DOI: 10.5455/msm.2014.26.177-181
Received: 25 JFebruary 2014; Accepted: 15 May 2014

Published online: 21/06/2014
Published print: 06/2014

ORIGINAL PAPER Mater Sociomed. 2014 Jun; 26(3): 177-181

Spatial Analysis and Geographic Variation of Fatal and Injury Crashes in Mazandaran Province from 2006 to 2010

Jamshid Yazdani-Charati[1], Hasan Siamian[2], Elham Ahmadi-Basiri[3]

Department of Biostatistics, Health Sciences Research Center, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran[1]
Health information Management Department, Health Sciences Research Center, School of Allied Medical Sciences, Mazandaran University of Medical Sciences, Mazandaran, Sari, Iran[2]
Health Sciences Research Center, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran[3]

Corresponding Author: Elham Ahmadi Baseri, MSc in Biostatistics, Health Sciences Research Center, Faculty of Health, Mazandaran University of Medical Sciences, Sari, Iran. Email: siamian46@gmail.com

ABSTRACT

Background: Road safety and traffic accidents change in time and space. Although, time variations have always been considered the subject being focused by researchers, the effect of spatial correlation and spatial components on the risk of accident have been less investigated. Due to its specific geographical position, Mazandaran Province is one of the highest traffic provinces. This study aims to investigate the factors influencing suburban crashes of Mazandaran province by considering the spatial correlation. **Methods:** This study is aggregated (descriptive -analytical) and the study period was 2006 to 2010. Social and environmental factors effects on the risk of accidents have been studied considering the correlation structure of the regions and regardless of this structure with Poisson regression, negative binomial and Full Bayes hierarchical models. Geographical pattern of risk distribution for the observed values of SMRs and the estimated values after smoothing have been plotted and analyzed. **Results:** Comparing the measures of models goodness of fit indicates that hierarchical Bayes model fits the data better. Plotting the geographical pattern, the north central parts of the province have been identified as the high-risk areas. Human factors were identified as the important factors for the risk of accident. **Conclusions:** The purpose of this procedure is to separate the random effect of residuals correlation. Using this method, the measure of the model goodness of fit got reduced reflecting a better model than the prototype model. The significance of the structured spatial effect shows the existence of unknown explanatory variables with correlated structure whose identification and control can reduce the risk of accidents.
Key words: Full Bayes hierarchical model, spatial correlation, Negative binomial model, Crash risk.

1. INTRODUCTION

One of the biggest problems the world is facing right now is traffic accidents and its consequences. So that it is stated that the main death factor of the young people aged 15-29 globally is traffic accidents and this factor is like an incurable disease such as AIDS or malaria surpassing them. The studies revealed that traffic accidents and the death toll or losses resulting from them, especially in the developing countries are a growing alarming concern while the investments and the studies conducted in this field don't meet the present needs (1).

Roadway accidents are of major factor behind deaths and severe life and financial loss whose heavy social, cultural and economic effects threat human communities (2). In Iran 25% of losses is the result of unnatural deaths due to roadway accidents and it is estimated that more than 22000 would lose their lives annually because of roadway accidents (3).

Mazandaran province is one of the highest traffic provinces countrywide due to its geographical position and the following are Tehran, Isfahan and Khorasan with the highest costs due to accidents. Based on the studies performed, there are 3000 accident prone points in the country and Haraz in Mazandaran (*Ostān-e Māzandarān*) is one of the four high accident-prone axes (2).

Since the environmental factors namely the road and climatic conditions and demographic structure like age, gender are of the important and effective ones in creating crashes and the resulting losses and the regions adjacent to each other have environmental and social conditions close to each other, it is expected that the occurrence of accidents from each area is under the influence of its adjacent areas (4, 5).

Identifying high-risk and low-risk regions is a great help to allocate future resources, facilities and to control future events and since environmental and social conditions are of the factors influencing event occurrence and its aftermath outcomes, taking environmental and social conditions into account for analyzing accidents data is highly significant. Regarding the

environmental and social conditions nature, structure and characteristics, the analysis of such data requires special statistical methods (6, 7).

The main limitation of the classical models used for such data analysis is that the spatial correlation between the observations is ignored (8).

