IOSR Journal of Business and Management (IOSR-JBM)
e-ISSN: 2278-487X, p-ISSN: 2319-7668. Volume 17, Issue 5.Ver. III (May. 2015), PP 60-70
www.iosrjournals.org

Analysis of relationship between road safety and road design parameters of four lane National Highway in India

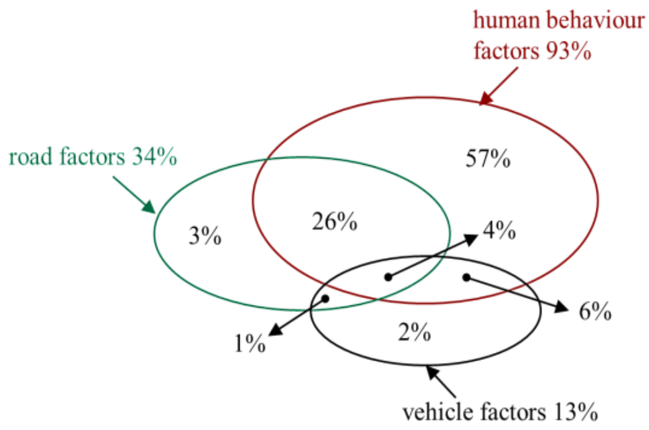
Ravi Shenker[1], Arti Chowksey[2] and Har Amrit Singh Sandhu[3]
[1] *Highway Safety & Engineering, Civil Engineering Department, DCRUST, Murthal, Sonepat, HR, India.*
[2] *Civil Engineering Department, DCRUST, Murthal, Sonepat, HR, India.*
[3] *Civil Engineering Department, PEC University of Technology, Chandigarh, India.*

Abstract: *The road safety depends on humans, vehicles, and road conditions and these factors influence the road safety separately or in combination. It is then essential to investigate the relationship between road accident rates and road design parameters to prevent road accidents and to provide a safe and comfortable driving environment to users.*
The objective of this paper to investigate the relationship between road accident rates and the road design parameters. Also, to develop an accident prediction model to predict the accident rates on four-lane National Highway in Karnataka.
The road accident prediction model was developed by using multiple linear regression analysis. The accident rates were found to be significantly related to road design parameters of study stretch of highway, such as carriageway width, horizontal curvatures, road roughness index & junction (Entry & Exist). The developed model is helpful for the design of safe highways. Additionally, it will contribute to identifying the potentially hazardous locations on highways and to the treatment of safety improvements.
Keywords: *Traffic Accident, Road Safety, Accident Prediction Model, Multiple Regression Model.*

I. Introduction

Road accidents are one of the significant causes of disability, injury and death in the world. The trauma caused by road accidents is unimaginable - physical, mental, financial and many a time irrevocable. India has the highest road traffic accident rates worldwide with over 140,000 deaths annually. Every hour, nearly 14 lives are lost due to road accidents in India. There is consensus forming among the general public due in part to emphatic reinforcement of the accident statistics by traffic authorities that the human element is the key causal factor of road accidents occurrence.

A large number of in-depth accident investigation studies provide a more complete picture of the real accidents causes. One example is shown in Figure-1. This diagram depicts the link between individual areas of the road safety system. It indicates that road accidents are usually the combination of the driver, the road, the vehicle. Human factor seems to be the dominant cause of accidents compared to the others. Drivers are often involved in accident because of their own errors, but also because they are affected by a combination of highway and/or vehicle elements. It is certainly not only the driver who bears responsibility for the occurrence of accident. Henderson (1971) suggested that focusing too much on the driver as the cause of an accident often masked the ability to see other causes that could reduce accident rates and accident severity. However, the number of accidents can be significantly reduced if the road factor is evaluated properly and highway design is made properly.



(Source: PIARC Road Safety Manual, 2003)

Figure -1: Accident Factors

An accident prediction model is a handy tool to help highway engineers to predict the accident rates as a function of road design parameters over a highway segment.

II. Literature Review

Various research papers have been studied which were available on geometric design standard, which impacts on traffic accidents, traffic analysis, accident analysis & prevention.

Chikkakrishna, Parida & Jain (2013) – used Poisson-Gamma and Poisson-Weibull modelling techniques to analyse road traffic crashes on a stretch of National Highway 58 in India. The study shows that median opening, traffic flow, access road & road-side developments, such as Industrial, Commercial, Residential and School are significantly related to road accident.

Singh & Suman (2012) – Regression analysis was used to study of accident of NH-77 and to develop the accident model. This study presented that road accident rate depends on the AADT & the condition of road or shoulders or both.

Rokade, Singh, Katiyar, and Gupta (2010) - Predicted an accident model using Multiple Linear Regression Analysis for Bhopal city based on the factors influencing road accidents were Road cross-section dimensions, traffic volume , speed, road shoulder width, lighting conditions ,traffic signs and traffic signals.

Jacobs (1976) – used multiple linear regression analysis to analyse the various road, found that the accident rates was significantly related to the number of junction, horizontal curvature, vertical curvature & surface irregularity in Kenya Road and road width & road junctions in Jamaica Road.

Mustakim and Fujita (2011) – used multiple linear regressions to establish the crash prediction models. In first model, the accident point weighting is significantly related to access point, approach speed, annual average daily traffic and vehicle gap. In second model, the accident point weighting is significantly related to access point, approach speed, motorcyclist, motorcar, vehicle gap and traffic light.

