International Journal of Engineering &Technology, 7 (4) (2018) 6313-6321

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.29057
Research paper

Factors influencing Malaysian drivers' tendency on underreporting

Nik Sarah Nik Zamri , Zamira Hasanah Zamzuri *, Kamarulzaman Ibrahim

School of Mathematical Sciences, Faculty of Science & Technology, Universiti Kebangsaan Malaysia
**Corresponding author E-mail: zamira@ukm.edu.my*

Abstract

The reported accidents data on sites is essential to measure the effectiveness of any road safety policy. If an accident remains unreported, the vital information about the characteristics of the accident will be lost. The choice of appropriate safety measures may, in turn, be adversely affected by the loss of the data. This current study is aimed to determine the reasons for the underreporting of accidents in Malaysia. Random samples comprising 2000 individuals were interviewed in person and data on their demographic background and experience involved in accidents were recorded. Out of 1422 accidents, 662 accidents were recorded as unreported accidents. The findings indicate that in general, Malays aged between 60 and 69, living in rural area, who are single and are motorcyclists tend not to report the accidents.

Keywords: *Underreporting; Road Accidents; Malaysia.*

1. Introduction

In recent year, the occurrence of road traffic crashes has become an important global issue as it causes high mortality rate. World Health Organisation estimated that almost 1.25 million people die each year due to road traffic crash. Evidence revealed that up to 20-50 million people undergo non-deadly injuries in which a number of them experienced some form of impairment due to the trauma (WHO 2018). Road accidents result in financial wastes to sufferers, their relations, and the country in general. Huge amount of expenditure is required for treatment, rehabilitation and case inquiry and decrease in productivity (e.g. in wages) faced by the sufferers who deceased or disabled because of injuries, as well as their relations who require leave for absence (or school) in order to look after the injured. Road traffic crashes significantly influence human lives and economy. It has been projected that the worldwide global loss caused by traffic injuries almost around 518 billion US dollars. This amount is approximately equivalent to 1 to 4% of the gross national products (Note and Lanka, 2014).

The problem of not reporting all reportable accidents could contribute to inaccuracy in estimating the true accident rate in a particular site. Accurate and comprehensive accident data are required to establish a road safety research program (Kowtanapanich et. al 2007). Consequently, the treatment provided for the site may not be effective. Over 20 years back, Hauer and Hakkert revealed that a complete list of traffic crashes was lacking (Hauer and Hakkert, 1988). The investigation of underreporting crash data resulted in partial approximation with the utilization of crash prediction models, which consequently lead to futile treatments. Previous reports demonstrated that crashes were underreported in all developed nations, however, the rate of unreporting is poorer in developing nations. On that account, several factors including age of the victim, function of the victim (e.g., role as the driver or the passenger, number of crashed vehicles and severity of the crash affect the likelihood of reporting involved (Hauer and Hakkert, 1988). Majority of the evaluation associated with road safety are relied upon the reported crash statistics. Hence, the problem of underreporting could restrict the capability to regulate the road safety. Commonly, there is a concern on the imprecision of the road traffic injury data due to underreporting or failure to report. The information relating to traffic accidents and victims is an important tool to control and detect road safety problems. For instance, a French regional research, which evaluated the difference between hospital and road traffic police data revealed that the data reported by the police was merely 37 percent of non-deadly hospitalized injuries (Aptel et al., 1999).

A study conducted in Australia indicated that with the utilization of a linked data system, almost 40–45 % of hospital records for road crash casualties lack matching police record (Rosman, 2001). Similarly, estimation of child road accident injuries in Japan via the linkage of Police and Insurance Associations records was much the same with Australia (Nakahara and Wakai, 2001). Moreover, a countrywide underreporting study conducted in the UK (Ward et al., 2006) reported a significant dissimilarity of underreporting between diverse stages of injury severity. In addition, the study revealed that approximately 20 percent of injuries categorized as severe by the Police were treated and discharged (i.e. slight injuries), whereas those treated by the hospital as severe (around 8%) were recorded as minor in the police record. In addition, several studies demonstrated key inconsistencies with regard to the degree of underreporting for each category of road user. Around 20 % of injured pedestrians underreported are in San Francisco, United States of America (Scortiono et al., 2005). Besides, underreporting of road accident injuries by police in France is greater amid the younger (0–17) age compared to the elder (55+ years) age categories (Amoros et al., 2006). Meanwhile, it has been projected that there is an expansion of underreporting of bicyclists and motorcyclists involved accidents in the United Kingdom Cryer et al. (2001).

Copyright ©2018 Nik Sarah Nik Zamri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

One of the earliest studies conducted based on a probability sampling of emergency room visits to 42 hospitals was in northeastern Ohio in the United States of America (Barancik and Fife., 1985). The researchers demonstrated that the police crash reports for only 55 percent of the patients treated for injuries received in a motor vehicle crash; among those hospitalized, 74 percent were matched. Furthermore, there were no report on discrete data aimed at bicyclists and pedestrians. A current research conducted on bicyclists whom were treated in a cross-section of hospital emergency wards located in North Carolina revealed that police reports were available for 60 % of motor vehicle crashes in Western Australia (Stutts et al., 1990). Moreover, the study demonstrated that 69% and 74% of pedestrians and bicyclists, respectively were admitted to hospitals had official police records.