The recent advances in spatial modeling techniques have enabled the researchers to study the discussions critical to unknown variables affecting spatial data and spatial correlation data. The most efficient model to adjust the classical models is Bayes hierarchical model. These methods make spatial smoothing and integrating information possible when the study areas have some rare incidents like motor vehicles accidents (9, 10).

This study has been done with the aim to explain the geographical and social pattern of suburban fatal and injury crashes and to determine the high-risk points considering the areas correlation structure to allocate the resources better and to take control actions in order to reduce and prevent the forthcoming incidents in Mazandaran.

2. METHODS

In order to collect information in this study from the accidents number statistics and aggregated variables necessary in analyzing it, "reports, documents and evidence study "has been applied. The data has been collected based on the theoretical framework and the research hypotheses.

The information related to all Mazandaran suburban areas fatal and injury crashes during 2006-2010 and the province economic, social and cultural planning indices during the mentioned years have been gathered from the statistics and information bureau of the province governor-general office planning deputy and police patrol. Also 2006-2010 weather condition related information as the other necessary variables in this research have been collected from Mazandaran - based meteorological headquarters.

Based on Agdon and Tailor theory, to estimate the values and to set and rank crash hazardous areas, several methods can be employed that differ from each other in terms of significance and accuracy. Such methods are as the following:

a) The number of accidents per road length unit: in this procedure, only the number of accidents is assumed in each road length unit during a certain period and the comparisons are done in each road length unit. This procedure drawback is that traffic volume and accident severity have no effect on the computations.

b) The number of accidents per the vehicles number per road kilometer: in this method, the number of accidents and traffic volume are considered with each other and the rates are calculated in terms of accidents per the vehicles number per million road kilometer.

The methods used in various studies depend on the availability of each one of these pieces of information (11).

In this research, the accidents rate has been estimated per 100 km road of each town.

Regarding this fact that the diverse geographical areas usually include the populations that may have different structures, ,instead of using raw rates, standardized rates are applied to analyze such data that compare the disease, incidence or death status regardless of the social and geographical structure effect in several geographic boundaries. For ecologic studies, the indirect standardization method is utilized (12).

To standardize crashes data indirectly, the province total accidents in the province roads km per year (k/y) has been considered as the standard level and the accidents expected number in every province for that year has been achieved taking this level into account. By dividing the observed cases number by the expected cases number in every province, a level is gained that is called the Standardized Mortality Rate shown as SMR .To determine high-risk and low-risk areas, this level is compared with number 1.

Since the accidents and resultant casualties are numerical data, to analyze the aggregated variables effect like demographic structure and environmental factors effect on accidents risk and the resultant casualties, Poisson regression model as the most common one for such data has been used. Of the important characteristics with Poisson distribution is having equal mean and variance, if this assumption does not hold, that is the data have over dispersion and in Poisson regression model, dispersion parameter is meaningful, we have employed negative binomial substitution model. First, to choose more important and more effective variables over the response, significance level 0.1% has been taken and Poisson regression model and negative binomial model have been fitted to the data and the meaningful variables at this level have been selected for the subsequent models .In Poisson model and negative binomial model, the descriptive variables coefficients and over dispersion parameter have been estimated by likelihood maximum method. These models have been executed in SAS version 9.2.

To study the descriptive variables effect on the accidents risk, by considering the areas spatial correlation structure, the variables meaningful in the initial negative binomial model at 0.1% have been introduced in hierarchical Bayes model. To consider the spatial correlation between the various areas and to analyze the temporal steady effect, the nested conditional auto regression has been applied. According to the previous conducted studies and Wakefield's proposition, through considering the a priori function of the sector being unaware of uniform distribution and with intercept and normal distribution with mean and variance 1000 for the regression variables coefficients, we have fitted the hierarchical Bayes model to the data.

For spatial unstructured effect, based on the previous studies, we considered a priori distribution and for super parameter indicating unstructured effect distribution accuracy, we have considered gamma a priori distribution with parameters 0.5 and 0.0005. For super parameter spatial structured effect indicating the spatial structured effect distribution, we considered gamma a priori distribution with parameters 0.5 and 0.0005.

