A Hybrid of Structural Equation Modeling and Artificial Neural Networks to Predict Motorcyclists’ Injuries: A Conceptual Model in a Case-Control Study

Shila HASANZADEH 1, *Mohammad ASGHARIJAFARABADI 1,2, Homayoun SADEGHI-BAZARGANI 1,2

1. Road Traffic Injury Research Center, Tabriz University of Medical Sciences, Tabriz, Iran
2. Department of Statistics and Epidemiology, Faculty of Health, Tabriz University of Medical Sciences, Tabriz, Iran

*Corresponding Author: Email: m.asghari862@gmail.com

(Received 12 Feb 2019; accepted 25 May 2019)

Abstract
Background: To model, the predictors of injuries caused the hospitalization of motorcyclists using a hybrid structural equation modeling-artificial neural network (SEM-ANN) considering a conceptual model.

Methods: In this case-control study, 300 cases and 156 controls were enrolled using a cluster random sampling. The cases were selected among injured motorcyclists in refereed to Imam Reza Hospital and Tabriz Shohada Hospital, Tabriz, Iran since Mar 2013. The predictability of injury by motorcycle-riding behavior questionnaire (MRBQ), Attention-deficit/hyperactivity disorder (ADHD) along with its subscales and motorcycle related variables was modeled using SEM-ANN. By SEM, linear direct and indirect relationships were assessed. To improve the SEM, the ANN was utilized sequentially to account for the nonlinear and interaction effects that is not supported by SEM.

Results: The predictors of injury were: MRBQ, ADHD, and its subscales, marital status, education level, riding for fun, engine volume, hyper active child, dark hour riding, cell phone answering, driving license (All \( P \) less than 0.05). In addition, the findings reveal the Mediating role of MRBQ for the relationship between underlying predictors and injury. Furthermore, ANN showed higher specificity (95.45 vs.77.88) and accuracy (90.76 vs.79.94) than usual SEM which lead us to introduce the second and third order effect of MRBQ into the modified SEM.

Conclusion: The hybrid model provided results that are more accurate; considering the results of the modeling, having intervention programs on ADHD motorcyclists, those have the hyperactive child, and those who answer their cell phones while driving, and improving the motorcyclists’ goal is highly recommended.

Keywords: Motorcyclists; Traffic injury; Structural equation modeling, Neural networks

Introduction

Road Traffic Injury (RTI) is one of the most important health issues (1). According to the WHO’s report, about 2.5 million deaths are caused by RTI. As a global problem, RTI causes 1.27 million deaths and 20-50 million injuries, annually (2). RTI ranks third among causes of death in Iran (3). While Iran accounts for less than 1% of the world's population, 2.5% of traffic accidents occur in Iran (4). This emerges effective and continuous action for RTI, especially in Iran.
Motorcycle riders are amongst more vulnerable group among Iranian. They are 7.03 and 5.5 times more vulnerable for injuries than the car and truck drivers, respectively. In Iran, the motorcycle-related injuries comprise 42% of RTI and motorcyclists comprise more than 51% of RTI deaths (5). Iran also is ranked one in deaths by RTI among the 15-29 yr youth motorcyclists (6). Therefore, the motorcyclists are in the top priority to be investigated to reduce the motorcyclists’ RTI and deaths in Iran (7).

Deficit/Hyperactivity Disorder (ADHD) is one of the main causes of traffic accidents (8, 9). Drivers with ADHD are three to four times more likely to have accidents than the other drivers (10). The study on this issue would be a valuable action in traffic safety (11). The (generalized) linear models may be insufficient and be prone to error. These models were used in all previous studies (12-14).

Recently, a hybrid methodology of SEM-ANN was introduced which may be suitable in these situations (15). SEM estimates the multiple equations simultaneously via a conceptual model and takes into account the measurement error (16-18), and investigates the mediator/moderator role of variables (19). The ANN covers the non-linear/interaction effect and does not need the restrictive assumptions of linear models such as normality, homogeneity of variances and the independence of residuals (17). Consequently, to cover these deficiencies and to provide a better prediction, hybrid SEM-ANN would be a proposition.

The aim of this study was to predict injuries caused by the hospitalization of motorcyclists using a hybrid SEM-ANN model considering a conceptual model.