Next, Santos et al., (2017) studied on traffic accidents in city of Belo Horizonte, Brazil. They compared the official traffic accident data to unofficial data collected from Waze, a mobile app that help users to avoid traffic congestion and they found that 7% of accidents reported by official sources also have been reported by users of Waze but most of the accidents reported by Waze were classified as having lower severity hence implied that most of lower severity accidents are not officially reported.

In Malaysia, the police and hospital records were collected to determine their matching and reporting rates in the Melaka Tengah district in the state of Melaka (Kamaluddin et al., 2018). They found that the hospital database contains 86% of police cases, while the police database contains only 4.1% of hospital cases. Both the reporting and matching rates increased with the level of injury severity. Another study by Mustaffa & Hokao (2012) was that they developed a database system for road traffic accident by integrate data from four cores involving in accident database which are hospital, police, fire and rescue and the civil defense in the state of Johor Bahru in Malaysia. The application of the research decreases the accident fatalities.

Road safety has been recognized as a long-established social accountabilities of the Malaysian Government and public (Mustafa, 2005). Fundamentally, road safety has been established within the government departments, private agencies and voluntary organisations. The main strategy is through the Cabinet Committee of Road Safety chaired by the Prime Minister. Moreover, the establishment of the National Road Safety Plan aided in the research related to road safety, behavioural modification of road users, road engineering and vehicle safety as well as medical treatment and safety administration.

There have been rapid population and economic growth industrialisation and motorisation since 1970 in Malaysia. These factors have significantly associated with rise in road accidents rate in Malaysia. According to a press release in the New Straits Times, the Transport Minister reported that 7152 people had died in road accidents in 2016. He also cited that the agonizing truth is that the rate of road accidents is rising annually (Babulal, 2017).

The key components of a successful road safety improvement program are the availability and the knowledge of basic facts about road accidents. In particular, essential pieces of information concerning the where, when, why and who of road traffic accidents are necessary prior to implementing any effective solution (OECD 1984).

Underreporting or failure to report the road traffic accidents has frequently increased the attention on the imprecision of data and its impact on road safety policy-making and development. The World Health Report has highlighted the necessity for precise and comprehensive information and scientific methodologies with regard to the prevention and control of road traffic injuries (Peden, 2004). A previous study demonstrated that generally, the exact figure of road crashes is indefinite, and practically entire studies of road crashes comprising greater than single form of data compare only two sources including police and hospital records (Elvick and Mysen.,1999). In this case, the level of comprehensiveness of these datasets is incomplete. The aforementioned two sources fail to include those who do not go to the hospital or to the police, causing in an additional underestimation of underreporting. Evidence has indicated that community-based studies usually provide precise death and injury rates (Sethi et al., 2004). A survey conducted by Ibrahim and Silcock (1992) including the entire highway authorities in England with regard to issue of underreporting of accidents. The study reveals that the highway authorities have contributed effectively for inspecting and rectifying the accident data.

Recent studies by Samuel et al (2012) demonstrated the underreporting of road traffic damages and mortality rate in Malawi. The study applied capture-recapture analysis for approximating the incidence of road traffic deaths by means of two partial data sources. Nevertheless, the authors proposed spatial analyses or third data source to enhance such estimates. Moreover, a previous research applied the capture-recapture technique with the use of traffic police and hospital injury surveillance statistics (Abegaz et al (2014). Neither of the two sources autonomously contributed towards precise coverage of road traffic incident linked to deaths and injuries. Furthermore, reinforcement both systems is obligatory for acquiring precise information on road accidents and human casualties in Ethiopia.

Periyasamy et al (2013) performed a community survey to approximate and compare the underreported road traffic injuries and its degree. Moreover, comparison was performed between features of reported and unreported to the police in district of Kandy, Sri Lanka. The results of underreporting ranging from 33% to 56% in the district of Kandy indicate the actual problem of underestimation of traffic injury. Underreporting was highly evident among non-deadly, non-critical injuries, high-income and middle-aged vulnerable road users. In general, underreporting is affected by the habits to elude legal procedures as well as time or money concerns. The current study emphasized the importance of improving the awareness on reporting. Moreover, this study is also aimed to form a conducive environment to increase reporting for establishing accurate data for effective and well-organized policy development.

A study was conducted in Hong Kong on miscategorization with regard to severity of injury amongst road fatalities in police reports (Tsui et al., 2009). The findings revealed that age, the injury severity scale and position of the victim largely influence the probability of police injury misclassification. The study suggested that the accurate and widespread data on injury features accessible in trauma records should be utilized for the documentation and measurement of preconceptions with regard to injury severity grading.

