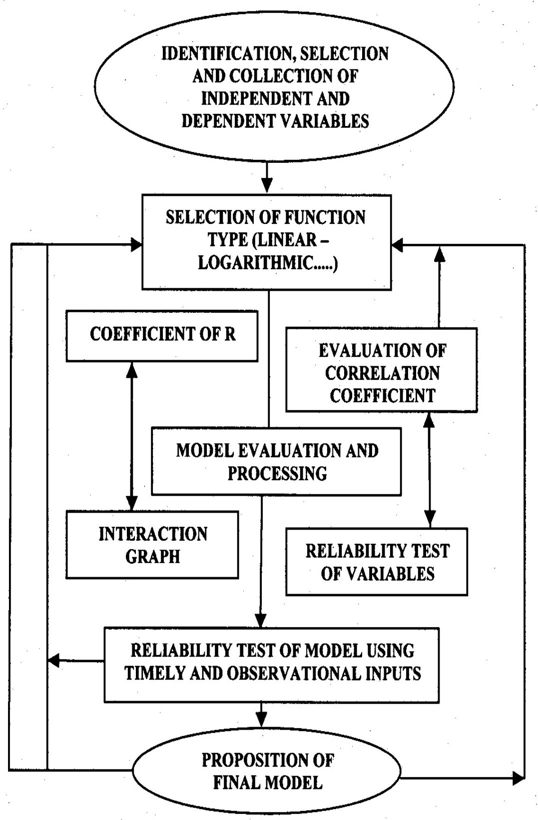

Figure 1: Model Development Procedure

5- Model Development Procedure

The modeling process is shown in Figure 1. To analyze the model, multi linear regression analysis is used with the help of SPSS Software (Statistical Package of Social Science). To obtain the model, linear regression method is used and different function forms is used in a linear fashion. After preliminary analysis, appropriate model is obtained with respect to control criterion. After evaluating and examining different conventional functions for accident prediction, the three following types of functions were chosen for the evaluation of road safety.
Results of the analysis of these functions are given in Table 2, in which the equation gives good explanation of different variables used in the model.
General form of equation to evaluate the

model is based on the following terms:

$$Y = A + \sum (B_i . X_i) \tag{7}$$

$$LnY = A + \sum (B_i . X_i) \tag{8}$$

$$Y = A . X^B . EXP(\sum (B_i . X_i)) \tag{9}$$

where:

Y = contingent variable (number of annual accidents in unit of road (AI)),

X = independent variable, Average Daily Traffic (ADT),

$\boldsymbol{X_i}$ = other variables are independent and are mentioned in Equation 5, and

A, B, $\boldsymbol{B_i}$ = Constant Coefficients.

Table 3: Matrix of correlation coefficients between dependent and independent variables for Eq. 9

	Ln AI	SW	RSW	FC	APK	TC	ALT	PCV	RH	Ln ADT
Ln AI	1.00	-0.21	0.05	-0.16	0.16	-0.14	0.19	-0.29	-0.28	0.61
SW	-0.21	1.00	0.00	-0.09	-0.09	-0.47	0.41	-.59	-0.27	0.14
RSW	0.05	0.00	1.00	-0.05	-0.05	-0.15	0.11	-0.05	-0.14	0.48
FC	-0.06	0.29	-0.32	-0.19	-0.19	0.35	0.46	0.39	-0.69	-0.14
APK	0.16	-0.09	-0.05	1.00	1.00	-0. 31	0.15	-0.32	-0.18	0.31
TC	-0.14	-0.47	-0.15	-0.31	-0.31	1.00	-0.39	0.03	-0.11	-0.42
ALT	0.19	0.41	0.11	0.15	0.15	-0.39	1.00	0.00	-0.58	0.22
PCV	-0.29	0.56	-0.05	-0.32	-0.32	0.03	0.00	1.00	-0.16	-0.26
RH	-0.28	-0.27	-0.14	-0.69	-0.18	-0.11	-0.58	-0.16	1.00	-0.49
Ln ADT	0.61	0.14	0.48	-0.14	0.31	-0.42	0.22	-0.26	-0.49	1.00

6- Model Evaluation

The results of preliminary processing of models are shown in Table 2.