Materials & Methods

Participants and procedures
This case-control study was conducted on 300 cases and 156 controls. The cases were selected among injured motorcyclists in referred to Imam Reza Hospital and Tabriz Shohada Hospital since Mar 2013. For cases, the sampling continued consecutively until completing the sample size. The control group was selected based on the postal code classification system of the city, which was used in the safe community project of Tabriz. From 150 codes within this project, 50 postal codes were selected randomly to be included in this study. Each postal codes constituted a cluster, and 25 households were surveyed in each cluster. Based on the numbers on the right side of the postcode; households were selected and entered the study. Then, the researcher went to that house to complete his/her questionnaires. If there is more than one motorcyclist at home, the researcher chooses a motorcyclist that would match the age group of other subjects.

Study Size
The sample size was estimated to be of 300 subjects for the case group and 300 for the control group, considering the relationship between MRBQ and injury, 95% confidence level and a power of 80%. Since the number of subjects was low in the control group, 156 samples participated in this group. In this case, given the fact that data collection was done before, access to new samples did not possible.

Data collection
A checklist of demographic information and socioeconomic status, and a valid screening ADHD questionnaire were used. The ADHD has four subscales, including A (inattention), B (hyperactivity, impulsivity), C (A+B), and D (ADHD index) (20). Furthermore, a valid and reliable MRBQ was used to assess the behavior of drivers (9).

Study variables
The predictors of injury were: age (yr), marital status (married/single), level of education (university/other), motorcycle goal (fun/other), helmet usage (yes/no), cell phone answering (no/other), driving recode (>1yr/≤1yr), motorcycle glider (yes/no), income category (<500TT/500-1000TT/1000+TT), referring to psychiatrist (yes/no), having hyper ac-
tive child (no/yes), engine volume (>125cc/≤125cc), having driving license (yes/no), riding per week, dark hour riding.

Inclusion and exclusion criteria
Inclusion criteria for the case group of traumatized patients, motorcycle driver, vigilant people when filling in the questionnaire and injured people in Tabriz. Exclusion criteria from the study for the case group were those who have low awareness due to brain stroke, the injured motorcycle who has been transferred from other cities in the province, and deliberate trauma due to the motorcycle accident. Inclusion criteria for the control group were men who had motorcycles without being injured. Exclusion criteria for the control group were those with a history of admission in the hospital due to trauma.

Ethical approval
The study protocol was approved by the Ethical Committee of Tabriz University of Medical Sciences (ethics code: TBAMED. REC.1394.783). In this study, information obtained from participants is used for scientific purposes, meanwhile, the privacy was preserved. All participants filled and signed the informed consent and assent. For the illiterate people, the informed consent form was read by the researcher or someone to whom he/she trusts. Then instead of signing, fingerprints were taken from participants.

Conceptual model
To construct the conceptual model, the relationships found in the various studies (12-14) were considered as the clues and the competing models were modified by supporting data. The relationship between injury with endogenous variable (ADHD) and the subscales, MRBQ as well as the marital status, level of education, riding only for fun, driving license, driving precedence, dark hour riding, engine volume, having hyperactive child, cell phone answering, riding per week, group income, referring to psychiatrist was evaluated.

Statistical analyses
Statistical analysis was done by STATA software [ver.13] (Stata Corp, College Station, Texas 77845 USA) and STATISTICA [ver. 12.5] (STATISTICA, Stat Soft, Inc. USA). Data were presented using mean (SD), median (min-max) for the numeric normal and non-normal variables respectively and frequency (percent) for categorical variables. Generalized structural equations model (GSEM) was used to investigate the relationships in the conceptual model. For GSEM, a multi-stage process has been used: model specifications, model identification, model estimation, model testing and model modification (21). In addition, the ANN was used to evaluate linear and nonlinear relationships and interactive effects. Different steps of ANN include selecting important variables as input, preprocessing data, dividing datasets, training, testing and validation samples, neural network paradigms, evaluation criteria, neural network training, and implementation. A multilayer perceptron (MLP) model was used and model validity evaluated using sensitivity, specificity and precision (17). Finally, the two models were compared by the diagnostic indices including sensitivity, specificity, and accuracy, (based on a cutoff point of 0.5 for the probability of occurrence of injury), and area under the ROC curve (AUC) of the two model as well.