Several studies conducted in Malaysia revealed that incompleteness of accidents records due to underreporting of accidents is a main issue leading to increased accident rate. Manan & Varhelyi (2012) studied that serious injuries were not reported up to 6 times whereas slight injuries up to 14 times. The Ministry of Transport (2016) stated that 7152 people died in road accidents in Malaysia. A previous study conducted on motorcycle fatalities revealed that the motorcyclists are associated with greater than 50% of road accident fatalities in Malaysia (Manan & Varhelyi, 2012). Notably, Malaysia is the nation with highest road fatality risk (per 100,000 population) amongst the ASEAN nations. Thus, accident reporting in Malaysia should be enhanced through developing an injury recording system and database built on hospital records. This is crucial to match police records for acquiring precise findings on injuries.

The survey method depends on data, which are generated through human memory. On that account, (Schacter, 1999) has cautioned that the human memory is frail. A number of studies dismissed the utilization of self-reported data because of its unproven poor quality. Nonetheless, Chan (2009) claimed that the alleged poor quality of self-reported data is solely a metropolitan legend. The respondents may reveal imprecise data in some instances as they are compelled by social interest, however, this could be a rare scenario. For instance, it is not likely that the respondents would provide fake information regarding their demographics, like gender and ethnicity. Secondly, although it has been evidenced that respondents incline to provide false information in the investiga-

tional studies, this problem is uncommon in methods applied in field studies and naturalistic settings. Additionally, there are many successful self-reported methods of distinct psychological constructs that have acquired to build validity proof via both convergent and discriminant validation. For instance, affectivity disposition, Big five personality traits, proactive personality, self-efficacy, goal orientations, perceived organizational support, and many others. A survey involving community and road traffic safety practitioners have also been carried out, in addition to linking of police and hospital records as a commonly used approach for estimating the average reporting rates of accidents.

The specific aims of this research are:

a) To identify the characteristics of Malaysian drivers and their unreported road accidents.
b) To determine the influences of drivers' variables on their reporting or unreported of road traffic accidents.
c) To investigate the reasons for unreported road accidents.

2. Method

The current study applied the stratified based sampling method for a general representation of Malaysian population. Based on the stratified sampling (stratified sampling 2018), the Malaysian residents are subdivided into a number of non-overlapping groups, namely strata. As such, a sample is selected based on the designated design within individual stratum. The samples were stratified into variables such as age, ethnics and gender. The study was carried out from 2016 to 2017. The questionnaire developed consisted of two sections all of which involved in motor vehicle accidents. Section A is a demographic section in which the variables including age, gender, ethnics, religion, residential location, residential, marital status, years of driving and average of driving were measured. Section B consists of the driver's behaviour, in which the variables involved are the frequency of accidents encountered by the drivers as well as the details of the accident. These include year of accidents, type of road users, vehicle damaged, injury, death, the presence of objects, cause of the accidents, type of road, hospitalized and whether the accident occurred at a known location. Information on gender and age of the casualties was also obtained. The study data consists of 2000 respondents, which are all motor-vehicle involved accidents. In addition to the information on the severity of the accident, data regarding vehicle type, location of the accidents, road condition and objects (if any) involved were also obtained. Based on all the data collected, two sets of data were developed, which are driver dataset and accident dataset. The driver dataset was used to analyse the demography of the drivers, the number of accidents reported and the number of unreported accidents. The accident dataset was used to analyze the criteria for each accident.

Table 1: Demographic Characteristics of Malaysian General Population and Current Research Data

Factor		Percentage for Malaysian Population (%)	Percentage from survey (%)
Age	15-64	69.7	100
Ethnicity	Malay	68.8	67.7
	Chinese	23.2	21.2
	Indians	7.0	8.9
	Others	1.0	2.2
Gender	Male	51	48
	Female	49	52

The current study data is relatively a general representation of the Malaysian population data. This is because the percentage of ethnicity and gender distribution were relatively similar. It should be noted that percentage of Chinese subjects in the current survey was slightly smaller compared to Malaysian population. In contrast, the percentage of Indian and others were slightly larger in the current study compared to general population. Moreover, the female participants were higher than male. It should be noted that all the study participants were aged 15-64. In this sample survey, sampling procedures is intended to enhance the reliability of observations, facilitate replication studies and permit statistical analysis of data and generalizations to larger populations (McClintock, 1974).

Table 2: Descriptive Statistics on Driver Dataset

Factor		Percentage (%)
Age	18-29	76.5
	30-39	6.0
	40-49	2.6
	50-59	1.6
	60-69	13.3
Gender	Male	48
	Female	52
Ethnicity	Malay	67.7
	Chinese	21.2
	Indian	8.9
	Others	2.2
Religion	Islam	70.1
	Hindu	7.9
	Christian	4.3
	Buddha	17.3
	Others	0.4
Residential Location	City	65.4
	Rural	34.6
Race by residents	Majority Malay	56.6
	Majority Chinese	8.0
	Majority Indian	3.0
	Ethnics mixtures	32.4
Marital status	Single	74.3
	Married	25.7

Table 2 demonstrates that the participants aged between 18 and 29 has the highest participation. This could be due to the fact that the current study focuses more on people in this age group. Moreover, this study had more female subjects than male subjects. Higher number of study participants were concentrated in the city area compared to rural area. Furthermore, there were no single individuals compared to the married.