To determine the accidents dispersion status, models comparison, determining high-risk points and clusters by considering various provinces and regardless of this effect, the accidents risk estimation map has been drawn using these two methods in Arc. GIS software.

3. RESULTS

Out of total 2652 fatal and 14659 injury crashes, during 2006-2010 in Mazandaran province, the highest fatal crashes have occurred in the year 2006 with 579 cases (21.83%) and the highest injury ones also have happened in the same year with 3166 cases (21.6%).

Amol town in the first 4 study years has had the highest fatal

crashes compared with the other towns and in 2010, it has had the maximum of such number following Sari town. Overall, for 5 years, Amol and Sari have had the maximum fatal crashes and the minimum number belongs to Ramsar and Joybar.

The highest % of fatal and injury accidents have been reported in the years 2006 and 2008 with 16.07 and 13.49 %, respectively ,from Amol ,in the year 2007 with 25.39% from Babol and in the years 2009 and 2010 with 26.73 % and 26.41%,respectively ,from Sari .The maximum accidents in road length unit in all the study years have been related to Babolsar.

In ordinary Poisson regression, the value is a model proportional evaluation standard. This value has been obtained 4.89 for model over the accidents data that differs a lot from number 1, and then the ordinary Poisson model doesn't fit these data well and indicates high variability among the data. In Poisson regression model, the dispersion index value is given by this relation .This value equals 2.44 for ordinary Poisson model over the accidents data .This value also confirms the existence of dispersion in the data. Via considering the dispersion in the observations assumption, the negative binomial model has been fitted to data. The dispersion parameter value in this model is 0.161. By fitting the negative binomial model to these data, temporal constant effects, the percentage of the individuals ranging from 15-25 years old, population density, rural roads length, asphalt roads length, freezing days % and the average temperature got meaningful at level 0.1(P-value<0.1).

By excluding the variables that didn't get meaningful in the initial model and through fitting the negative binomial model over the meaningful variables at level 0.1,the effects of the years 2006, 2007 and 2009 got meaningful relative to the year 2010 at level 0.05 while the effect of the year 2008 didn't get meaningful .The effect of the variables 15-25 year-old individuals, population density, rural roads length, asphalt roads length, freezing days % and the average temperature also got meaningful at level 0.05((P-value<0.05).

The risk of accidents in the years 2006, 2007 and 2009 has increased compared with the year 2010. The young population %,the province population density ,the average asphalt roads length and the average temperature have increased the accidents risk while the average rural roads length and the average freezing days % have lowered the accidents risk. Irrespective of the structure, the correlation between the observations of the climatic conditions and human factors has the highest effect on the suburban accidents risk. Considering the AIC goodness of fit in the initial model is 890.04 and in the model including the meaningful variables in the initial model, it is 877.86, it is concluded that the model including the reduced explanatory variables has better fitness to the data. To analyze the effect of the adjacent areas correlation structure on variables' results getting meaningful in the initial negative binomial at level 0.1, they were introduced in hierarchical Bayes model. By putting the data in the hierarchical Bayes model, the effect of all meaningful variables in the negative binomial model got meaningful except for the temporal steady effect of the years 2008 and 2009. Except for the freezing days % variable, the other meaningful variables coefficients in both models were almost close estimations. 1% increase in the freezing days number would decrease the accidents risk exp (0.0556) =1.57 times.

From the standard deviation significance of non-structured error term in the study years, it is concluded that a part of Poisson additional variability is associated with uncorrelated heterogeneity and thus the model is overdispersed. The Poisson additional variability confidence intervals 95% due to the spatial structured error in the study years don't include zero, thus this variability isn't meaningful and it is concluded that in the study years, there is spatial correlation between the observations and part of the Poisson additional variability is expressed with spatial correlation. Estimating the spatial structured effect standard deviation in various study years don't differ much from each other, then spatial variability in the study years isn't much different.

Comparing the AIC and DIC goodness of fit in table 1, it is concluded that the hierarchical Bayes model has better fitness to the data. This is as a result of additional variability related to spatial correlation that unlike the negative binomial model, full Bayes model can explain it.