Results
The final sample consisted of 456 male motorcycle riders with the mean age of participants was 30 (SD=10.8) yr. More than two-thirds of them did not have a motorcycle license, and 42% used the helmet. Almost 18% of people were riding motorcycles only for fun purposes and 73% of motorcycles had an engine volume of =>125cc. Other information on variables is presented in Table 1. The distribution of the main outcome variable, injury, and the main predictors, as well as ADHD and ADHD subscales, are reported in Table 2.
Table 1: Characteristics of the study's participants (n=456)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Control (n=156)</th>
<th>Case (n=300)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n/Mean (%/SD)</td>
<td>n/Mean (%/SD)</td>
<td></td>
</tr>
<tr>
<td>Having driving license (Yes)</td>
<td>235 (70.6)</td>
<td>98 (29.4)</td>
<td>0.005</td>
</tr>
<tr>
<td>Driving record (&lt;1yr)</td>
<td>28 (53.8)</td>
<td>24 (46.2)</td>
<td>0.044</td>
</tr>
<tr>
<td>Marital status (Married)</td>
<td>150 (71.8)</td>
<td>59 (28.2)</td>
<td>0.035</td>
</tr>
<tr>
<td>Academic education (University)</td>
<td>276 (69.7)</td>
<td>120 (30.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Riding just for fun (Fun)</td>
<td>238 (64.3)</td>
<td>132 (35.7)</td>
<td>0.028</td>
</tr>
<tr>
<td>Engine volume (&lt;=125cc)</td>
<td>66 (52.8)</td>
<td>59 (47.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>referring to psychiatrist (Yes)</td>
<td>15 (39.5)</td>
<td>23 (60.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cell phone answering (Yes)</td>
<td>113 (52.6)</td>
<td>102 (47.4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smoking (Yes)</td>
<td>115 (62.5)</td>
<td>69 (37.5)</td>
<td>0.157</td>
</tr>
<tr>
<td>Income group (500-1000TT)</td>
<td>131 (60.7)</td>
<td>85 (39.4)</td>
<td>0.024</td>
</tr>
<tr>
<td>Hyperactive child (Yes)</td>
<td>96 (57.8)</td>
<td>70 (42.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Annual fin (Yes)</td>
<td>80 (58)</td>
<td>58 (42)</td>
<td>0.013</td>
</tr>
<tr>
<td>Motorcycle glider (Yes)</td>
<td>26 (76.5)</td>
<td>8 (23.5)</td>
<td>0.340</td>
</tr>
<tr>
<td>Having glasses (Yes)</td>
<td>37 (53.6)</td>
<td>32 (46.4)</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Table 2: Summary and descriptive statistics for quantitative data (N=456)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Control (n=156)</th>
<th>Case (n=300)</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD) (Min-Max)</td>
<td>Mean (SD) (Min-Max)</td>
<td></td>
</tr>
<tr>
<td>ASS</td>
<td>4.2 (3.8) (0-20)</td>
<td>7.1 (4.6) (0-21)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BSS</td>
<td>7.9 (4.2) (0-22)</td>
<td>8.7 (4.9) (0-25)</td>
<td>0.037</td>
</tr>
<tr>
<td>CSS</td>
<td>12.1 (7.0) (0-41)</td>
<td>15.8 (8.5) (0-42)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DSS</td>
<td>8.9 (5.2) (0-27)</td>
<td>10.6 (5.4) (0-26)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MRBQ</td>
<td>2.2 (1.1) (1-4)</td>
<td>2.9 (1.1) (1-4)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>ADHD</td>
<td>21.0 (11.4) (0-68)</td>
<td>26.4 (13.3) (2-67)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>29.4 (11.8) (13-79)</td>
<td>30.7 (8.6) (16-59)</td>
<td>0.214</td>
</tr>
<tr>
<td>Hours Driving</td>
<td>115.8 (98.0) (0-540)</td>
<td>148.2 (113.5) (0-480)</td>
<td>0.002</td>
</tr>
<tr>
<td>Dark Hour Riding</td>
<td>24.7 (51.6) (0-420)</td>
<td>55.1 (61.6) (0-300)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Riding Per Week</td>
<td>5.7 (2.1) (0-20)</td>
<td>5.8 (1.6) (0-7)</td>
<td>0.945</td>
</tr>
</tbody>
</table>

MRBQ: Motorcycle Rider Behavior Questionnaire; ADHD: Attention Deficit Hyperactivity Disorder; ASS Score: subscale A measuring inattention; BSS Score: subscale B measuring hyperactivity, impulsivity; CSS Score: subscale C the sum of A and B subscales; DSS Score: subscale D measuring ADHD index; MRBQ ranges over (0, 192) and ADHD ranges over (0, 90); ASS and BSS ranges over (0, 27); CSS ranges over (0, 54); DSS ranges over (0, 36)
The results indicate a moderate level of total ADHD and the subscales, and MRBQ in the study population.

