CHAID ANALYSIS OF VEHICULAR ACCIDENTS: A BASIS FOR A MITIGATION POLICY

Carla Carmela P. Perez
ccperez@psu.edu.ph
Pangasinan State University

Abstract - The study investigated the occurrence of the offense committed in the major roads of Binalonan (Binalonan-Pozorrubio, Binalonan-Urdaneta, Binalonan-Laoac and Binalonan-San Manuel) from the period of 2013-2015 using Chi Square Automatic Interaction Detector (CHAID) Analysis. This study aimed to find the best predictor that foretells the offense committed in Binalonan, Pangasinan specifically in determining the characteristics profile of the respondent in terms of the time, vehicle type, age, sex, civil status, and place, identifying the significant relationship between the independent variable and dependent variable, and identifying the best characteristics profile of the respondent that predicts the offense committed in Binalonan, Pangasinan. This study made use of descriptive method of research which is defined as description, recording, analysis and interpretation of the presented data.

Based on the findings, majority of the offenses committed cases have increasing probability that the offender on vehicular accidents end up from reckless imprudence resulting to damage to property. Also, there is a significant relationship between the characteristics profile of the respondents and the offense committed due to vehicular accidents. Among all the characteristics profile of the respondents only time was detected as best predictor of the offense committed due to vehicular accidents. Thus, National Policy should do a better idea and strategy in terms of time which is greatly related to vehicular accidents in order to reduce the offense committed due to vehicular accidents in particular location for the preparation of the future prediction.

Keywords – Reckless driving, Classification Tree, offense, Binalonan Pangasinan

I. INTRODUCTION

Road accidents have become a major issue of concern worldwide. Accident is an event, occurring suddenly, unexpectedly and inadvertently under unforeseen circumstances. A road accident is commonly defined as the collision of vehicles, pedestrian or with an object that will result to death, injury and property damage with at least one moving vehicle involvement. In the Philippines, road accidents is the leading cause of deaths. According to the World Health Organization (WHO), 1.2 million people worldwide die from road accidents. In the Philippines, vehicular accident is the number 1 type of accident that results to an increasing number of deaths and number 5 among the top causes of death in the country. Most of the road accidents that cause death of people primarily involve drivers or those victims. These road accidents can be attributed to the defective roads and non-standard road signs.

Many researchers have come up with the causes, effects and recommendations of vehicular accidents. These causes include drunken driving, machine failure and over speeding (Sagberg, Fosser, and Saetermo, 1997). Many research studies have been carried out about the many reasons of road accidents in the Philippines which includes drunken driving under the (influence of alcohol and drugs), over speeding, mechanical malfunctions (happens when the driver does not regularly check the vehicle) and overtaking (happens when the drivers go ahead of another vehicles). In short, driver/driving errors are the main reasons of the continuously increasing number of vehicular accidents. There are numerous suggested solutions, various inventions by the government, nongovernmental organizations and other road stakeholders in curtailing road accident and its effects on human lives and properties. The mere increase in the number of accidents is not enough for one to conclude that there is really an increase in vehicular

accidents; hence the need to analyze the accidents' data statistically is a must to check whether there is any evident increase in road accidents as years go by resulting to large number of people losing their lives (Oppong, 2012).

Reckless driving has not been defined as a crime in statutory books but is only penalized as a quasi-offense under article 365 of the Revised Penal Code as reckless imprudence resulting to damage to property. According to Atty. Manuel J. Laserna Jr., reckless imprudence is generally defined by our penal law as doing or failing to do an act from which material damage results by reason of inexcusable lack of precaution on the part of the person performing or failing to perform such act, taking into consideration his employment or occupation, degree of intelligence, physical condition and other circumstances regarding persons, time and place. Reckless imprudence resulting to homicide, reckless imprudence resulting to physical injury, reckless imprudence resulting to damage to property are the offenses committed in particular incidents as vehicular accidents.