Table 2: Results of preliminary regression analysis of model's functions

Equation	R	R^2	R^{2a}
7	0.774	0.598	0.418
8	0.687	0.472	0.234
9	0.929	0.863	0.801

Variable (R^2) as a correlation coefficient to justify dispersal of relevant variables is used as the result of independent variables. The high values of R^2 in the above table, is indicative of proximity of data from observation and model's prediction of median inputs. Coefficient (R^{2a}) is another primary research quantity used to explain the dispersal of relevant variables as a result of independent variables.

In this regard, the analysis of model in Equation 9 is an appropriate model.Results obtained from the correlation coefficient (R) between dependent and independent variables are shown in Table 3.

The degree of correlation between accidents and amount of daily traffic variable shows the importance and effect of this variable in the model. Low correlation between independent variables shows that each one of these independent variables had no effect on other variables and the value of both independent variables is protected independently, that is the result of choosing proper variables.

Since regression coefficient provides an intuitive understanding of the importance of variables, the low value of some dependent coefficient variables with dependent variables, does not reduce the reliability of variable, but it creates doubts about the importance of the variable that should be tested again in t-student. Variable (t) expresses whether or not a variable is important in predicting accidents. The test values for Equation 9 shown in table 4.

Table 4: Student t test values for variables in Equation 9

variables	Constant	sw	RSW	FC	APK	TC	ALT	PCV	RH	Ln ADT
T test value	-2.377	-4.25	-6.23	-1.7	0.029	4.63	6.25	5.21	3.79	7.9

With respect to the result of analysis of experiment (t), it was shown that except for the variable of access per kilometer (APK), the rest of the variables have reliability of 90 percent and higher. By eliminating the variable of access per kilometers (APK), the obtained coefficients from the final regression model is given in Equation 10.

$$AI = 0.01144 \times (ADT)^{2.41} \times EXP(0.45 \times RH + 0.047 \times PCV + 3.41 ALT + 2.626 \times TC - 0.83 \times FC - 0.203 \times RSW - 0.735 \times SW) \quad (10)$$

Results of this analysis are used for evaluating safety condition of roads with different functional class and terrain condition. In order to identify which variables had the most effect on the model results, using SPSS software in modeling process, the process started from a model that only had fixed value and in every step added a variable to the model that caused the most change in (R2). Results indicated that average daily traffic (ADT), width of shoulder (SW) and width of road surface (RSW) had the most effects in accident rate, respectively. ADT variable increases the accident rate and variables SW and RSW cause to reduce accident rates. To evaluate the model, a comparison was made between the model results and the values obtained from the field data. Figure 2 illustrates the model outputs with the observed data.

7- Conclusion

In this research, an accident prediction model was developed in which the effects of different traffic and geometric design parameters on the road safety and their contribution on road accident can be investigated.

Based on the model Results, if the traffic volume increases by 33 percent, accident