**SEM modeling**
A model was fitted on data taking into account the relationship of MRBQ, ADHD and underlying factors with injury in two models: MRBQ as mediator and do not as mediator. According to the AIC, BIC, the results showed the better fit for model with MRBQ as mediator (BIC=12167.3, AIC=11932.3 in model with MRBQ as mediator vs. BIC=12214.52, AIC=11979.53 in model without mediator). Hence in continue, the model with the mediator was used and interpreted. In the next step, to consider the subscales of the ADHD, a latent ADHD was fitted in another model (Fig. 1 and 2).

![Diagram](https://example.com/diagram.png)

*Fig. 1: The conceptual model of relationship between motorcyclerelated variables, ADHD, MRBQ with Injury
MRBQ: Motorcycle Rider Behavior Questionnaire; ADHD: Attention Deficit Hyperactivity Disorder.
* P<0.05 and ** P<0.01

The results indicated significant linear and direct relationships between odds of injury and cell phone answering (OR= 2.22, 95% CI= 1.33 to 3.71, \( P= 0.010 \)), hyperactive child (OR= 1.65, 95% CI= 1.07 to 2.55, \( P= 0.057 \)), dark hour riding (OR= 1.011, 95% CI= 1.005 to 1.014, \( P= 0.001 \)) and MRBQ (OR= 1.27, 95% CI= 1.01 to 1.62, \( P= 0.092 \)), while significant inverse relations was observed between injury and being married(OR= 0.428, 95% CI= 0.271 to 0.676, \( P= 0.002 \)), academic education (OR= 0.29, 95% CI= 0.16 to 0.52, \( P= 0.001 \)). Additionally, MRBQ mediates the relation between injury and ADHD (Fig. 1).

The latent ADHD showed a significant relationship with CSS and DSS and the other relationships remained the same as those mentioned in the above model (Fig. 2).
Fig. 2: The conceptual model of relationship between motorcycle related variables, latent ADHD, MRBQ with Injury
MRBQ: Motorcycle Rider Behavior Questionnaire; ADHD: Attention Deficit Hyperactivity Disorder; CSS Score: subscale C measuring (ASS+BSS); ASS Score: subscale ASS measuring inattention; BSS Score: subscale B measuring hyperactivity, impulsivity; DSS Score: subscale D measuring ADHD index; * $P<0.05$ and ** $P<0.01$

**ANN modeling**
Initially, 19 variables were selected as the network input. The optimal ANN model consisted of 3 layers, each containing 5 neurons. The model used the perceptron approach (minimum = 1 and maximum=17 hidden units). It should be mentioned that 16 of 19 variables were influential variables. A comparison of sensitivity, specificity, accuracy and AUC between SEM and ANN models, indicated higher values of accuracy, sensitivity and AUC for ANN than SEM and approximately the same value of specificity in these models (Table 3) specifying better prediction for ANN.

**Table 3: Comparison between SEM and ANN**

<table>
<thead>
<tr>
<th>Variable</th>
<th>ANN</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity (%)</td>
<td>79.45</td>
<td>80.82</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>95.45</td>
<td>77.88</td>
</tr>
<tr>
<td>Correct (%)</td>
<td>90.76</td>
<td>79.94</td>
</tr>
<tr>
<td>AUC (%)</td>
<td>94.22</td>
<td>85.93</td>
</tr>
</tbody>
</table>

Regarding Specificity, Sensitivity lower and higher Correct ANN Model is better

In ANN model, the linear relationship assumption was assessed by the lowest method and the results indicate that there were nonlinear relationships between MRBQ and odds of injury (Fig. 3).
Fig. 3: Nonlinear relationship between MRBQ and odds of Injury indicated in ANN

**Modified SEM modeling based on ANN results**

Regarding the presence of nonlinear relationships of the variables, which was indicated by ANN, in the next step, the modified SEM model was re-fitted by including the second order and the third order of MRBQ in the model. The results showed significant relationships between the second order and the third order of MRBQ with odds of injury in the model (Both \( P<0.001 \)) (Fig. 4).

Fig. 4: The conceptual model of relationship between motorcycle related variables, ADHD, MRBQ, MRBQ2 and MRBQ3 with Injury.

MORBQ: Motorcycle Rider Behavior Questionnaire; MRBQ2: second order MRBQ; MRBQ3: third order MRBQ; ADHD: Attention Deficit Hyperactivity Disorder.