Mayora & Rubio (2003) - Developed negative binominal multivariable crashed-prediction model for the Spanish National Network two lane rural roads. The highway variables that have the highest correlation with crash rates in Spain´s two lane rural roads were: Access density, average sight distance, average speed limit and the proportion of no-passing zones. Access density is the variable that influences most the rate of head-on and lateral collisions, while in run-off the road and single vehicle crashes sight distance was decisive.

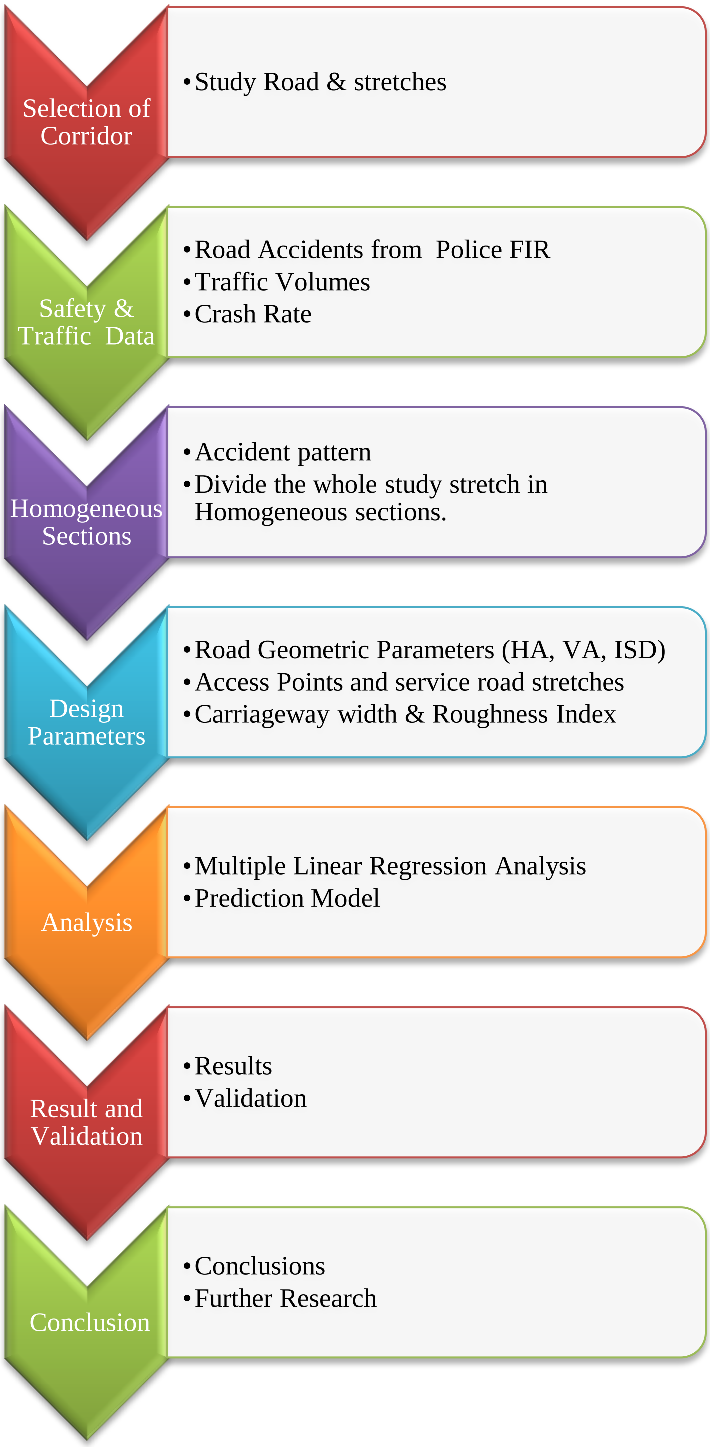


Figure -2: Process Map

III. Modeling Methodology

In this paper, accident rates have been considered as a dependent variable and Road width, Vertical Curvature, Horizontal Curvature, Road Roughness Index, No. of Junctions, Sight Distance and Service Road as independent variables. The Process map as given in Figure-2 describes the methodology adopted for this analysis.

The road design parameters data have been collected from 82 Km stretch and accident data have been collected over the period of 2008-2011 from selected highway. The collected data was split into two parts, first part was used to development of accident prediction model and second part was used to validate the model .

3.1 Accident Rates

The accident rates for roadway segment are calculated as:

$$A_R = \frac{C \times 10{,}00{,}00{,}000}{V \times 365 \times N \times L}$$

Where,

A_R = Roadway accident rates for the road segment expressed as accident per 100 million vehicle-Km of travel-year.

C = Total number of roadway accidents in the study period

V = Traffic volumes using Average Annual Daily Traffic (AADT) volumes

N = Number of years of data

L = Length of the roadway segment in Km

3.2 Homogeneous Sections

On the basis of the accident pattern, the study road was divided into number of homogeneous sections. The cumulative difference area method was used to create homogeneous section of whole stretches.

The Cumulative difference area is a variant of the cumulative sum methods. It is founded in the statistic Z_X that represent the difference between the cumulative areas under the curve of a data series and the cumulative mean area mathematically; it is expressed as below Equation.