Table 3: General Characteristics of the Road Accident Data

Factor		Percentage (%)
Type of road users	Drivers	62.9
	Motorcyclist	15.0
	Passengers	19.8
	Cyclist	0.5
	Pedestrians	1.8
Vehicle damaged	Yes	81.9
	No	18.1
Injury	Yes	32.8
	No	67.2
Hospitalized	Yes	28.2
	No	71.8
Objects	Yes	82.9
	No	17.1
Death	Yes	1.2
	No	98.8
Type of roads	Road junctions	38.7
	Straight lane road	29.5
	T-junction	15.2
	Highway	9.8
	Corner	4.1
	Roundabout	1.5
	Others	1.2

Table 3 demonstrates the general characteristics of the accident data. The results indicate that most of the users are car drivers and higher accident occurred at road junctions.

A logistic regression model was used to estimate probability of unreported accidents. A simple logistic model has the form:

$$logit\ (Y) = ln(odds) = \ ln\frac{\pi}{1-\pi} = \beta_0 \ + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_i x_i \ for\ i = n$$

Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables. It is a form of binomial regression and it is used to conduct when the dependent variables is dichotomous (binary). Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

Confusion matrix was used to validate the performance of the model and to obtain the error rate. It is a summary of results on a classification problem. The confusion matrix shows the ways in which classification model is confused when it makes predictions. It is a two-dimensional matrix, indexed in one dimension by the true class of an object and in the other by the class that the classifier assigns. We can summarize the table as below:

	Event	No event
Event	True positive	False positive
No event	False negative	True negative

This gives us:

a) true positive for correctly predicted event values
b) False positive for incorrectly predicted event values.
c) True negative for correctly predicted no-event values.
d) False negative for incorrectly predicted no-event values.

3. Discussion

Figure 1 demonstrates the unreported accidents and reported accidents. The drivers' and accident dataset were used in this study. From the drivers' dataset, the total accidents were 1422 out of 2068. Out of 1422 total of accidents, 662 accidents were recorded as unreported accidents and 760 were recorded as reported accidents.

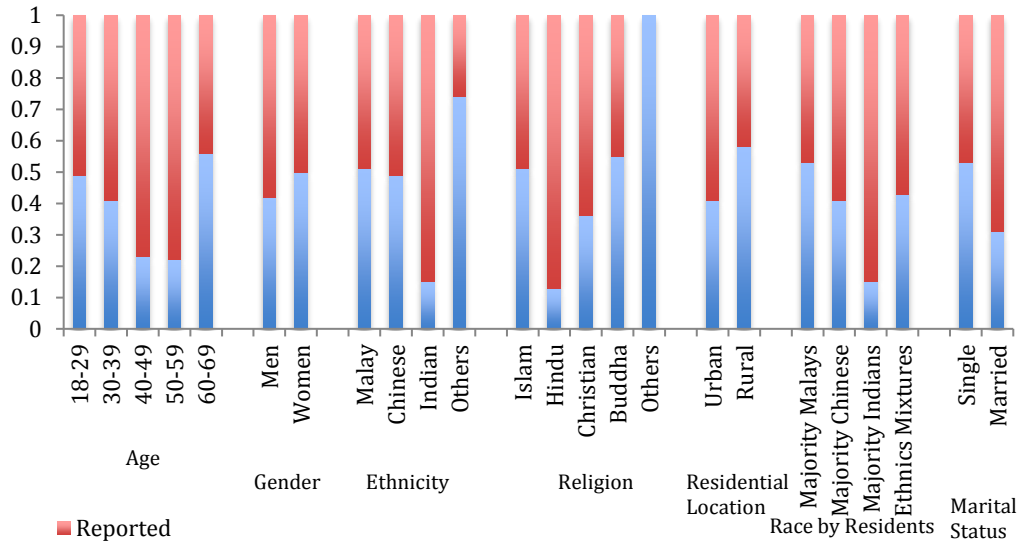


Fig. 1: Profile of the Unreported and Reported Accidents Based on Drivers' Attribute.