CHAID is a Classification Tree technique that not only evaluates complex interactions among predictors, but also displays the modeling results in an easy-to-interpret tree diagram. The "trunk" of the tree represents the total modeling database. CHAID then creates a first layer of "branches" by displaying values of the strongest predictor of the dependent variable. CHAID automatically determines how to group the values of this predictor into a manageable number of categories (e.g. we may start with ten categories of age, and CHAID might collapse these ten categories down to only four or five statistically significant different age groups).

CHAID then creates additional layers of branches off of each age grouping, using the strongest of the remaining predictors. It continues this branching procedure until the final branches or "twigs" of the tree have been generated. If CHAID is being used to generate a predictive market segmentation model, then these terminal branches are the final market segments.

This study investigated the occurrence of the offense committed in the major roads of Binalonan which are the Binalonan-Pozorrubio, Binalonan-Urdaneta, Binalonan-Laoac and Binalonan-San Manuel from the period of 2013-2015. Chi Square Automatic Interaction Detector Analysis (CHAID) was used to identify and analyze the characteristics profile of the respondents which affects the offense committed in Binalonan. The use of CHAID algorithm was provided to solve the problem.

II. STATEMENT OF THE OBJECTIVES

This study aimed to find the best predictor that foretells the offense committed in Binalonan, Pangasinan with the following specific objectives:

1. To determine the characteristics profile of the respondent in terms of the following:
   a. Time;
   b. Vehicle Type;
   c. Age;
   d. Sex;
   e. Civil Status; and
   f. Place.
2. To identify the significant relationship between the independent variable and dependent variable.
3. To identify the best characteristics profile of the respondent that predicts the offense committed in Binalonan, Pangasinan.

III. METHODOLOGY

This study made use of descriptive method of research which is defined as description, recording, analysis and interpretation of the presented data. This type of research is the most commonly used. The basic reason for carrying out descriptive research is to identify the cause of something that is happening (DJS Research Ltd, 2010). The process of descriptive research goes beyond mere gathering and tabulation of the information and data. It involves interpretation of the significance of what is being described.

The data used in this study was secondary data which was obtained from the Binalonan Police Station. The data of this study considered accidents data, a three-year record of accidents in Binalonan for the period of 2013

to 2015. The data is treated for the appropriate codes that was used in the study.

IV. FINDINGS

Characteristics Profile of the Respondents

The attributes of the respondents considered in this study were Time, Place, Age, Sex, Civil Status and Vehicle type. The data were analyzed and described using the frequency counts and percentages.
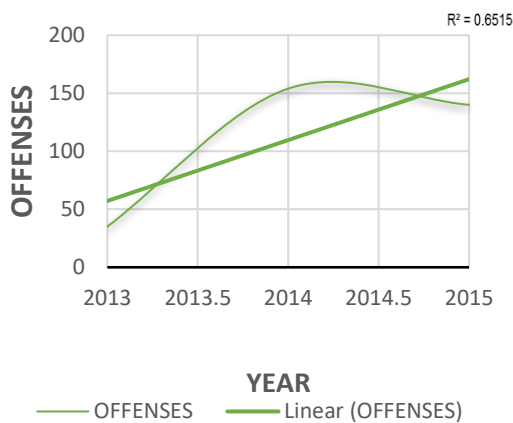

Figure 1: Trend Line for Offenses in Binalonan from 2013-2015

As shown in figure 1, the graph shows that there is no sign of seasonality with an upward trend, because the fluctuation of the series is rising and the value of R-squared is 0.6515, which is a good fit of the line to the data, therefore, increasing trend exists.

In the Distribution of Characteristics Profile of respondent according to place, there were 329 overall total of offenses committed in Binalonan. From Binalonan to San Manuel Road, there were 88 or 62.8% of the respondents who committed reckless imprudence resulting to damage to property with the total of 140 offenses committed on that place. From Binalonan to Laoac Road, there were 19 or 73.1% of the respondents who committed reckless imprudence resulting to physical injury with the total of 26 offenses committed on that place. From Binalonan to Pozurrubio Road, there were 70 or 69.3% of the respondents who committed reckless imprudence resulting to damage to property with the total of 101 offenses committed on that

place. From Binalonan to Urdaneta Road, there were 37 or 59.7% of the respondents who committed reckless imprudence resulting to damage to property with the total of 62 offenses committed on that place.