Where x_i is the distance between an i^{th} data point and the first data point, n is the n^{th} accidents, n_t is the total number of accidents observed r_i is the value of the segmented

$$Z_X = \sum_{i=1}^{n} a_i - \left[\frac{\sum_{i=1}^{n} a_i}{L_p}\right]\sum_{i=1}^{n_t} x_i \quad ; \quad a_i = \left[\frac{r_{i-1} + r_i}{2}\right]x_i$$

characteristic of the accidents and L_p is the total length of the section.

Geometrically, Z_X represents the difference between the cumulative area until the measure point r_i and the mean cumulative area within the same length. This can be observed in Figure-3. The segment edges are defined by the points where the sign of the slope of Z_X changes.

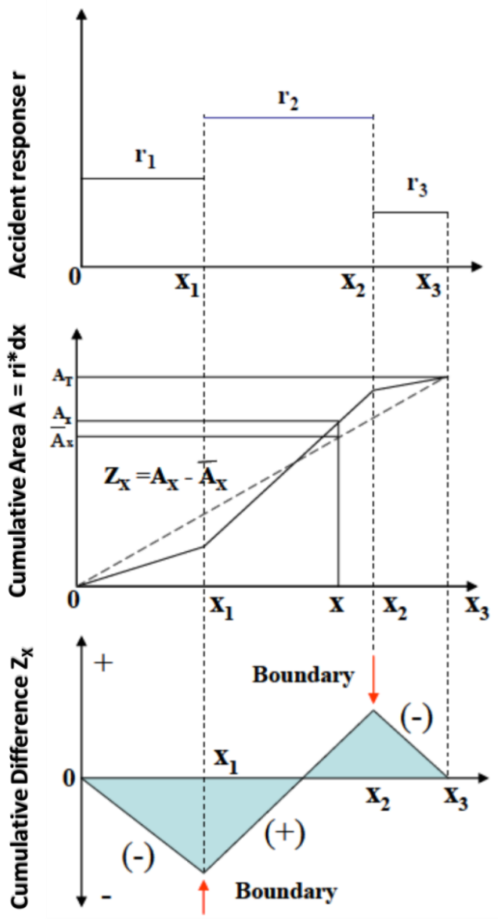


Figure -3: Graphic representation of the CDA method

3.3 Multiple linear regression Analysis

In case of prediction of the value of dependent variable (accident rate), first we identified variables and then to find out random sample n-size for the chosen values of dependent variables.

Suppose that k phenomenon is identified as independent variable (predictor), or X_i, i = 1, 2... k, and Y as dependent random variable. The whole multiple linear model can be presented as one equation for voluntarily dependent variable Y_i,

$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + \varepsilon_i$ ... (i)

Where:

Y_i - dependent random variable,

x_1, x_2, x_k - values of independent variable,
β_0, β_1, ., β_k- model parameters (regression coefficient)
ε_i - a supporting element, or a random error which has normal distribution, zero mean and constant variance.
Multiple linear regression model (i) consists of two parts:
− determined (Y_i')
$Y_i' = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$.......... (ii)
− stochastic (ε_i), so that from (i) we can get:
$\varepsilon_i = Y_i - Y_i'$ .... (iii)
Determined part of the linear regression model is an average value of dependent variable (Y_i) for the given values of independent variables:
$Y_i' = E(Y_i)$
$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$ …….(iv)
and other values of Y_i show average values $E(Y_i)$.
The whole regression model (i) was estimated by the sample regression model:
The whole regression model (i) was estimated by the sample regression model:
$\hat{y}_i = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$ ...,...... (v)
Where we have:
$\hat{y}_i$ − adjustable or foreseen value of dependent variable Y_i,
$x_1, x_2, \ldots, x_k$– values of independent variables,
$b_0, b_1 \ldots b_k$ – estimations of unknown parameters $\beta_0, \beta_1, \ldots, \beta_k$.

We should choose the multiple linear regression model which presents in the most suitable way the relationship between observed phenomena. It can be achieved by minimizing a sum of square equations of empirical points from the regression model (for example: regression plane when k=2), or:
$\sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \min$.................. (vi)
Where, ei – is random error in sample.

Multiple linear regression model as statistical model does not mean only mathematical expression but also assumptions which supply the optimal estimation of parameters $\beta_0, \beta_1, \ldots, \beta_k$. These assumptions are usually connected with random error:
− the random error has normal distribution,
− it is equal zero (on the average)
− supporting elements have equal variances.
In case when k=2, multiple linear regression model is regression plane equation in samples (the easiest example of multiple linear regression):
$\hat{y}_i = b_0 + b_1 x_1 + b_2 x_2$ .... (vii)

In order to establish adaptation of the estimated regression model by empirical data, we use standard error of the sample regression which represents the estimation of standard deviation of the random error σ_ε. It is market by S_ε, and it is presented as square root of repetition, or:
$S\varepsilon = \sqrt{\sigma^2} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-k-1}} = \sqrt{\frac{SSE}{n-k-1}}$ ... (viii)
Where, SSE is a sum of square root aberration of the empirical points of regression model (Error Sum of Squares).

The standard error of regression as absolute measure of the unexplained variability is not convenient for comparison. That's the reason why we use relative indicator − coefficient of multiple determination R^2. It is presented as a measure of explained variability and it is calculated by this equation:
$R^2 = \sqrt{\frac{\sum (\hat{y}_i - \bar{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}} = \frac{SSR}{SSy}$ .... (ix)
Where, SSR presents Regression Sum of Squares (explained variability) and SSy presents the total Sum of Squares (total variability).