Based on Figure 1, people aged between 60 and 69 had the highest tendency to not report the accident (0.56) followed by people age between 18-29 (0.49). Disparities in the driving behavior could be due to various behavioral, physiological and psychological factors, which eventually caused the variation in the accident injuries of distinct age categories. For instance, parallel with increase in the age, there might be greater driving experience and intensities of acceptable risk. Moreover, as the results of ageing, the physical/ reflex degradation problems could also lead to increased reaction/perception duration, faster onset of fatigue, decline in the capability to process data, decrease in strength and mobility, and degradation of vision and hearing (Smith et al., 1993). The gender subgroup analysis revealed that female drivers are inclined to not report the accidents (0.49) compared to the male drivers (0.42). Findings on the effect of gender with regard to accident reporting rates is inconsistent. As such, a previous study demonstrated that there was no significant difference in reporting rates between men and women (Barancik and Fife., 1985; Aptel et al., 1999). Nevertheless, Rosman and Knuiman (1994) reported that male casualties had significantly lesser reporting rates. In contrast, there were no significant differences with regard to gender and ethnic subgroups with respect to reporting rates in another study by Rosman and Knuiman's. According to ethnic and religion based subgroup analyses, other and Islam groups had the highest frequency of not reporting the accidents. A higher percentage (85.1%) of Indian ethnics, who are Hindus tend to report the accidents. Additionally, those living in the rural areas tend to not report the accidents (0.58). This could be linked to low road safety and enforcement in rural areas (Manan., 2012). Helmet compliance is also quite low in the rural areas. Besides that, people who live in area populated by majority of Malays had no report on the accidents. Furthermore, single individuals had the highest frequency to not report the accidents in this study. This finding was inconsistent with the study by Shinara et al. (2016), which revealed that married person tend not to report the accidents. This possibly because a driver with a partner tend to be influenced by the partner, meanwhile single individuals are most likely to make official report to the police regarding the accidents to obtain assistance and support (Shinara et al, 2016).

Figure 2 illustrates that highest frequency of motorcyclists had not reported the accidents followed by the car drivers. Data on crash have shown that bicycles users had higher underreporting rate in contrast remaining road users. This is because bicycle users are frequently not documented by the police as the accidents occur without the motor vehicles. For example, this scenario is common when a cyclist smashes a fixed object or falls, on or off the road (DeCamposa et al., 2016). Despite of involvement of a motor vehicle, most of the crashes are not documented in police records. A year prospective research data on the cycling behavior and crashes of 1087 adult commuter riders in Brussels, Belgium revealed that the police only documented 7% of the crashes despite 19% of crashes involved a collision with a car (De Geus et al., 2012). In metropolitan such as Munster, Germany, which has many bicycle users than drivers, the hospital admission records comprise double amount of people with injury due to bicycle crashes compared to the police records (Juhra et al., 2012). The accidents were reported (64.3%) only if there is a risk of injury, whereas if injury did not occur, the accidents are likely to be not

be reported (57%). Non-deadly injuries were underreported mainly because of limited understanding with regard to significance of reporting by the injured road users (Periyasamy et al., 2013).

The main concern of road crash sufferers for their unwillingness to make police report is to elude the penalties for the crash if they were at fault. Furthermore, roundabout had the highest number of unreported accidents while the number of reported accidents are the highest at the highway. Evidence has revealed that divided highways have the lower chance of occurrence of high severity crashes (Patil et al., 2012). Nonetheless, accidents still occur in the divided highways and being reported, however, the occurrence is still lower than the accidents on road junctions. It should be noted that there were only 14 accidents that caused death in this study in which 5 deaths for unreported accidents and 9 deaths for reported accidents. For the unreported accidents, the death may possibly had happened more than 30 days after accidents.

Over the past decade, enabling an enhanced protection for the vehicle occupants suffering from side-impact collisions has emerged as the main concern of government, industry and academic researchers. Moreover, the side-impact issue emphasizes on the side-impacts with immovable roadside obstacles including guard rails, trees and utility poles. A number of these obstacles such as trees are considered as natural objects found by the roadside. Additional objects such as utility poles are positioned sideways of the roadside yet these objects are not required for roadway function. Furthermore, several objects such as luminaire supports, guard rails and signs, are positioned along the roadway for an explicit objective associated with the roadway function. All the objects that are positioned alongside the roadside should be assessed via full-scale vehicle crash tests to assess the efficiency of the object. This is essential to reduce the hazards to vehicle occupants who may possibly strike the object (Ross et al., 1993). Absence of objects influence the drivers to not report any accidents. A previous study revealed that poles (together with light poles, overhead sign poles, railway, trees, utility and traffic) suffered a substantial number of crashes (11.7%) (Holdridge, 2005). Generally, most of the reported accidents are likely to be hospitalized. If no hospitalization, no accidents will be reported. Notably, traffic accidents have caused hospitalizations with high costs (Baker et al., 1974). Therefore, people tend to neglect reporting of the accidents if there is no hospitalization.

