DEVELOPMENT OF ACCIDENT PREDICTION MODELS FOR KHULNA METROPOLITAN CITY

Junayed Pasha*[1] and Quazi S Hossain[2]

[1] *Student, Khulna University of Engineering & Technology, Bangladesh, e-mail: junayed729@gmail.com*
[2] *Professor, Khulna University of Engineering & Technology, Bangladesh, e-mail: sazzad1999@yahoo.com*

ABSTRACT

This study involves various aspects of road accidents in Khulna Metropolitan City. In this study, road accident data for the five police stations of this city were collected for the years 2000 to 2014. The data were analyzed to understand the nature of road accidents in Khulna Metropolitan City. Smeed's and Andreassen's models were evaluated by regression analysis. It was found that Smeed's model is applicable for Khulna Metropolitan City whereas Andreassen's model is not. Finally, an alternative model was developed to understand the extent of the causes of road accidents, using the concept of Smeed's model.

Keywords: *Road accident, Smeed, Andreassen, Khulna Metropolitan City*

1. INTRODUCTION

Road traffic crash is an event on the public road involving a vehicle and resulting into harm to people and damage to properties. Road traffic accidents and their resulting fatalities may be regarded as a growing social and economic problem, especially in developing countries where the resources are limited. The World Health Organization has predicted that traffic fatalities will be the third leading cause of death worldwide by the year 2020 (Murray & Lopez, 1994).

At present for large number of accidents, it is so important to develop a road accident model considering the basic factors. Accident prediction models are invaluable tools that have many applications in road safety analysis. However, there are certain statistical issues related to accident modelling that either deserve further attention or have not been dealt with adequately in the road safety literatures.

Bangladesh is a developing country of South Asia. It consists of an area of 147,570 square kilometers, containing a high population density of 1,033.5 per square kilometer. Khulna is the 3rd largest city in Bangladesh with an area of 45.65 square kilometres. 1,500,689 people live in this city. The population density is 67,994 per square kilometer in the city. There are 1215 roads, totalling a length of 356.64 kilometers (Khulna City Corporation, 2015). For this city, very few research works have been carried out with a detailed spectrum of analysis of road accident models for a full understanding of accident problems and thereby developing effective countermeasures. Loss of lives and property damages are expected to continue if suitable measures are not taken. Therefore, it is necessary to develop road accident models for Khulna Metropolitan City.

The main objectives of this study are:

1. To collect the accident data of the five police stations of Khulna Metropolitan City.
2. To evaluate Smeed's and Andreassen's models for this city through analysing the collected data.
3. To develop an alternative road accident model by regression analysis.

2. LITERATURE REVIEW

Before carrying out the analysis and modelling of accidents, it is required to conduct a review of literature on the related aspects. In this regard, the following road accident models were studied.

Based on the study of data from twenty countries, Smeed (1968) approximated the per vehicle fatality rate vis-à-vis the per capita vehicle ownership model as follows:

$$F/N = 0.0003\ (N/P)^{-0.67} \quad (1)$$

Where, F = number of road traffic fatalities; N = number of vehicles and P = population. While developing this accident model, Smeed took the number of fatalities as dependent variable and the number of motorized vehicles and population as independent variables.

Later on, other authors tried to validate or update the model based on newer data. The model was found to be valid with some changes in parameters (Adams, 1987). Fortunately, the increasing trend of the total number of fatalities started to change towards a decreasing trend in some countries from the 60's. For the UK, Smeed's prediction was moving correctly and had approximately the right magnitude until about 1966. Since 1966, Smeed's prediction has continued to rise, while the real road fatalities have fallen quite reliably.

Valli (2005) made an attempt to develop relationships among the parameters, namely, road accidents, the number of registered motor vehicles and population. The data on road accidents in terms of total number of accidents (A), fatalities (F), injuries (I) and the number of registered motor vehicles (N) were used from road accidents in India. The model developed by Smeed does not however, explain the discrepancy for the data on India, as the data do not a give good fit to the model. It was, therefore, proposed to investigate the data and develop appropriate models based on the concept of Smeed's and Andreassen's model. The models developed by Valli and Sarkar (1997) relating fatalities (F) to vehicles (N) and population (P) in India for the years 1960 to 1991 is as follows:

$$F/N = 0.00073\ (N/P)^{-0.4} \quad (2)$$

The parameter's coefficient i.e, -0.4 in equation (2) is not the same as that of Smeed's model. The constant term i.e, 0.00073 also differs from Smeed's model i.e, 0.0003. The regression analysis by Valli (2005) was carried out using Smeed's model for the years 1970 to 2001 for India. The equations are given below:

$$A/N = 0.0008\ (N/P)^{-0.75} \quad (3)$$
$$F/N = 0.0003\ (N/P)^{-0.58} \quad (4)$$
$$I/N = 0.0014\ (N/P)^{-0.57} \quad (5)$$