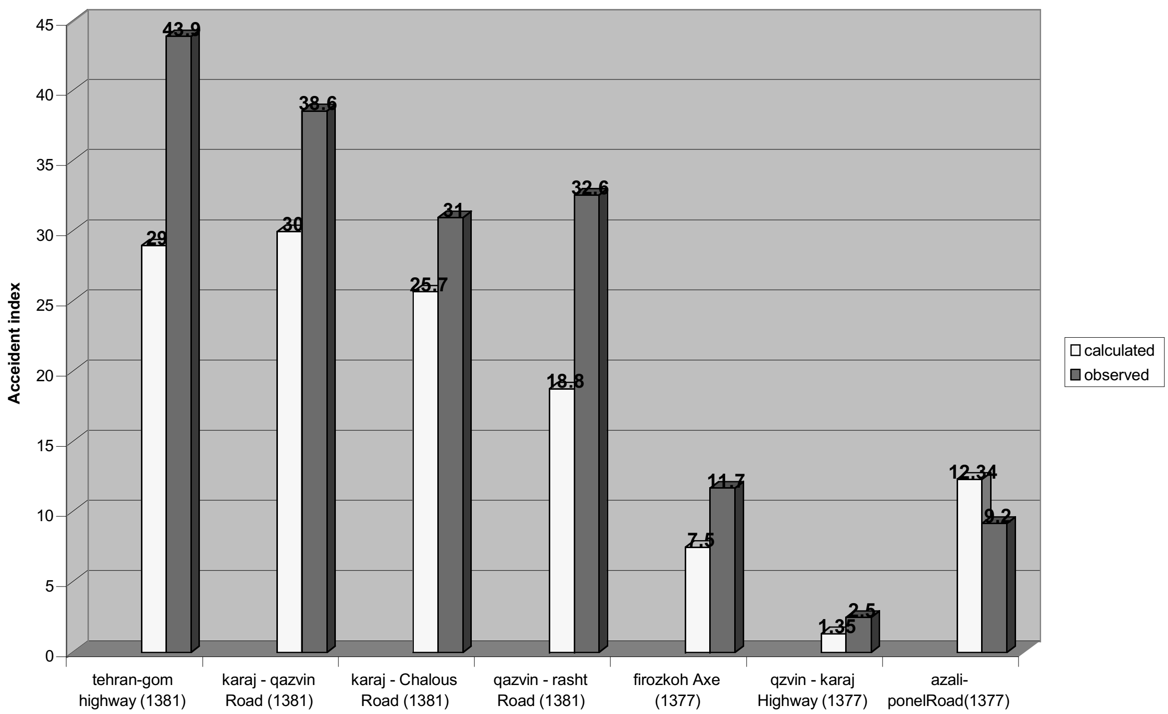

Figure 2: Comparison of model results with the observed data

probabilities will increase by 100 percent, and increasing one meter to the width of shoulder, will improve the road safety by 50 percent. Adding one lane to traffic lanes will result in 50 percent reduction in accidents. But the effect of these variables out of standards given in the model is not clear. Formation of any land use in road environment decreases the road safety, significantly.

In the case of road classification (FC), it is worth to emphasize that separation of traffic lanes in each direction from road to highway, improves safety and reduces accidents by 60 percent.

Based on the obtained results, commercial vehicles are the major causes of road safety deterioration. An assumed 20 percent reduction in commercial vehicle fleet, will result accident reduction by 60 percent.

Shoulders are the other important factors in road safety. Use of road restraints, signs and sign boards especially in mountainous areas and high density accident areas is vital. Also the effect of road environment is significant in accidents in a way that mountainous roads are 13 times more unsafe than other roads. Moreover, in this study it was shown that adverse weather conditions and the number of cross roads per kilometer do not have major effects in safety of rural roads.

8- References

[1]. World Health Organization, "Road Safety not Accident", 2002.

[2]. Miaou, S.P., Hu, P.S., Wright, T., Rathi, A.K., Davis, S.C.,. Relationship Between Truck Accidents and Highway Geometric Design; a Poisson Regression Approach. Transportation Research Record 1376, 10-18., 1992.

[3]. Mohamadshah, Y.M., Paniati, J.F., Hobeika, A.G. Truck Accident Models for Interstate and Two-lane Rural Roads. Transportation Research Record 1407, 35-41. , 1993.

[4]. Hadi, M.A., Aruldhas, J., Chow, L.F., Wattleworth, J.A., Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression. ,Transportation Research Record, 1500, TRB, National Research Council, 169-177., 1993.

[5]. Ivan, J.N., O'Mara, P.J., Prediction of Traffic Accident Rates Using Poisson Regression, Presented in the 1997. Transportation Research Board Meeting, Washington, DC. 1997.

[6]. Karlaftis, M.G., Tarko, A., Heterogeneity Consideration in Accident Modeling., Accident Analysis and Prevention 30 (4), 425-433,1998.

[7]. Vogt, A., and J. Bared., Accident Models for Two-Lane Rural Roads Segments and Intersections. Report RD-98-133. FHWA, U.S. Department of Transportation, 1998.

[8]. Harwood, D.W., Council, F.M., Hauer, E., Hughes, W.E., and Vogt, A. Prediction of the Expected Safety Performance of Rural Two- Land Highways Publication No. FHWA-RD-99-207, U.S. Department of Transportation, Turner- Fairbanks Highway Research Centers, Mc Lean, VA 22101-2298, July 2000.

[9]. Persaud, B.N. Estimating Accident Potential of Ontario Road Sections. In Transportation Research Record 1327, Transportation Research Board, Washington, DC, 1991, PP. 47-53.