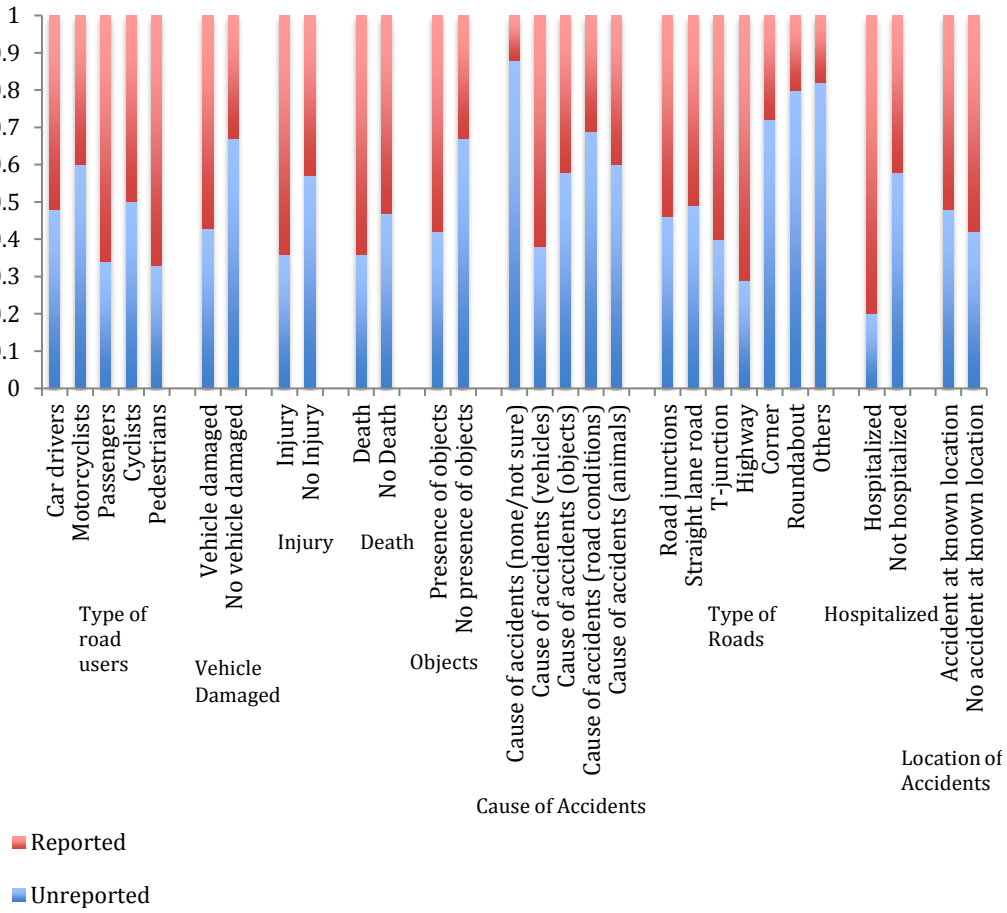


Fig. 2: Profile of Unreported and Reported Accidents Based on the Accidents Attributes.

Logistic regression

A number of previous studies on the highway safety area have applied logistic regression models to identify various factors. In one such study in Iowa evaluated the injury severity in head-on highway crashes, where age and gender were investigated as predictors (Mercier et al., 1997). Furthermore, other research conducted in Hawaii demonstrated that the personal and behavioral aspects of the predictors linked to automobile crash and injury severity (Kim et al., 1995). Additionally, another study, which was also conducted in Hawaii applied logistic regression models to identify demographic and temporal factors associated with impaired motorcycle crashes (Kim et al., 2000). The study also evaluated the predictors related to utilization of safety belt amongst drivers who involved in crash and front seat passengers. This study also includes misclassification errors in the outcome variable (Let et al., 1999). Furthermore, a study carried out in Pennsylvania on run-off-road crashes applied a logistic regression modelling to approximate the impacts of driver, environmental and highway factors (McGinnis et al., 1999).

A logistic regression model was fitted into accident dataset consisting of a previously specified set of fifteen dichotomous predictors: age, gender, ethnicity, religion, residential locations, race of residents, marital status, type of road users, vehicle damaged, total loss, the presence of objects, type of roads, hospitalized, location of the accident and summon (Steyerberg, 2001). The logistic model was fitted with all the variables in which the AIC and deviance were recorded. The process was repeated until all the variables were significant. On that account, the Akaike Information Criteria

(AIC) enables evaluation of the quality of model. It should be noted that the model is good if the value of AIC is small. In this case, deviance is used to assess of the goodness of fit of a generalized linear model.

The variables are labeled as below:

x_1 : Location of accidents
x_2 : Vehicle damaged
x_3 : Hospitalized
x_4 : Total loss
x_5 : Objects involved
x_6 : Marital status
x_7 : Age
x_8 : Gender
x_9 : Ethnicity
x_{10} : Religion
x_{11} : Residential location
x_{12} : Race of residents
x_{13} : Type of road users
x_{14} : Type of roads
x_{15} : Summons

The model for the variables are:

M0 : $y \sim \beta_0$
M1 : $y \sim \beta_0 + \beta_1 x_1$
M2 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2$
M3 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$
M4 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$
M5 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5$
M6 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6$
M7 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7$
M8 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8$
M9 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9$
M10 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10}$
M11 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11}$
M12 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12}$
M13 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13}$
M14 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14}$
M15 : $y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8 + \beta_9 x_9 + \beta_{10} x_{10} + \beta_{11} x_{11} + \beta_{12} x_{12} + \beta_{13} x_{13} + \beta_{14} x_{14} + \beta_{15} x_{15}$