Andreassen (1985) produced the following relationship:

$$A = Constant * (N)^{M1} * (P)^{M2} \quad (6)$$
$$F = Constant * (N)^{M1} * (P)^{M2} \quad (7)$$
$$I = Constant * (N)^{M1} * (P)^{M2} \quad (8)$$

Where, M1 and M2 are model parameters. Other symbols bear their usual meanings.

The models describing the changes in road fatalities have used other parameters, such as vehicle-kilometers travelled and Gross Domestic Product (GDP). The research carried out by Oppe (1991) found that the long-term development of traffic fatalities in the highly motorized countries follows a law-like pattern determined by the growth of motorization and the decline of the fatality rate per vehicle-kilometers of driving.

The change from the increasing to the decreasing trend could be observed in several countries. Kopits and Cropper (2005) have found that the income level at which traffic fatality risk (F/P) first declines is $8600 (1985 international prices), regardless of how the time trends are specified. This is the approximate income level attained by countries such as Belgium, the United Kingdom, and Austria in the early 1970s, South Korea in 1994, and New Zealand in 1968.

Akgungor and Dogan (2009) stated that the increase in the number of traffic accidents is related to economic and social outcomes. They advised the Turkish force authorities to develop new alternative transportation policies for reducing traffic accidents. In Turkey, from 2006 to 2008, approximately 2.5 million traffic accidents were recorded and nearly 14,000 people lost their lives and more than 541,000 people got injured.

Akgungor and Dogan (2009) obtained different exponential values i.e, B1 and B2 in equation (9) for Germany, France, England, USA etc.

$$F = Constant * (N)^{B1} * (P)^{B2} \quad (9)$$

The symbols bear their usual meanings in this equation.

3. METHODOLOGY

3.1 Data Collection

The study area is under the five police stations of Khulna Metropolitan City, namely, Khulna Sadar, Daulatpur, Khalishpur, Khanjahan Ali and Sonadanga. Road accident data for the years 2000 to 2014 were collected from the Head Quarter of Khulna Metropolitan Police (KMP), Accident Research Institute of Bangladesh University of Engineering and Technology (BUET). The numbers of registered vehicles for each year were collected from Bangladesh Road Transport Authority (BRTA). The yearly populations of the areas under the five police stations were collected from the population census performed by Bangladesh Bureau of Statistics (BBS, 2011).

3.2 Evaluation of Existing Models

Based on the collected data, Smeed's and Andreassen's models were available to use. These models were evaluated by regression analysis. The linear regression analysis was performed by using the software SPSSv16.

3.2.1 Smeed's Model

In the present study, to evaluate Smeed's model for the years 2000-2014 for Khulna Metropolitan City, the following equations were used:

$$A/N = C\,(N/P)^{M} \quad (10)$$
$$F/N = C\,(N/P)^{M} \quad (11)$$
$$I/N = C\,(N/P)^{M} \quad (12)$$

These equations can be written in the following form:

$$\ln(A/N) = \ln(C) + M \ln(N/P) \quad (13)$$
$$\ln(F/N) = \ln(C) + M \ln(N/P) \quad (14)$$
$$\ln(I/N) = \ln(C) + M \ln(N/P) \quad (15)$$

Where, A is the number of road accidents, F is the number of road traffic fatalities, I is the number of injuries, N is the number of registered vehicle and P is the yearly population. Here, M and C are model parameters.