RELATIONSHIP OF CHARACTERISTICS PROFILE OF The Respondent And Offense Committed
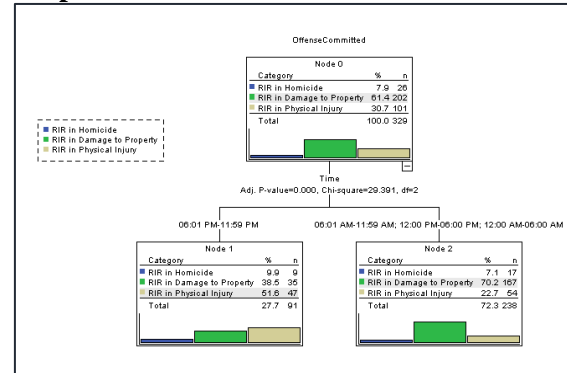

Figure 2: Classification Tree of Offense Committed and the Characteristics Profile of the Respondent as Predictors

Based on Figure 2, by using CHAID method, the tree shows that among all the characteristics profile of the respondents only time was detected as significant and best predictor of the offense committed with Chi-Square (χ^2) =29.391 and significance value (P-value) that is equal to 0.000.

The root node (node 0) shows that 202 or 61.4% of the respondents committed a violation of reckless imprudence resulting to damage to property out of 329 respondents. The root node (node 1) shows that 47 or 51.6% of them committed a violation of reckless imprudence resulting to physical injury during night time from 06:01 PM to 11:59 PM out of 329 respondents due to the dark environment caused by lack of road lights and more drivers involved in alcohol. On the other hand, 238 out of 329 committed a violation of reckless imprudence resulting to damage to property during day time from 12:00 AM to 06:00 PM due to an increasing number of vehicles traveled than night.

Table 1
Target Category: Homicide
Gains for Nodes

Node	Node		Gain		Response	Index
	N	%	N	%		
1	91	27.7%	9	34.6%	9.9%	125.1%
2	238	72.3%	17	65.4%	7.1%	90.4%
Total	329	100%	26	100%		
Growing Method: CHAID						
Dependent Variable: OFFENSE_COMMITTED						

Table 1 shows the summary of information about the terminal nodes in the model. The terminal nodes show the frequency and percentage of those who committed homicide with time as predictor. Based from node 1, 9 or 9.9% committed a reckless imprudence resulting to homicide. On the other hand, according to node 2, 17 or 7.1% committed reckless imprudence resulting to homicide in terms of time.
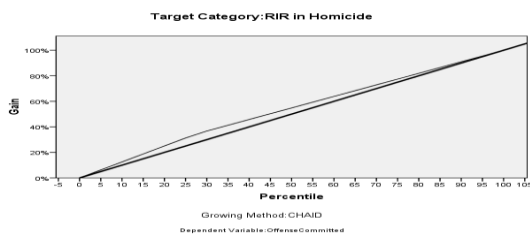

Figure 3: Graph Showing the Gain

The graph in Figure 3 shows that it begins from 0% and ends at 100% as it goes from one end to the other. It implies that the model is fairly good one since the gain chart rose suddenly toward 100% and then leveled off.
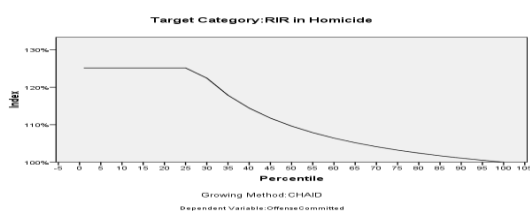

Figure 4: Graph Showing the Index

The graph in Figure 4 shows the cumulative index chart. A cumulative index chart tends to start above 100% and gradually falls until it reaches 100%. For a good model, the index value should start well above 100%, remain on a high plateau as the line moves along, and then trail off sharply toward 100%. Since figure 11 shows these characteristics then it is a good model.