The results of article showed that the suburban accidents observed SMR values during 20006-2010. From Sari standing between the province eastern and central areas to Noshahr locating between western and central areas, the accidents risk has been fluctuating in different years. A high-risk cluster in the central areas northern parts includes Mahmoudabad, Joybar and Babolsar seen throughout the 5 years. In different years, Amol, Babol, Qaemshahr and Sari have joined this cluster and formed a wider cluster of high-risk areas in the province central parts.

4. DISCUSSION

The differences observed between various points can be under the influence of demographic structures, environmental factors and socioeconomic indices. The effect of these factors on crashes risk has been analyzed by using regression models.

Hierarchical Bayes model considers the differences due to unknown factors in the model through taking the spatial structured effect into account (9).

In classical models, all of the unknown resources for variance-covariance calculation and the estimations standard deviation aren't considered, thus the standard deviation is underestimated and most of parameters get meaningful incorrectly. Bayes models are calculated for all unknown resources and the estimations standard deviation gets higher than what usually occurs in classical models (4).

In the initial negative binomial model fitted to data in the presence of all explanatory variables, the important and influential explanatory variables of the main roads length and road density didn't get meaningful. Since the total road length influences the offset term determination (definition) and it can explain more variability related to the road conditions variables, this result can be justified. The other reason for this result can be attributed to few observations number against the relatively high number of the explanatory variables that leads to the random variability of data against the feasible variability explained by explanatory variables getting high (9).

The accidents negative binomial model's temporal effect significance and its insignificance in the long time in Bayes model can be assigned to the standard deviation underestimation in negative binomial model and these effects getting meaningful in negative binomial model. This result is in line with the results obtained by Jonathan Aguero on road accidents in Pennsylvania during 1996-2000. In his research, though the temporal effects got meaningful in binomial model, in Bayes model the effects of the years 1998 to 2000 didn't get meaningful compared with

the year 1996 (9). Unlike the two models differences in terms of nature and structure and some results, in most of the studies, the models' similarities are more important than their differences. One of the significant similarities of the two models also seen in the majority of the studies is almost identical coefficients of the meaningful variables in the two models (9-13).

The 15-25 individuals % as the representative of young drivers in both models has positive meaningful effect on the accidents risk. This result is consistent with that of the study conducted by Jonathan Aguero in Pennsylvania and the research performed by Ali Lotfi-Darvish in Florida State (9-13).

Population density introduced as a critical factor to state travelling amount and socioeconomic conditions has meaningful effect on accidents risk in both models. Its coefficient as expected is positive but it's insignificant. The only study investigating population density effect on accidents risk is the one conducted in Florida, whose effect didn't get meaningful (13).

High population density in a region indicates more population relative to its area and usually implies more non-aboriginal residents compared to the other regions, the issue can indicate higher traffic relative to the other regions .In our study in various years, Babolsar, Ghaemshahr(Qaemshahr), Babol and Mahmoud Abad are the high-risk towns. These towns have higher population density compared with the other ones.

Regarding the variable coefficient of the rural roads %, it is concluded that accidents risk has been lower in rural roads. The rural roads unevenness relative to the urban ones can be a factor requiring being cautious and observing security measures particularly controlling speed and subsequently, lowering accidents risk. This result has also been achieved in the research done by Amoros about the accidents risk in French cities. In the research by Jonathan Aguero in Pennsylvania also; the mean accidents risk in rural two-way roads has been lower than the other kinds of road (14-15).

Considering the asphalt roads length, this variable increases the accidents risk. Though this variable's coefficient estimation is negligible, its effect is meaningful. In the study by Srenio eso et al. that compared negative binomial model and Lindley negative binomial model and these models and used these models for India and Michigan State accidents data, the asphalt roads length in India had enhancement effect on the accidents while in Michigan State, it led to accidents risk reduction (16). This result in our study can be somehow attributed to driving culture and ethics among people. Since the asphalt roads possess more appropriate conditions relative to the dirt and sand roads..., driving on these routes is less cautiously done by the drivers especially about controlling speed relative to the other roads, thus the accidents risk rises. In the research done about Haraz axis crashes, there is a meaningful relationship between asphalt binder course roads inappropriateness and the accidents (17). Because in this study, no separation has been done between theses roads in terms of the binder course quality, the result gained cannot be assigned as a factor.