3.2.2 Andreassen's Model

Smeed's analysis was heavily criticized by Andreassen for model accuracy. It was argued that Smeed's model could not be applied universally to all countries. The relationship that Andreassen produced is of the following form:

$$A = C * (N)^{M1} * (P)^{M2} \quad (16)$$
$$F = C * (N)^{M1} * (P)^{M2} \quad (17)$$
$$I = C * (N)^{M1} * (P)^{M2} \quad (18)$$

These equations can be written in the following form:

$$\ln(A) = \ln(C) + M1 \ln(N) + M2 \ln(P) \quad (19)$$
$$\ln(F) = \ln(C) + M1 \ln(N) + M2 \ln(P) \quad (20)$$
$$\ln(I) = \ln(C) + M1 \ln(N) + M2 \ln(P) \quad (21)$$

Where, M1, M2 and C are model parameters. Other symbols bear their usual meanings.

3.3 Development of an Alternative Road Accident Model

Based on Smeed's model, an attempt was made again to develop relationships among the parameters, namely, the number of road accidents, fatalities, injuries; the number of registered motor vehicles and population for Khulna Metropolitan City. The model parameters were estimated and the model was evaluated by inverse regression analysis.

The mathematical form of the developed model is as follows:

$$A/N = C + M\,(P/N) \quad (22)$$
$$F/N = C + M\,(P/N) \quad (23)$$

$$I/N = C + M\,(P/N) \quad (24)$$

Where, M and C are model parameters. Other symbols bear their usual meanings.

4. RESULTS AND DICUSSION

4.1 Data Analysis

Road accident data of Khulna Metropolitan City for the years 2000 to 2014 is shown in Table 1. It can be seen that the population has decreased in a significant rate. This may be due to the relocation of the metropolitan area. Moreover, a significant percentage of the population has moved to Dhaka for a better income opportunity.

Table 1: Road accident data for the years 2000 to 2014

Year	Population	Registered motor vehicles	Reported accidents	Reported fatalities	Reported injuries
2000	864247	9235	42	25	44
2001	884445	10015	40	30	25
2002	869679	10895	48	24	32
2003	855303	11655	31	18	20
2004	841305	12475	15	8	12
2005	827674	13605	18	16	19
2006	814400	14985	42	24	35
2007	801469	15815	67	48	30
2008	788873	17220	30	14	24
2009	776600	19030	41	33	19
2010	764643	21400	38	32	22
2011	751237	24250	29	27	7
2012	739923	27590	19	14	7
2013	728894	30710	21	22	9
2014	718141	34300	24	15	25

On the other hand, the number of registered vehicles has increased over the years. Moreover, the rate of increment has also increased over the passage of time. This is because of the improvement of the people's economic capacity and also the improvement of the road network. The improvement of road geometrics and cautiousness of the road users might also have contributed to that. However, the number of accidents, fatalities or injuries have always remained discrete as accidents never follow any specific pattern. Still, the growth of motorization, urbanization and hence number of road users contribute to that. There are many other influencing factors.

4.2 Evaluation of Smeed's Model

Table 2: Smeed's model for Khulna Metropolitan City

Case	Formulation	R^2	Standard error	Significance
Accidents	$A/N = e^{-10.733} * (N/P)^{-1.160}$	0.670	0.408	0.000
Fatalities	$F/N = e^{-10.227} * (N/P)^{-0.921}$	0.498	0.464	0.003
Injuries	$I/N = e^{-12.717} * (N/P)^{-1.539}$	0.717	0.486	0.000

As for modelling, Smeed's model was evaluated for Khulna Metropolitan City. It was evaluated for accidents, fatalities and injuries, using linear regression. Table 2 represents Smeed's model for accidents, fatalities and injuries for Khulna Metropolitan City. It is seen that this model satisfies the condition of 95% confidence level, as the significance values are less than 0.05. Moreover, the values of coefficient of determination (R^2) are near or over 0.5 for all the cases which is acceptable (Abojaradeh, 2013).

Table 3 shows the variation between actual and predicted values from Smeed's model. It is seen that the percentage deviation between these values are quite high.