The coefficient of multiple determinations shows the percentage of variations of dependent variable Y which is described by common influence of independent variables which are involved in this model. During its calculation we should take care of the number of independent variables and of sample size. It is achieved by calculation of the adjusted coefficient of multiple determinations:
$R^2_{adj.} = 1 - \frac{n-1}{n-k-1} . (1 - R^2)$ .... (x)
Where: n − is the sample size and k − number of independent variables.

- Model Usage Testing

In order to use the estimated regression equation we firstly have to test the significance of given estimates.
This is zero and alternative hypothesis:
H_0: $\beta_0 = \beta_1 = \beta_2 = \ldots + \beta_k = 0$

H_A: at least one $\beta_i \neq 0$
According to this, we have laid, zero hypothesis in that way that a linear connection between observed phenomena variations does not exist, or that $x_1, x_2, \ldots, x_k$ has not influence on Y.
If we start from the assumption that the total variability of dependent variable is conditioned by the variability of independent variables involved in the model and by the unexplained variability, we can write:
SSy = SSR + SSE................................ (xi)
Where, SSy − presents the total Sum of Squares (total variability),
SSR − Regression Sum of Squares (explained variability),
SSE − Error Sum of Squares (unexplained variability),
We apply F-test, and test and test the possibility of the regression model usage by analysis of variance. The table of this analysis is presented here:

Sources of variation	Degrees of freedom	Sum of squares	Mean squares	F- ratio
Regression	k	SSR	$MSR = \frac{SSR}{k}$	$F = \frac{MSR}{MSE}$
Error	$n\text{-}k\text{-}1$	SSE	$MSE = \frac{SSE}{n-k-1}$	
Total	$n\text{-}1$	SSy		

The decisions rules:
if $F \geq F_{\alpha;k;n-k-1}$, we reject null hypothesis,
if $F < F_{\alpha;k;n-k-1}$, we accept null hypothesis.
According to this, if the realized value of the F-test is lesser than theoretical, or we accept null hypothesis, we come to a conclusion that the linear influence of independent variables on dependent variable doesn't exist.

- Regression Coefficients Testing

If we use estimated regression model to estimate and predict the values of dependent variable Y, we have to test the significance of estimates of each parameter apart (β_i, i=1, 2, … k).
In the easiest case of the multiple linear regression k = 2 we test the estimate significance of two parameters β_1 and β_2. The null and alternative hypotheses are presented here:

I	$H_0: \beta_1 = 0$	II	$H_0: \beta_2 = 0$
	$H_A: \beta_1 \neq 0$		$H_A: \beta_2 \neq 0$

Test statistics:

$$t_1 = \frac{b_1}{S_{b_1}}$$

$$t_2 = \frac{b_2}{S_{b_2}}$$

Has the t-distribution with n − k − 1 degrees of freedom.
If $|t_i| < t_{\alpha/2,\ i=1,\ 2,}$ the value of the test statistics has fallen in the field of accepting null hypothesis. In that case we accept null hypothesis that independent variables (X_1, i.e. X_2) does not influence dependent variable Y.
Generally in multiple regression model we apply testing:
$H_0: \beta_i = 0$
$H_A: \beta_i \neq 0$
(for i = 1, 2, …, k); the test statistics

$$t_i = \frac{b_i}{S_{b_i}}$$

has the t-distribution with degrees of freedom n − k − 1. We accept null hypothesis if $|t_i| < t_{\alpha/2}$.
We can also mention that when testing the significance greater realized value of the t-test statistics does not mean that the variable which corresponds has greater relative influence on dependent variable.
In order to determine mathematical model of multiple regression and to test it and we have to do many calculations.

For these estimates to be acceptable it is necessary to test the hypothesis that the value computed for each regression coefficient is unlikely to have arisen by chance. To check this, the standard error of each regression coefficient is computed and tested for significance at the 5% level, variables with non-significant co-efficients has been eliminated from the analysis.

The computer program (Microsoft Excel) is used in the multiple regression analysis and manually eliminate the non-significant variables and test such variables with other combinations and replace them where necessary. This technique is known as 'stepwise' regression analysis.

3.4 Model Error Estimates

Mean Absolute Deviation (MAD): This criterion has been proposed by Oh et al. (Oh et al., 2003) to evaluate

$$MAD = \frac{\sum |Actual - Forecast|}{n}$$

the fit of models.
MAD closer to zero value is considered to be best among all the available models.

Mean Absolute Percentage Error (MAPE): Lower MAPE values are better because they indicate that smaller percentages errors are produced by the forecasting model. The following interpretation of MAPE values was

$$MAPE = \frac{\frac{\sum |Actual - Forecast|}{Actual} \times 100}{n}$$

suggested by Lewis (1982) as follows: MAPE is less than 10% is highly accurate forecasting, 10% to 20% is good forecasting, 21% to 50% is reasonable forecasting and 51% and above is inaccurate forecasting.

IV. Data Collections

The Study Highway traverses through two districts viz. Dharwad and Belgaum. Belgaum is located at 15.87°N and 74.5°E. It has average elevation of 751 meters above sea level. The Highway passes through plain and rolling terrain in most of its length. However, it traverses through hilly terrain in some stretches. This National Highway No. 4 runs between Chennai and Mumbai. In the Chennai-Bangalore section km 0.0 is at Chennai and increases towards Bangalore. Between Bangalore and Mumbai km 0.0 is at Bangalore and increases towards Mumbai.
Following information have been extracted from site and various reports section wise of selected stretch of highway (see Table-1, 2 & 3).