Among the variables related to the weather conditions, freezing day's % and average temperature have respectively decreased and increased the accidents risk. The main reason behind the accidents risk increase as average temperature rises is the vehicles traffic over-rising due to the good weather condition and the drivers' tendency to drive under such nice weather condition while under adverse climatic conditions and freezing, due to the roads being closed by the authority organizations and the drivers' reluctance to drive under such conditions, the accidents risk decreases. These results are compatible with that one conducted on Karaj-Chalous axis but don't agree with the research done on Sanadaj-Marivan axis and Canada (18, 19, 20).

The variable coefficient of freezing days % in Bayes model has had high variation in terms of magnitude relative to negative binomial model. Since there are rarely freezing days in the province and in special towns and this variable value has had high variation in the diverse years and regions, it has been influenced by the model and its coefficient estimation has changed a lot.

The dispersion parameter value in the initial negative binomial model in the presence of all variables is 0.161 and with respect to confidence interval 95 %(0.1092 and 0.2135), the assumption based on its being zero has been rejected. That means 16% of the observed differences among the accidents risk is related to unknown factors. The dispersion parameter value has increased by excluding the insignificant variables and its confidence interval is 95 %(0.1156 and 0.2257).Although, in the reduced negative binomial model, higher data variability hasn't been explained by the existing variables, regarding the AIC goodness of fit , the reduced model has been more fitted to the data.

5. CONCLUSIONS

The major goal of the study is to analyze the demographic structure effect and environmental factors and temporal variations on the accidents risk and its resultant losses (casualties) and to state its geographical distribution in Mazandaran province and to determine the factors influencing these variations in several consecutive years that can be used a guide by the planners to execute the preventive interventions and control.

In Poisson regression model as the most common one to explain numerical data such as the accidents data, due to the data overdispersion, the model fitness gets reduced. To state such data better, some substitution models like negative binomial model is utilized. When these models are applied to explain data in geographical scale, the areas proximity and correlation effect is ignored. To control the spatial proximity effect on the spatial scale data and the model better fitness to data, the smoothing methods have been introduced that the most robust one is hierarchical full Bayes model.

Therefore, hierarchical Bayes model including spatial autocorrelation detects the spatial communications of smaller local units better since factors such as weather conditions vary more in smaller temporal and local intervals. To observe their variability in the considered question, studying in smaller temporal and local scales is recommended due to increasing the variables explanatory power.

The study results suggest that the accidents risk has a cluster clear spatial pattern. The various years variations don't modify (adjust) the spatial pattern and the risk doesn't show much variation at different times .Thus, it is recommended to apply interventions to control and lower the accidents risk and their resultant losses in long-term.

Concluding about the relationship between the response and the explanatory variable is strongly under the influence of choosing the model. Ignoring overdispersion in the analysis results in the standard deviation underestimation and it may lead to the less important variables get highly meaningful in the model.

Bayes model better fitness to data relative to Poisson and

negative binomial models indicates that the original source of data variability is spatial correlation. Regarding this subject, it is recommended that by studying more to identify the influential environmental variables in the accidents risk and to implement them in the models with spatial structure, to analyze their effect on the model and to take control measures associated with them.

One of the goals of the present study is to determine the accidents high-risk and low-risk regions .Through identifying these regions, the planners are recommended to identify the behaviors that cause this value being low in these regions by analyzing the socioeconomic conditions of low-risk areas and execute and control them in high-risk regions.

With respect to the obvious effect of human and environmental factors on the accidents risk and their due losses, to apply the correct scientific programs and to appropriately invest in order to lower the accidents statistics and to promote the roads security through giving the necessary education to the youth as the most important human factor affecting the accidents, to maintain the roads and to post the essential signs seem necessary.

The real differences in temporal trend are better revealed in the models with time and location (space) interactions. It is recommended to execute and analyze these complicated methods on data in the subsequent studies.

To better comprehend and control various variables effect on the accidents risk, introducing traffic volume, the number of the vehicles involved and the type and severity of the accidents by applying the appropriate weight in the model seems more suitable.

An important and influential factor in the conditional autoregression model fitness to spatial data is proximity matrix .Analyzing the models sensitivity by putting various proximity matrices regarding the spatial units intervals and explanatory variables in the model is recommended in order to study the spatial proximity.