Table 3: Comparison of actual and predicted values from Smeed's model for Khulna Metropolitan City

Year	Accidents			Fatalities			Injuries		
	Actual	Predicted	% deviation	Actual	Predicted	% deviation	Actual	Predicted	% deviation
2000	42	39	-7.14	25	22	-12.00	44	30	-31.82
2001	40	40	0.00	30	22	-26.67	25	30	20.00
2002	48	38	-20.83	24	22	-8.33	32	28	-12.50
2003	31	37	19.35	18	22	22.22	20	26	30.00
2004	15	36	140.00	8	22	175.00	12	24	100.00
2005	18	35	94.44	16	22	37.50	19	23	21.05
2006	42	34	-19.05	24	21	-12.50	35	21	-40.00
2007	67	33	-50.75	48	21	-56.25	30	20	-33.33
2008	30	32	6.67	14	21	50.00	24	19	-20.83
2009	41	31	-24.39	33	21	-36.36	19	17	-10.53
2010	38	30	-21.05	32	21	-34.38	22	16	-27.27
2011	29	28	-3.45	27	21	-22.22	7	14	100.00
2012	19	27	42.11	14	21	50.00	7	13	85.71
2013	21	26	23.81	22	21	-4.55	9	12	33.33
2014	24	25	4.17	15	20	33.33	25	11	-56.00

A comparison of actual and predicted values from Smeed's model in terms of accidents, fatalities and injuries are shown in Figures 1 to 3. It is observed that predicted and actual values are not perfectly well fitted. Still, the significance values are less than 0.05. So, it satisfies the model criteria.

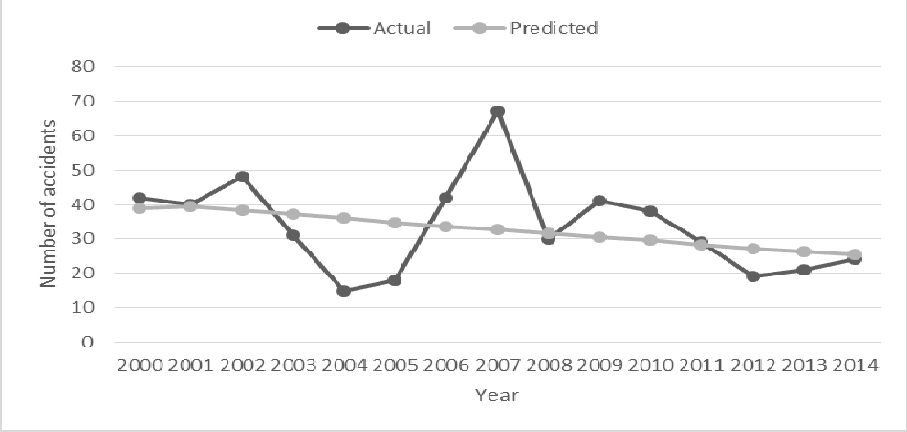

Figure 1: Comparison of predicted and actual values using Smeed's model for accidents

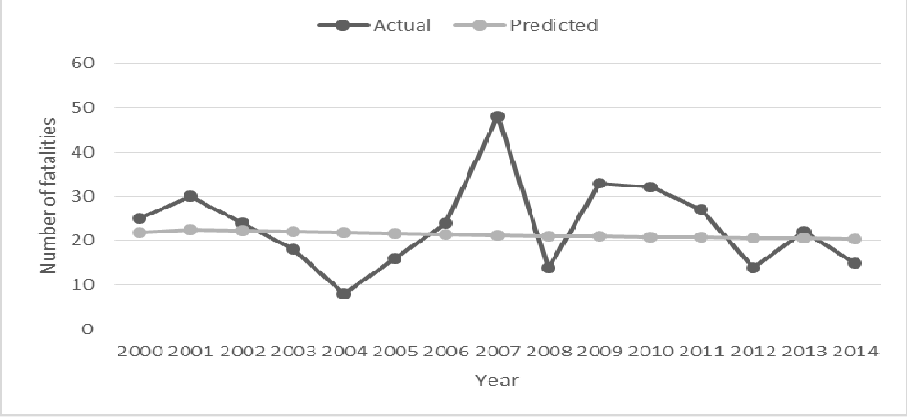

Figure 2: Comparison of predicted and actual values using Smeed's model for fatalities