1) Road accidents (2008-2011).
2) The annual average daily traffic (2010).
3) Section length (m)
4) Average Horizontal Curvature (degrees/km)
5) Average Vertical Curvature (m/km)
6) Average Sight Distances (m)
7) Average carriageway width (m)
8) Road Roughness (mm/km)
9) Service Road (%)

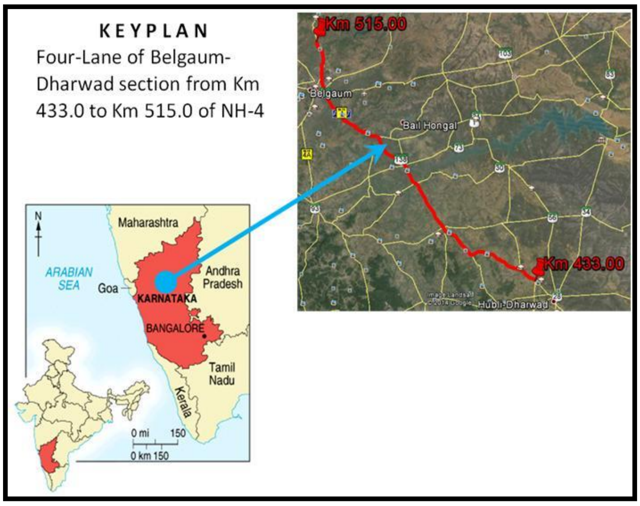


Figure -4: Key Plan

4.1 Site Selection

The study stretch of Project Highway is situated between km 433.00 (end of Hubli-Dharwad bypass) and km 515.00 (after crossing Belgaum city and Kakati Industrial area) of Bangalore-Mumbai section of NH-4 as presented in Figure-4. The total length of the Project Highway is about 82.00 km.

4.2 Traffic Volume (AADT)

The data (primary and secondary) collected as per IRC SP 19-2001 and has been analysed to obtain information on ADT, Seasonal Variation and AADT. AADT was taken about 10258 (Year 2010) throughout the section for this study.

4.3 Road Accident Rates (A_R)

All accidents data used in this study were collected over the period of January 2008 to December 2011 for Belgaum - Dharwad section of NH-4. The accident rates per 100 million vehicle-Km of travel- year have been calculated section wise.

4.4 Horizontal Curvature (H_C)

The horizontal curvature was measured by the average degree of Curvature per Kilometre as shown in Figure-5.

Average curvature of section AB,

$$= \frac{\emptyset 1+\emptyset 2+\emptyset 3+\cdots+\emptyset n}{\text{Distance AB (km)}} \text{ (expressed as degrees/km)}$$

4.5 Vertical Curvature (V_C)

The vertical curvature of the road was measured in terms of meters of Rise and Fall per Kilometre as shown in Figure-6.

Average rise and fall of section AB,

$$= \frac{h1+h2+h3+\cdots+hm}{\text{Distance AB (km)}} \text{ (expressed as m/km)}$$

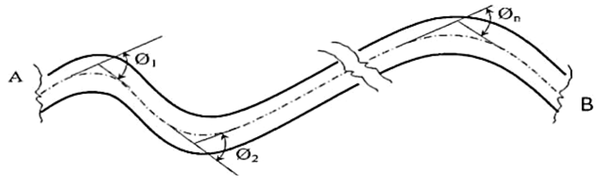


Figure- 5: Methods of Specifying the Curvature of a Road Section

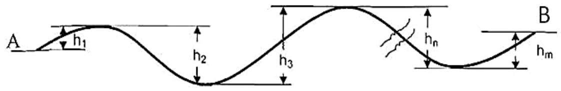


Figure-6: Methods of Specifying the Longitudinal Profile

4.6 Sight Distance (S_D)

Sight distance available from a point is the actual distance along the road surface, over which a driver from a

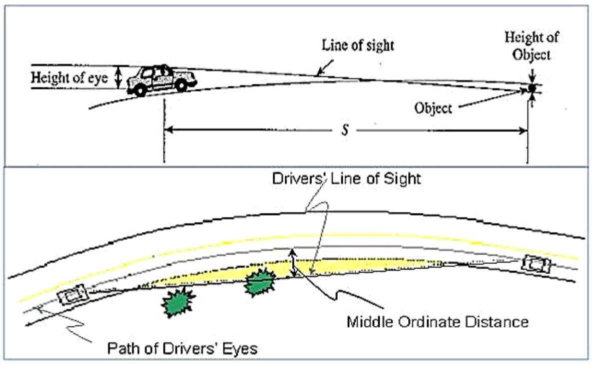


Figure-7: Sight Distance

SN	Sight Distance	Driver's Eye Height	Height of object
1	Stopping Sight Distance	1.2m	0.15m
2	Intermediate Sight Distance	1.2m	1.2m
3	Overtaking Sight Distance	1.2m	1.2m

specified height above the carriage way has visibility of stationary or moving objects.

Distance is the minimum sight distance available on a highway at any spot having sufficient length to enable the driver to stop a vehicle travelling at design speed, safely without collision with any other obstruction. The minimum ISD for four lane highway is 360m for the speed 100kmph. The intermediate sight distance calculated from existing ground surface along the alignment as per driver & object height as per IRC 73-1980 & IRC SP 23-1993.