Table 2
Target Category: Damage to Property
Gains for Nodes

Node	Node		Gain		Response	Index
	N	%	N	%		
2	238	72.3%	167	82.7%	70.2%	114.3%
1	91	27.7%	35	17.3%	38.5%	62.6%
Total	329	100%	202	100%		
Growing Method: CHAID						
Dependent Variable: OFFENSE_COMMITTED						

Table 2 shows the summary of information about the terminal nodes in the model. The terminal nodes show the frequency and percentage of those who committed to damage to property with time as predictor. Based from node 1, 35 or 38.5% committed reckless imprudence resulting to damage to Property. On the other hand, according to node 2, 167 or 82.7% committed reckless imprudence resulting to damage to Property in terms of time.
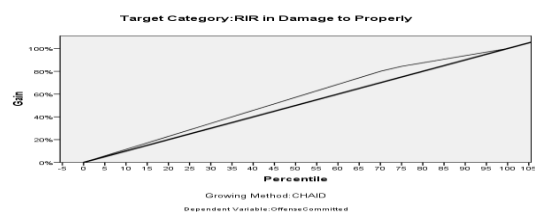

Figure 5: Graph Showing the Gain

The graph in Figure 5 shows that the line begins from 0% and ends at 100% as it goes from one end to the other. It implies that the model is fairly good one since the gain chart rose suddenly toward 100% and then leveled off.
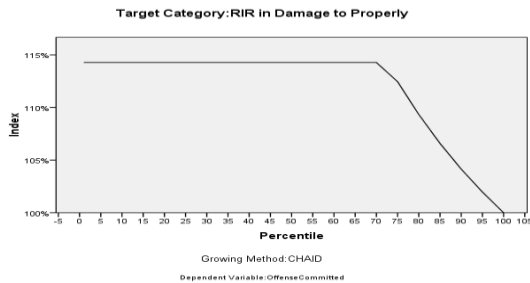

Figure 6: Graph Showing the Index

The graph in Figure 6 shows the cumulative index chart. A cumulative index chart tends to start above 100% and gradually falls until it reaches 100%. For a good model, the index value should start well above 100%, remain on a high plateau as the line moves along, and then trail off sharply toward 100%. Since figure 13 shows these characteristics then it is a good model.

Table 3
Target Category: Physical Injury
Gains for Nodes

Node	Node		Gain		Response	Index
	N	%	N	%		
1	91	27.7%	47	46.5%	51.6%	168.2%
2	238	72.3%	54	53.5%	22.7%	73.9%
Total	329	100%	101	100%		

Growing Method: CHAID
Dependent Variable: OFFENSE_COMMITTED

Table 3 shows the summary of information about the terminal nodes in the model. The terminal nodes show the frequency and percentage of those who committed physical injury with time as predictor. Based from node 1, 47 or 51.6% committed reckless imprudence resulting to physical injury. On the other hand, according to node 2, 53 or 22.7% committed reckless imprudence resulting to physical injury according to time.
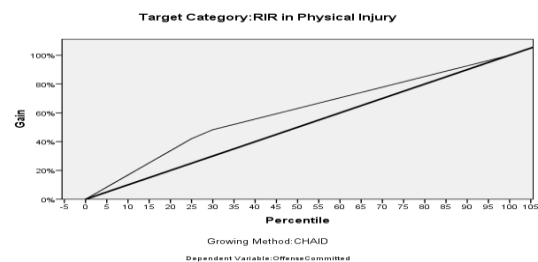

Figure 7: Graph Showing the Gain

The graph in Figure 7 shows that the line begins from 0% and ends at 100% as it goes from one end to the other. It implies that the model is fairly good one since the gain chart rose suddenly toward 100% and then leveled off.
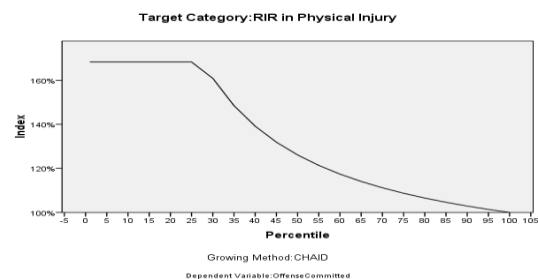

Figure 8: Graph Showing the Index

The graph in Figure 8 shows the cumulative index chart. A cumulative index chart tends to start above 100% and gradually falls until it reaches 100%. For a good model, the index value should start well above 100%, remain on a high plateau as the line moves along, and then trail off sharply toward 100%. Since figure 15 shows these characteristics then it is a good model.