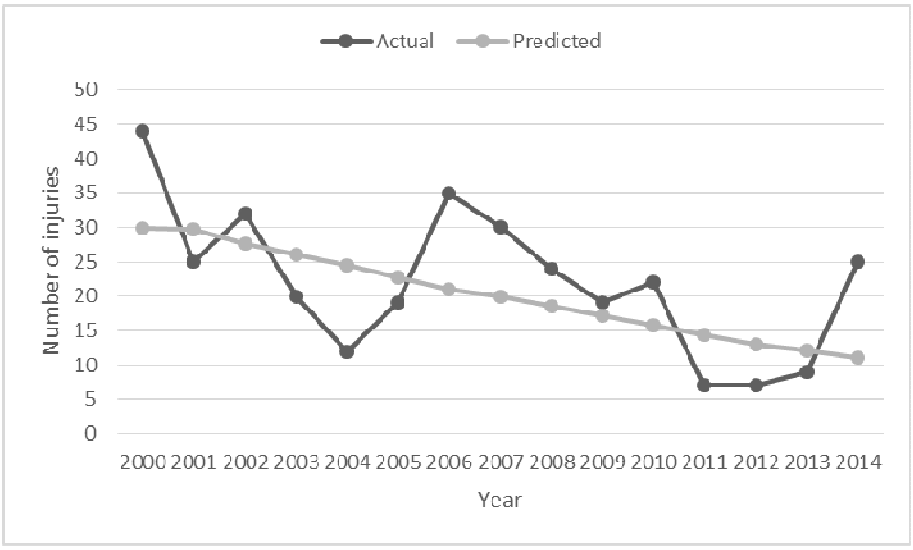

Figure 3: Comparison of predicted and actual values using Smeed's model for injuries

Based on the above analysis, it is found that the coefficients of Smeed's model are significant at 95% confidence level. Therefore, Smeed's model is applicable for Khulna Metropolitan City.

4.3 Evaluation of Andreassen's Model

Andreassen's model was also evaluated for Khulna Metropolitan City. Linear regression was performed for accidents, fatalities and injuries. Table 4 represents Andreassen's model for accidents, fatalities and injuries for Khulna Metropolitan City. It is seen that this model does not satisfy the condition of 95% confidence level, as the significance values are more than 0.05. Again, the values of coefficient of determination (R^2) are less than 0.5. Therefore, Andreassen's model is not applicable for Khulna Metropolitan City.

Table 4: Andreassen's model for Khulna Metropolitan City

Case	Formulation	R^2	Standard error	Significance
Accidents	$A = e^{147.530} * N^{-1.855} * P^{-9.273}$	0.169	0.411	0.329
Fatalities	$F = e^{150.509} * N^{-1.643} * P^{-9.672}$	0.048	0.470	0.744
Injuries	$I = e^{165.401} * N^{-2.446} * P^{-10.200}$	0.365	0.491	0.065

4.4 Development of an Alternative Road Accident Model

Based on Smeed's model, an attempt was made to develop an alternative road accident model for Khulna Metropolitan City. It was evaluated for accidents, fatalities and injuries, using inverse regression. Table 5 represents the developed model for accidents, fatalities and injuries for Khulna Metropolitan City. From the table, it is seen that this model satisfies the condition of 95% confidence level, as the significance values are less than 0.05. Moreover, the values of coefficient of determination (R^2) are near or over 0.5 for all the cases which is acceptable.

Table 5: Developed accident prediction model for Khulna Metropolitan City

Case	Formulation	R^2	Standard error	Significance
Accidents	$A = 4.675 * 10^{-5} (P)$	0.608	0.001	0.001
Fatalities	$F = 2.505 * 10^{-5} (P)$	0.457	0.001	0.006
Injuries	$I = 4.285 * 10^{-5} (P)$	0.707	0.001	0.000

Table 6 shows the variation between actual and predicted values from the developed model. It is seen that the percentage deviation between these values are quite high.

Table 6: Comparison of actual and predicted values from the developed model for Khulna Metropolitan City

Year	Accidents			Fatalities			Injuries		
	Actual	Predicted	% deviation	Actual	Predicted	% deviation	Actual	Predicted	% deviation
2000	42	40	-4.76	25	22	-12.00	44	37	-15.91
2001	40	41	2.50	30	22	-26.67	25	38	52.00
2002	48	41	-14.58	24	22	-8.33	32	37	15.63
2003	31	40	29.03	18	21	16.67	20	37	85.00
2004	15	39	160.00	8	21	162.50	12	36	200.00
2005	18	39	116.67	16	21	31.25	19	35	84.21
2006	42	38	-9.52	24	20	-16.67	35	35	0.00
2007	67	37	-44.78	48	20	-58.33	30	34	13.33
2008	30	37	23.33	14	20	42.86	24	34	41.67
2009	41	36	-12.20	33	19	-42.42	19	33	73.68
2010	38	36	-5.26	32	19	-40.63	22	33	50.00
2011	29	35	20.69	27	19	-29.63	7	32	357.14
2012	19	35	84.21	14	19	35.71	7	32	357.14
2013	21	34	61.90	22	18	-18.18	9	31	244.44
2014	24	34	41.67	15	18	20.00	25	31	24.00

A comparison of actual and predicted values from the developed model in terms of accidents, fatalities and injuries are shown in Figures 4 to 6. It is observed that predicted and actual values are not perfectly well fitted. Still, the significance values are less than 0.05. So, it satisfies the model criteria.