Table 1: Homogeneous Sections

Sections	From KM	To KM	Total Length KM
1	433.0	442.5	9.5
2	442.5	451.0	8.5
3	451.0	455.0	4.0
4	455.0	457.0	2.0
5	457.0	465.0	8.0
6	465.0	476.0	11.0
7	476.0	488.0	12.0
8	488.0	494.0	6.0
9	494.0	501.0	7.0

Table 2: Accident Rates

Sec.	Stretches KM	Total Length KM	Accident Rates (Yr-2008)	Accident Rates (Yr-2009)	Accident Rates (Yr-2010)	Accident Rates (Yr-2011)	Accident Rates (Yr-2008-10)
1	433.0-442.5	9.5	92.776	81.530	92.776	87.153	89.027
2	442.5-451.0	8.5	25.137	56.559	31.421	28.279	47.132
3	451.0-455.0	4.0	20.031	13.354	26.708	13.354	20.031
4	455.0-457.0	2.0	40.062	93.479	40.062	40.062	57.868
5	457.0-465.0	8.0	13.354	20.031	13.354	13.354	15.580
6	465.0-476.0	11.0	60.700	70.412	70.412	58.272	67.175
7	476.0-488.0	12.0	22.257	35.611	48.965	15.580	35.611
8	488.0-494.0	6.0	80.125	48.965	57.868	62.319	62.319
9	494.0-501.0	7.0	87.755	49.601	57.232	34.339	64.863
10	501.0-515.0	14.0	47.693	62.955	51.509	53.416	54.052

4.7 Number of Junctions (J_N)

In Project Road, junctions are counted along both sides of the road, thus a crossroads would be counted as two junctions while a five-way road would counted as three junctions. The slip in & slip out also considered as a junction (access point). The expressed as in number of junction per Kilometre on a road section.

4.8 Carriageway width (C_W)

The existing pavement (excluding earthen shoulder) width is measured in meters and expressed in average width per kilometre of the section.

4.9 Road Roughness (R_I)

Roughness Index is used to define a characteristic of the longitudinal profile of a travelled wheel track and constitutes a standardized roughness measurement. The Roughness of the road is expressed in mm/km.

Table 3: Road design parameters

Section Number	Average Road Width (m)	Vertical Curvature (m/km)	Horizontal Curvature (deg/km)	Roughness Index (mm/km)	No of Junctions /km	Sight Distance (m)	Service Road %
1	17.82	15.67	32.30	1881	2.84	274	94%
2	17.93	22.83	43.39	2054	1.41	269	99%
3	18.03	6.77	10.45	1927	1.25	290	100%
4	18.00	22.60	56.82	2163	2.00	270	100%
5	14.95	14.69	33.89	2221	2.13	262	77%
6	15.13	12.89	20.95	1910	2.82	275	95%
7	15.23	8.94	23.85	1948	1.25	291	90%
8	16.21	29.27	56.79	2127	1.50	249	70%
9	14.34	17.17	25.22	1957	3.43	268	89%
10	15.75	13.27	14.29	2071	4.14	263	85%

4.10 Service Road (S_R)

The service road length measured in Km along the project road and expressed in percentage (%) with respect to project road.

V. Analysis

A multiple linear regression analysis was used to establish and quantify relationships between one dependent variable and one or more independent variables.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6 + b_7X_7$$

Where,
Y= dependent variable
X_1, X_2…..= independent variable
b_0 = regression constant
b_1, b_2…..= regression coefficient
For development of accident model, the dependent and independent variable are presented in Table-4.

Table 4: Dependent and Independent Variables

Sr. No.	A_R	R_W	V_C	H_C	$\frac{R_I}{R_W^2}$	J_N	S_D	S_R
Variables→	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	89.027	17.82	15.67	32.30	5.92	2.84	274	94%
2	47.132	17.93	22.83	43.39	6.39	1.41	269	99%
3	20.031	18.03	6.77	10.45	5.93	1.25	290	100%
4	57.868	18.00	22.60	56.82	6.68	2.00	270	100%
5	15.580	14.95	14.69	33.89	9.94	2.13	262	77%
6	67.175	15.13	12.89	20.95	8.34	2.82	275	95%
7	35.611	15.23	8.94	23.85	8.40	1.25	291	90%
8	62.319	16.21	29.27	56.79	8.09	1.50	249	70%
9	64.863	14.34	17.17	25.22	9.51	3.43	268	89%

VI. Result & Validations

6.1 Result

From regression analysis, an equation was derived, which related accidents rate per 100 million vehicle-kilometres per year to the road design parameters. The results obtained in analysis show how various features of the road considered separately are related to the accident rates.
The regression equation of factors related to the accident rate is as follows:

$$\mathbf{A_R = 867.7745 - 35.7649\, R_W + 1.229\, H_C - 40.4545\, \frac{R_I}{R_W^2} + 19.1260\, J_N} \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{Eq.-1}$$

Where,
A_R= Accident Rate per 100 million vehicle –Km-year
R_W= Average Road Width (m)
H_C= Horizontal Curvature (deg/km)
R_I= Roughness Index (mm/km)
J_N= No. of Junctions/km

Model Summary				
Model	Multiple R	R Square	Adjusted R Square	Standard Error
1	0.993632	0.9873	0.974607456	3.796329

For the goodness of fit, R2 is 0.9873 and adjusted R^2 is 0.9746 is large, it proves that the regressed model does provide a very good fit to the independent variable. R^2, that is the amount of variation in the accident rates accounted for by all the independent variables is 98.73% with probability value of 0.00048 and probability values (P-values) of independent variables are less than alpha (0.05) so the association is statistically significance at 5.0% level.