Table 4: The Best Subsets for Number of Variables

Number of Variables	Number of combinations	The best subsets	AIC	Deviances	The significant variables
1	15	x_5	2068.9	2064.9	Objects involved
2	105	x_4,x_5	2126.3	2100.2	Total loss, Objects involved
3	455	x_2,x_4,x_5	1966.2	1961.3	Vehicle damaged, Total loss, Objects involved
4	1365	x_2,x_3,x_4,x_5	1834.5	1821.0	Vehicle damaged, Hospitalized, Total loss, Objects involved
5	3003	x_1,x_2,x_3,x_4,x_5	1724.0	1715.4	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved
6	5005	x_1,x_2,x_3,x_4,x_5,x_6	1712.0	1698.0	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital status
7	6435	x_1,x_2,x_3,x_4,x_5,x_6,x_7	1659.4	1638.5	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital status
8	6435	x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8	1643.8	1634.3	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital status, Gender
9	5005	x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{11}	1591.2	1562.9	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital status, Gender, Residential location
10	3003	x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{11},x_{15}	1582.2	1544.3	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital status, Gender, Residential location
11	1365	x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{11},x_{14},x_{15}	1592.1	1543.2	Hospitalized,

12	455	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{11},x_{13},x_{14},x_{15}$	1575.3	1529.6	Total loss, Objects involved, Marital status, Residential location Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital status Vehicle damaged, Hospitalized,
13	105	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{10},x_{11},x_{13},x_{14},x_{15}$	1431.1	1374.3	Total loss, Marital status, Gender, Residential location Vehicle damaged, Hospitalized,
14	15	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_{10},x_{11},x_{12},x_{13},x_{14},x_{15}$	1426.3	1356.3	Total loss, Objects involved, Marital status, Gender, Residential location Vehicle damaged, Total loss,
15	1	$x_1,x_2,x_3,x_4,x_5,x_6,x_7,x_8,x_9,x_{10},x_{11},x_{12},x_{13},x_{14},x_{15}$	1426.3	1354.3	Objects involved, Age, Religion, Residential location, Race of residents, Type of roads

Table 4 shows the best subsets for the number of each variables. In the table above, the variables that are significant is the model with six number of variables. Although the AIC is not the lowest, but it is still the best model because of all the variables are significant. Although the lowest AIC is 1426.3 which is model 15, but we cannot take this as a good model because not all the variables are significant. Hence, we choose the model 6 as the best model since all the variables are significant which is location of accidents, vehicle damaged, hospitalized, total loss, objects involved and marital status.

Table 5 demonstrates the ranking of significant variables that influenced the status of underreporting.

Table 5: Ranking of the Significant Model

Model	AIC	Reduction in AIC ($AICm_i$ - $AICm_{i+1}$)	Deviances	Difference in deviances ($devm_i$ - $devm_{i+1}$)	Ranking	New Variables Added
M0	2866.9	-	2866.9	-	-	-
M1	2495.7	371.2	2491.7	375.2	1	Location of accidents
M2	2126.2	369.5	2120.2	371.5	2	Location of accidents, Vehicle damaged
M3	1969.2	157.0	1961.2	159.0	3	Location of accidents, Vehicle damaged, Hospitalized
M4	1830.5	138.7	1820.5	140.7	4	Location of accidents, Vehicle damaged, Hospitalized, Total loss
M5	1726.4	104.1	1714.4	106.1	5	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved
M6	1712.0	14.4	1698.0	16.4	6	Location of accidents, Vehicle damaged, Hospitalized, Total loss, Objects involved, Marital Status

The results indicate that all the variables are significant to each other. The ranking is based on the value that has the highest drop in AIC. For model 1, the ranking is 1 since the model resulted in low AIC up to 371.2. Model 6 is the best model as the AIC is the smallest. As the variable marital status, vehicle damaged, total loss, objects, hospitalized and location of accident are significant, the logistic regression equation in this study is expressed as follows:

$$\log\left(\frac{p}{1-p}\right) = -2.41 - 0.64\,\text{SINGLE} + 0.88\,\text{VEHICLE DAMAGED} + 0.0008\,\text{TOTAL LOSS} + 1.64\,\text{OBJECTS} + 1.55\,\text{HOSPITALIZED} + 0.34\,\text{ACCIDENT LOCATION}$$

Table 6: Estimated Coefficients and P-Values for the Fitted Logistic Model

Variable	Coefficient	P-Value
Single	-0.640 (0.140)	0.000
Vehicle Damaged	0.880 (0.160)	0.000
Total Loss	0.0008 (0.000)	0.000
Objects	1.640 (0.165)	0.000
Hospitalized	1.550 (0.171)	0.000
Accident Location	0.340 (0.140)	0.016

Table 6 demonstrates that for an increase in every one unit in SINGLE, the odds of reporting the accidents decreases by a factor of -0.640. The negative coefficient indicates that being a single person involved in accidents tends to lower the likelihood of reporting the accidents. For every one-unit increase in VEHICLE DAMAGED, the odds of reporting the accidents increases by a factor of 0.880. For TOTAL LOSS, the odds of reporting the accidents increases by a factor of 0.008. For every one-unit increase in OBJECTS, the odds of reporting the accidents increases by a factor of 1.640. The result indicates that the presence of objects tends to increase the probability of reporting the accidents. For every one-unit increase in HOSPITALIZED, the odds of reporting the accidents increase by a factor of 1.550. A positive coefficient implies that the hospitalized individual will report the accident. Lastly, the odds of reporting the accidents for ACCIDENT LOCATION increases by a factor of 0.340, indicating that the accident will be reported.