Table 4
Risk Estimate of Offense Committed

Estimate	Std. Error
.350	.026
Growing Method: CHAID	
Dependent Variable: OFFENSE_COMMITTED	

Table 4 shows the risk estimate of 0.350 which indicates that the category predicted by the model (homicide, damage to property and physical injury) is wrong for 35% of the cases. So approximately, 35% of the respondents are misclassified with standard error equal to 0.026

Table 5
Classification of Offense Committed

Observed	Predicted			
	Homicide	Damage to Property	Physical Injury	Percent Correct
Homicide	0	17	9	0.0%
Damage to Property	0	167	35	82.7%
Physical Injury	0	54	47	46.5%
Overall Percentage	0.0%	72.3%	27.7%	65.0%
Growing Method: CHAID				
Dependent Variable: OFFENSE_COMMITTED				

Table 5 shows the correct classification done by CHAID. Sixty-five percent (65.0%) of the cases were classified correctly. As shown on the table, 82.7% from damage to property and 46.5% from physical injury were classified correctly. On the other hand, 72.3% of offenses committed predicted by time are damage to property and 27.7 % of offense committed predicted by time is physical injury.

V. CONCLUSIONS AND RECOMMENDATIONS

Based on the findings, majority of the offenses committed cases have increasing probability that the offender on vehicular accidents end up from reckless imprudence resulting to damage to property.

There is a significant relationship between the characteristics profile of the respondents and the offense committed due to vehicular accidents.

Among all the characteristics profile of the respondents only time was detected as best predictor of the offense committed due to vehicular accidents.

However, National Policy should do a better idea and strategy in terms of time which is greatly related to vehicular accidents in order to reduce the offense committed due to vehicular accidents in particular location for the preparation of the future prediction.

VI. REFERENCES

Abdel-Aty, Mohamed, & Abdelwahab, H. (2004). Modeling Rear-end Collisions Including the Role of Driver's Visibility and Light Truck Vehicles using a Nested Logit Structure. Accident Analysis and Prevention, 36(3), 447–56.

Chang, L. Y., & Mannering, F. (1999). Analysis of Injury Severity and Vehicle Occupancy in Truck- and Non-Truck-Involved Accidents. Accident Analysis and Prevention, 31(5), 579–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10440555

De Lapparent, M., (2006). "Empirical Bayesian Analysis of Accident Severity for Motorcyclists in Large French Urban Areas". Accident Analysis and Prevention. 38 (2), 260-268.

Donate-Lopez, C., Espigares-Rodriguez, E., Jimenez-Moleon, J.J., Luna-del-Castillo, J.d.D., Bueno-Cavanillas, A., Lardelli-Claret, P., (2010). The Association of Age, Sex and Helmet use with the risk of Death for Occupants of Two- Wheeled motor vehicles involved in traffic crashes in Spain.

Accident Analysis and Prevention. 42 (1), 297-306.

Laserna (2013). Reckless Imprudence Definition. Retrieved from http://www.attylaserna.blogspot.com/2013/12/reckless-imprudence-right-of-way.html

Lev, D., Hershkovitz, E., Yechiam, E. (2008). Decision Making and Personality in Traffic Offenders: A Study of Israeli Drivers. Accident Analysis and Prevention. 40 (1), 223-230

March (2016). Chi-Square Automatic Interaction Detector Analysis. Retrieved from http://www.TREE-CHAID.pdf

Oppong (2012). Road Accidents in Ghana.. Retrieved from http://www.ir.knust.edu.gh/bitstream/123456789/4544/1/FINAL%20THESIS%202012,%20Oppong,%20Richard%20Asumadu.pdf

Savolainen, P., Mannering, F. (2007). Probabilistic Models of Motorcyclists' Injury Severities in Single- and Multi-Vehicle Crashes. Accident Analysis and Prevention. Vol. 39 (5), 955-963