Some more specialized techniques like geostatistical techniques are effective in predicting the values for the regions not having recorded environmental and weather conditions. These techniques can provide some information about small scale such as the road segment surface .Using such techniques is recommended in the future studies.

Acknowledgments

This article was extracted from Proposal no. 90-6 which accepted and granted by Mazandaran University of Medical Sciences, Sari, Iran.

CONFLICT OF INTEREST: NONE DECLARED.

REFERENCES

1. Pourmoallem N, Salimi MB. The Analysis and Assessment of Black Spots (From the Viewpoint of Accident Location Recognition, Costs of Damages and Remedial Works). Traffic Management Studies. 2009; 4(12): 25-48. (Persian)
2. Sadat Hoseini SM, Soleimani M. The Study of Fatal Highway Accidents' Causes (A Case Study of Esfahan). Traffic Management Studies. 2009; 4(14): 71-84. (Persian).
3. Karami S, Farajzadeh M. Road Accident Analysis by a Climatic Approach Using Geographic Information Ystems (Gis) Case Study: Firouz Koh- Sari Road. Modarres Human Sciences. 2005; 9(1): 151-166.
4. Song J, Ghosh M, Miaou S, Mallick B. Bayesian multivariate spatial models for roadway traffic crash mapping. Journal of Multivariate Analysis. 2006; 97(1): 246-273.
5. Ribeiro C, Turkman A, Cardoso J. Bayesian hierarchical models for road accidents in small-area. International Workshop on Spatio-Temporal Modeling. 2010; 30: 3-10.
6. Guthrie KA, Sheppard L, Wakefield J. A hierarchical aggregate data model with spatially correlated disease rates. Biometrics. 2002; 58(4): 898-905.
7. MacNab YC. Bayesian spatial and ecological models for small-area accident and injury analysis. Accident Analysis and Prevention. 2004; 36(6): 1019-1028.
8. Rytkönen MJ. Not all maps are equal: GIS and spatial analysis in epidemiology. International journal of circumpolar health. 2004; 63(1).
9. Aguero-Valverde J, Jovanis PP. Spatial analysis of fatal and injury crashes in Pennsylvania. Accident Analysis and Prevention. 2006; 38(3): 618-625.
10. Lawson A, Biggeri A, Boehning D, Lesaffre E, Viel J, Clark A, et al. Disease mapping models: an empirical evaluation. Disease Mapping Collaborative Group. Statistics in medicine. 2000; 19(17-18): 2217-2241.
11. European Union Road Federation. Good Practice Guidelines to Infrastructural Road Safety: European Union Road Federation (ERF), 2002.
12. Bailey TC. Spatial statistical methods in health. Cadernos de Saúde Pública. 2001; 17(5): 1083-1098.
13. Lotfi Darwiche A. A GIS safety study and a county-level spatial analysis of crashes in the state of Florida. USA: Florida: University of Central Florida Orlando, 2009.
14. Amoros E, Martin JL, Laumon B. Comparison of road crashes incidence and severity between some French counties. Accident Analysis and Prevention. 2003; 35(4): 537-547.
15. Aguero-Valverde J, Jovanis PP. (Eds). Identifying road segments with high risk of weather-related crashes using full Bayesian hierarchical models. 86th Annual Meeting of the Transportation Research Board, Washington, DC, 2007.
16. Geedipally SR, Lord D, Dhavala SS. The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. Accident Analysis and Prevention. 2012; 45: 258-265.
17. Ahadi M, Doagouyan D, Tavangar M. Evaluati A Review of Hgv Role In Road Accidents (a Case Study of Haraz Road). Traffic Management Studies. 2010; 5(17): 53-72.
18. Farajzadeh M, Baghdam O. The Evalution of Road Accident with Environmental Hazards Approch Using GIS; Study Area: Sanandaj - Marivan Axies. Modares Journal. 2005; 9(1): 1-13. (Persian).
19. Farajzadeh M, Gholizadeh M, Adabi Firozjaei A. The analysis of road accident with climatic approach: Karaj - Chalus road. Physical geography research journal. 2010; 73: 37-52.
20. Andreescu MP, Frost DB. Weather and traffic accidents in Montreal, Canada. Climate Research. 1998; 9(3): 225-230..