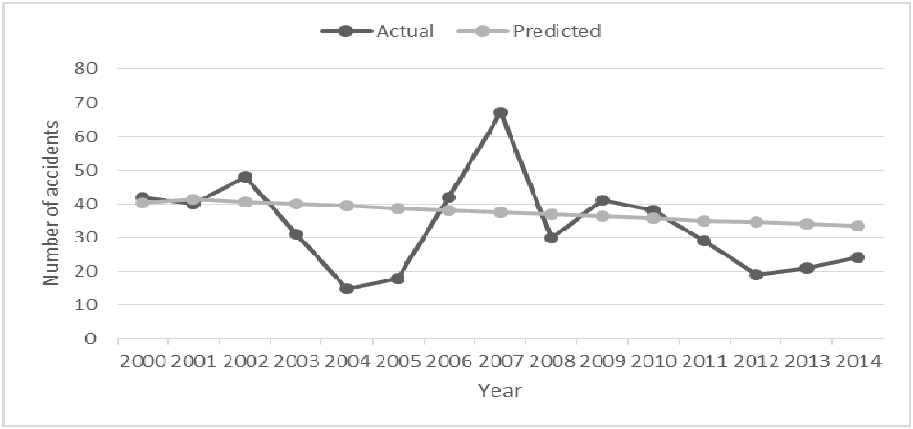

Figure 4: Comparison of predicted and actual values using the developed model for accidents

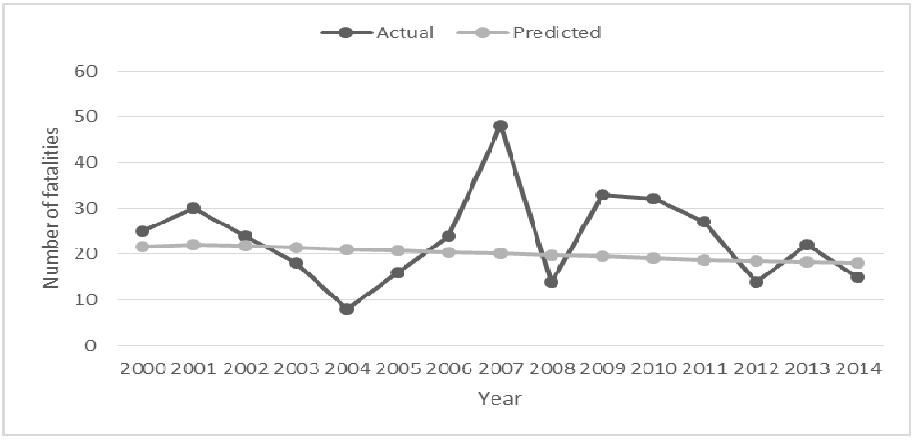

Figure 5: Comparison of predicted and actual values using the developed model for fatalities

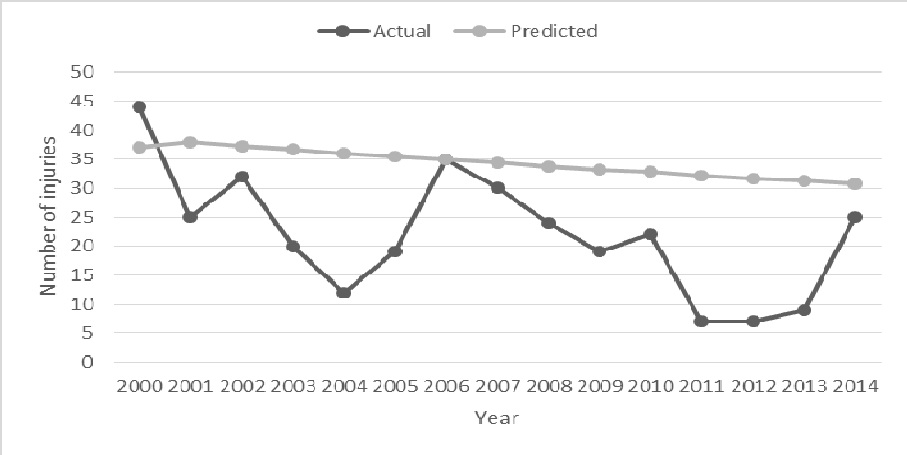

Figure 6: Comparison of predicted and actual values using the developed model for injuries