* \( P<0.05 \) and ** \( P<0.01 \)
Discussion

This study demonstrated that the better performance of hybrid SEM-ANN; the ANN alleviated the violation in linear relationship assumption, which is considered in the SEM, hence resulted in higher values of accuracy and sensitivity and hence better prediction for ANN. It may be concluded that the hybrid SEM-ANN improve the prediction of outcome as compared to each technique separately.

This study examined the new methodological of hybrid (SEM-ANN) modeling for the data of a case-control study in the field of traffic, which enables providing a deeper analysis and understanding Motorcyclists behavior. Our rationale in the application of hybrid SEM-ANN in describing the relationship between underlying predictors of the injury was:
1. Considering a simple interpretation of relationships via SEM
2. The inclusion of linear and nonlinear relationships via ANN.

This has been not considered in traffic data yet. There were several studies, which utilized the SEM and ANN separately/simultaneously: A study showed that the use of artificial neural networks might improve the accuracy of driving prediction (22). ANN is known as the black box; it models the relationships via a hidden layer (and not a clear path) (23). In addition, the GSEM models simultaneously the endogenous and exogenous variables (24). In another study, SEM was used to examine the relationship between traffic noise and health (25). The rationale of using hybrid SEM-ANN is that the ANN the best predict the outcome with higher accuracy and SEM investigate the relationship between variables in a causal context, therefore the combination of two methods would yield good results (18).

As other results of this study, better fit has been found for the model with MRBQ as the mediator; according to the AIC, BIC and the findings indicate that the MRBQ mediates the relation between injury and ADHD. To consider the subscales of the ADHD, a latent ADHD was fitted in another model, which showed a significant relationship of latent ADHD with CSS and DSS in the model. The results indicated significant linear and direct relationships between odds of injury and cell phone answering, hyperactive child, dark hour riding and MRBQ, while the significant inverse relation between injury and being married, academic education. It has been shown that depressive disorder, behavioral disorder and other behavioral problems in individuals with ADHD were more than those without ADHD (RR=1.23) (26). Besides, another study found the significant relationship between the accident and the subscales of ADHD (27). Moreover, our study supports the relationship between ADHD in children and ADHD in their parents.

In our study, there was the significant relationship between the injury and having the hyperactive child. The parents with ADHD children are at a higher risk of harm than the parents of children without ADHD; this requires childhood care and (27), The age is considered as risk factor for injury in motorcycles, because youth and adolescents and singles are more likely to have accidents than the other age categories (28, 29). Furthermore, a review of several studies supported being young as the main risk factor for injury (13, 30). Using the motorcycle in crowded times and in the case of focus reduction, high speed and non-compliance are the main reasons for injury in this age period (31). Driving in crowded hours, anger and mobile phone answering increase traffic accidents (32). Educated people are less likely to be injured than other people (6, 33). Young motorcyclists always utilize motorcycles just for fun, acrobatic movements and endanger themselves and others living with dangerous behaviors (34) which theses high-risk motorcycles are responsible for 90% of road accidents (35).

Study limitations

The absence of appropriate and standardized conceptual model for SEM was a major limitation of this study; we tried to build a conceptual model based on integrated evidence and was

Available at:  http://ijph.tums.ac.ir
supported by our data, which may be proposed as the initial conceptual model for future studies. However, the study was conducted only for a sample of motorcycle riders in Tabriz City and the surrounding areas, hence, we cannot generalize the results to other parts of Iran; other studies are required to further investigate the model and relationships. Other limitations of the study include self-declaration questionnaires, the possibility of the existence of recall bias in the case-control design; a cohort design to better assess the relationship of behavior in ADHD motorcyclists and the injury. Sample imbalance in the case and control groups was another limitation of the current study.

Conclusion

The hybrid SEM-ANN performed better than each one separately. The ANN alleviated the violation in linear relationship assumption of SEM so resulted in higher accuracy. Furthermore, considering the mediating role of MRBQ for the relationship between injury and ADHD as well as the underlying predictors of injury, useful guidelines might be developed for society, traffic users, policymakers, manufacturers, research centers, and environmental organizations for better planning and reducing traffic accidents. In addition, the hybrid model is recommended for the results with higher accuracy.

Ethical considerations

Ethical issues (Including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, redundancy, etc.) have been completely observed by the authors.

Acknowledgements

We would like to appreciate Research Deputy of Tabriz University of Medical Sciences who supported our study. This project was supported by Road Traffic Injury Research Center and Research Deputy of Tabriz University of Medical Sciences, Tabriz, Iran.

Conflict of interest

The authors declare that there is no conflict of interest.

References