ANOVA						
Model		df	SS	MS	F	Significance F
1	Regression	4	4482.931921	1120.73298	77.76327769	0.000479493
	Residual	4	57.64844351	14.41211088		
	Total	8	4540.580365			

Coefficients					
Model		Coefficients	Standard Error	t Stat	P-value
1	Constant	867.7744933	89.94501546	9.647833055	0.00064558
	$X_{1} = (R_W)$	-35.76493816	3.86388277	-9.2562172	0.000757457
	$X_{3} = (H_C)$	1.229043852	0.110890452	11.08340554	0.00037692
	$X_4 = \left(\frac{R_I}{R_W^2}\right)$	-40.45447581	3.683297496	-10.98322247	0.000390485
	$X_{5} = (J_N)$	19.12596449	1.897857139	10.07766291	0.000545416

6.2 Validation of model

This section presents the validation procedure for accident prediction model for highway. This method evaluates the accuracy of accident rates calculated from equation by analysis and the differences in observed and predicted values. These two separate aspects have been considered for model validation.

1) Validation of the models against additional years of accident data for the same stretch used in the prediction.

This validation is used to assess the models' ability to forecast accidents across time. The accident data has been collected from site in same method on study section for year-2011 and calculated the accident rates from Equation -1 for the same highway from Km 433.00 to Km 515.00 Km. Predicted accident rates and observed accident rates are tabulated in Table-5 and the mean absolute percentage error (MAPE) between observed value and predicted is about 45% & mean absolute deviation (MAD) is low. Also plot the graph between observed accident rates and predicted accident rates and presented in Figure-8 and the most of points are fairly closed to 1:1 line

Table 5: Model predictive ability for Year 2011

Sr. No.	**Stretches km**	$\mathbf{R_W}$	$\mathbf{H_C}$	$\mathbf{\frac{R_I}{R_W^2}}$	$\mathbf{J_N}$	**Predicted Accident Rates**	**Observed Accident Rates - 2011**	**MAD**	**MAPE**
1	433.0-442.5	17.82	32.30	5.92	2.84	84.874	87.153	12.81	45%
2	442.5-451.0	17.93	43.39	6.39	1.41	48.448	28.279		
3	451.0-455.0	18.03	10.45	5.93	1.25	19.935	13.354		
4	455.0-457.0	18.00	56.82	6.68	2.00	62.058	40.062		
5	457.0-465.0	14.95	33.89	9.94	2.13	13.397	13.354		
6	465.0-476.0	15.13	20.95	8.34	2.82	68.728	58.272		
7	476.0-488.0	15.23	23.85	8.40	1.25	36.574	15.580		
8	488.0-494.0	16.21	56.79	8.09	1.50	59.100	62.319		
9	494.0-501.0	14.34	25.22	9.51	3.43	66.484	34.339		
10	501.0-515.0	15.75	14.29	8.34	4.14	63.588	53.416		

2) Validation of the models against additional segment of same highway.

The most common way to validate the model is through cross-validation. This validation exercise is used to assess the models' ability to forecast crashes over a jurisdiction whose data were not employed in model development. Data from Km 501.00 to Km 515.00 are used to cross validation of predictive ability of model. The comparison of results between predicted accident rates from Equation -1, and observed accident rates have been presented in Table-6 and in Figure-9 for section from Km 501.00 to Km 515.00 on same highways in year 2008 to year 2011 and the variation in observed accident rates and predicted accident rates are low and results are close to observed accident rates.

Table 6: Comparison of Results on additional segment of road

Years	Stretches km	R_W	H_C	$\frac{R_I}{R_W^2}$	J_N	Predicted Accident Rates	Observed Accident Rates	Variation %
2008	501.0-515.0	15.75	14.29	8.34	4.14	63.588	47.693	-33%
2009	501.0-515.0	15.75	14.29	8.34	4.14	63.588	62.955	-1%
2010	501.0-515.0	15.75	14.29	8.34	4.14	63.588	51.509	-23%
2011	501.0-515.0	15.75	14.29	8.34	4.14	63.588	53.416	-19%

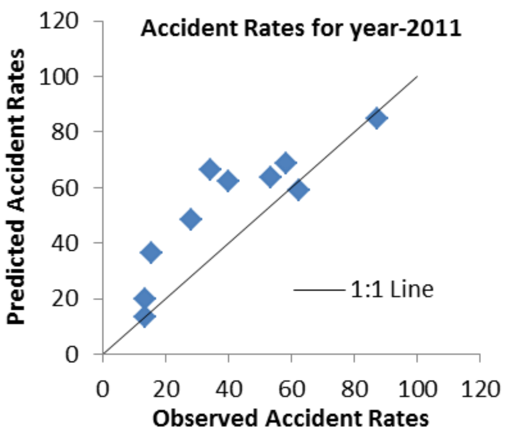


Figure-8: Predicted Vs Observed Accident Rates

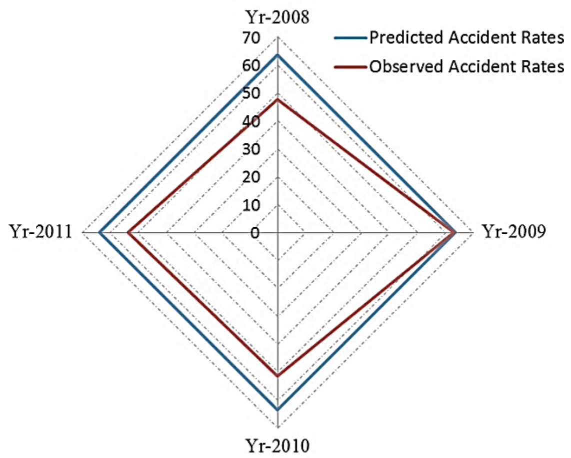


Figure-9: Year wise model comparison on additional segment of Highway

VII. Conclusions And Suggestions For Further Research

This study developed an accident prediction model to analyse the accident rates of highways. A multiple linear regression model was applied to the dataset which includes the historical accident data, highway geometrics & facilities data for the homogeneous highway segments.