Based on the above analysis, it is found that the coefficients of the developed model are significant at 95% confidence level. Therefore, the developed accident prediction model is applicable for Khulna Metropolitan City.

5. CONCLUSIONS

The data on road traffic accidents in Khulna Metropolitan city are not satisfactory. Police records are the only source of information on road accidents, although there is substantial under-reporting as many accidents are settled privately. However, the fewer data on accident reports at police stations are an indicator of lack of awareness of accident reporting. Police have no incentive to collect data accurately and there is no standard accident-reporting format. Based on the police data, it is not possible to make correct models and to implement safety measures.

The analysis based on Smeed's equations showed that the significance values were less than 0.05. Therefore, this model can be accepted for Khulna Metropolitan City. On the other hand, the analysis based on Andreassen's equations showed that the significance values were greater than 0.05. Therefore, this model cannot be accepted for Khulna Metropolitan City. At last, by analysing the data and using the concept of Smeed's model, an alternative model was developed which showed that the significance values were less than 0.05 Therefore, this model can also be accepted for Khulna Metropolitan City. Moreover, the values of standard error of this model were significantly less than those of Smeed's model (Table 2 & 5).

The data collected for selected police stations could not be fitted perfectly to express the accident prediction models in the best way. Still, it was tried to develop the closest fitting models based on the available data. A further attempt should be made for the prediction of road accidents in Khulna Metropolitan City for the future years.

REFERENCES

Abojaradeh, M. (2013). Traffic accidents prediction models to improve traffic safety in greater Amman area. *Civil and Environmental Research*, 3(2), 87-101.

Adams, J.G.U. (1987). Smeed's law: some further thoughts. *Traffic Engineering & Control*, 28(2), 70-73.

Akgungor, A.P. & Dogan, E. (2009). An artificial intelligent approach to traffic accident estimation: model development and application. *Transport*, 24(2), 135-142.

Andreassen, D.C. (1985). Linking deaths with vehicles and population. *Traffic Engineering & Control*, 26(11), 547–549.

Bangladesh Bureau of Statistics. Statistics and Informatics Division, Ministry of Planning. (2011). Population and Housing Census 2011. Retrieved from https://www.google.com.bd/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CBsQFjAAahUKEwiLpvDWxYjJAhUFGZQKHSuKDbc&url=http%3A%2F%2Fwww.bbs.gov.bd%2FCensus2011%2FKhulna%2FKhulna%2FKhulna%2520at%2520a%2520glance.pdf&usg=AFQjCNEp3OxFXwpMKW6auCKnX66e9G9EgQ&bvm=bv.107406026,d.dGo

Khulna City Corporation. (2015). Basis statistics. Retrieved from http://www.khulnacity.org/Content/index.php?pid=30&id=32&page=About_KCC

Kopits, E. & Cropper, M. (2005). Traffic fatalities and economic growth. *Accident Analysis & Prevention*, 37(1), 169-178.

Murray, C.J.L. & Lopez, A.D. (1994). A comprehensive assessment of mortality and disability from disease, inquires and risk factors in 1990 and projected to 2020. *The Global Burden of Disease*, Vol. 1.

Oppe, S. (1991). The development of traffic and traffic safety in six developed countries. *Accident Analysis & Prevention*, 23(5), 401–412.

Smeed, R.J. (1968). Variations in the pattern of accident rates in different countries and their causes. *Traffic Engineering & Control*, 10(7), 364–371.

Valli, P.P. & Sarkar, P.K. (1997). *Models for road accidents in India*. Highway Research Bulletin, Indian Roads Congress, 56, 1–11.

Valli, P.P. (2005). Road accident models for large metropolitan cities of India. *IATSS Research*, 29(1), 57-65.