The developed model has been related to road design parameters such as presence of vertical curvature, sight distance and presence of current service roads are rarely correlated to the accidents, road width, road roughness divided by square of road width are negatively correlated to the accidents and horizontal curvature & number of junctions are positively correlated to the accidents in this 4-lane highway. The accident rates were found to be significantly (5.0% level) related to road design parameters of study highway, such as carriageway width, horizontal curvatures, road roughness index & junction (Entry & Exist).

The road accident is a random phenomenon and it is very difficult to do the exact prediction of future trends of accidents by using any model or theory, but developed model is helpful for the design of safe highways. Additionally, it will contribute to identifying the potentially hazardous locations on highways and to the treatment of safety improvements.

There exists a limitation in this model due to the fact that human factors are not considered in the development of the prediction model. Accordingly, driver characteristics and driving behaviour due to the change of the traffic environment are not included in the developed model in this study. The further research should concentrate on the following:

1. Validation of the models on other four-lane highways.

2. Development of multiple linear regression accident prediction models based for other types of highway facilities (e.g., expressway, intersections).
3. Development of multiple linear regression accident prediction models increase number of datasets.
4. Development of accident-prediction models that account for human behaviour and vehicle factors in addition to road design parameters.

Acknowledgements

The author would like to thank National Highway Authority of India (NHAI) and my friends for carrying out the data extraction.

References

[1]. Chikkakrishna N K, Parida M, Jain S S (2013) - "Crash Prediction for Multilane Highway Stretch in India", Proceedings of the Eastern Asia Society for Transportation Studies, Vol.9.
[2]. Singh R K & Suman S K (2012) - "Accident Analysis and Prediction of Model on National Highways"- International Journal of Advanced Technology in Civil Engineering, Volume-1(2).
[3]. Rokade S, Singh K, Katiyar S.K and Gupta S (2010) - "Development of Accident Prediction Model", International Journal of Advanced Engineering Technology Volume-1(3).
[4]. Jacobs G D (1976) – "A Study of Accident Rates on Rural Roads in Developing Countries", Overseas Unit Transport and Road Research Laboratory Crowthorne Berkshire United Kingdom.
[5]. Mustakim F & Fujita M (2011) - "Development of Accident Predictive Model for Rural Roadway"- World Academy of Science, Engineering and Technology 58.
[6]. Mayora J M P & Rubio R L (2003) - "Relevant Variables for Crash Rate Prediction in Spain´s Two Lane Rural Roads"- Transportation Research Board, 82nd Annual Meeting.
[7]. PIARC Road Accident Investigation Guidelines For Road Engineers- August 2007
[8]. AP-R162- Relationship between Crash Risk And Geometric Characteristics Of Rural Highways
[9]. Farah H, Polus A and Cohen M A- Multivariate Analyses for Infrastructure-Based Crash- Prediction Models for Rural Highways
[10]. Koorey G F- Incorporating Safety Into Rural Highway Design- Department of Civil & Natural Resources Engineering University Of Canterbury
[11]. Tejeda S V, Tampier H de S & Navarro T E (2009)-Proposal of A Segmentation Procedure for Skid Resistance Data, The Arabian Journal for Science and Engineering, Volume 33 (1B).
[12]. IRC:38-1988- Guidelines for Design of Horizontal Curves for Highways and Design Tables (First Revision)
[13]. IRC:53-1982- Road Accident Forms A-1 and 4 (First Revision)
[14]. IRC:73-1980- Geometric Design Standards for Rural (Non-Urban) Highways
[15]. IRC:108-1996- Guidelines for Traffic Prediction on Rural Highways
[16]. IRC:SP:19-2001- Manual for Survey, Investigation and Preparation of Road Projects (Second Revision)
[17]. IRC:SP:23-1983- Vertical Curves for Highways
[18]. IRC:SP:30-1993- Manual on Economic Evaluation of Highway Projects in India (First Revision)
[19]. IRC:SP:16-2004- Guidelines for Surface Evenness of Highway Pavements (First Revision)
[20]. NHAI web site: http://www.nhai.org/
[21]. http://www.rtirn.net/docs/WDR_2012.pdf
[22]. Ministry of Road Transport and Highways web site: http://morth.nic.in/
[23]. Hines W. and Montgomery D., - Probability and Statistics in Engineering and Management Science, John Wiley and Sons, New Delhi, 1990.
[24]. Orlov M L, "Multiple Linear Regression Analysis Using Microsoft Excel", Chemistry Department, Oregon State University (1996).
[25]. Ofor T, Ackah B and Ephraim L "Statistical Models for Forecasting Road Accident Injuries in Ghana" University of Mines and Technology, Tarkwa, Ghana