4. Conclusion

The current study demonstrated that 46.5% of road accidents remained unreported during the period of this study. Moreover, the

respondents who did not report their road accidents were found to be generally of aged between 60 and 69, of Malay race, living in rural area, single and motorcyclists. Evidence has shown that the highest rate of motorcycle fatalities occurs in the main roads (Radin Umar., 1994). In contrast, majority of the accidents occur in the countryside were unreported.

The severity of a crash influences the reporting. On the other hand, underreporting commonly inclined by the human tendency to prevent from legal procedures and time as well as money concerns related to reporting of the crash. Moreover, underreporting was most pronounced amongst non-fatal, non-severe injuries and middle-aged at risk road users. The current research emphasizes the importance of enhancing the awareness with regard to reporting. The study also highlights the significance of establishing a conducive environment to increase reporting for developing a precise dataset aiming at effective and competent policy development.

Fundamentally, it is important to identify the occurrence of traffic accidents for the effective implementation of safety measures to avoid such circumstances. Further research is needed to evaluate the lifestyle factors that are related to increased risks of traffic accidents and underreporting. Importantly, society should play crucial role to report the accidents regardless of involving severity and fatalities.

Acknowledgements

We would like to express gratitude to Universiti Kebangsaan Malaysia and Ministry of Higher Education (MOHE) for sponsoring this study through grant GUP-2018-110.

References

[1] Abegaz, T., Berhane, Y., Worku, A., Assrat, A., Assefa, A., 2014. Road Traffic Deaths and Injuries Are Underreported in Ethiopia: A Capture-Recapture Method. PLoS One 9 7. https://doi.org/10.1371/journal.pone.0103001.

[2] Amoros, E., Martin, J.L., Laumon, B., 2006. Underreporting of road crash casualties in France. Accid. Anal. Prev. 38, 627–635. https://doi.org/10.1016/j.aap.2005.11.006.

[3] Aptel, I., Salmi, L.R., Masson, F., Bourde, A., Henrion, G., Erny, P., 1999. Road accident statistics: discrepancies between police and hospital data in a French island. Accid. Anal. Prev. 31, 101–108. https://doi.org/10.1016/S0001-4575(98)00051-7.

[4] Babulal, V., 2017. Number of fatal road accidents up in 2016, more than 7,000 lives lost [https://www.nst.com.my/news/2017/01/205090/number-fatal-road-accidents-2016-more-7000-lives-lost]. New Straits Times.

[5] Baker, S.P., O'Neill, B., Haddon, W., Long, W.B., 1974. The Injury Severity Score: a method for describing patients with multiple injuries and evaluating emergency care. J. Trauma Inj. Infect. Crit. Care 14, 187–196. https://doi.org/10.1097/00005373-197403000-00001.

[6] Barancik, J.I., Fife, D., 1985. Discrepancies in vehicular crash injury reporting: Northeastern Ohio Trauma Study. Accid. Anal. Prev. 17, 147–154. https://doi.org/10.1016/0001-4575(85)90017-X.

[7] Chan, D., 2009. Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences. CRC Press, New York.

[8] Cryer, P.C., Westrup, S., Cook, A.C., Ashwell, V., Bridger, P., Clarke, C., 2001. Investigation of Bias After Data Linkage of Hospital Admission Data to Police Road Traffic Crash Reports. Inj. Prev. 7, 234–241. https://doi.org/10.1136/ip.7.3.234.

[9] DeCamposa, C.I., Vasconcellos Romão, M.N.P., Pinto Ferraz, A.C., RAIA Junior, A.A., Simões, A., 2014. Hospitalizations by the Brazilian Health System Due to Traffic Accidents. Procedia - Soc. Behav. Sci. 162, 282–289. https://doi.org/10.1016/j.sbspro.2014.12.209.

[10] De Geusa, B., Vandenbulckeb, G., Panisc, L.I., Thomas, I., Degraeuwec, B., Cumpsa, E., Aertsensc, J., Torfsc, R., Meeusena, R., 2012. A Prospective Cohort Study on Minor Accidents involving Commuter Cyclists in Belgium. Accid. Anal. Prev. 45, 683–693. https://doi.org/10.1016/j.aap.2011.09.045.

[11] Elvik, R., Mysen, A.B., 1999. Incomplete Accident Reporting: meta-analysis of studies made in 13 countries 133–140. https://doi.org/10.3141/1665-18.

[12] Hauer, E., Hakkert, A.S., 1988. Extent and Some Implications of Incomplete Accident Reporting. Transp. Res. Rec. 1185, 1–10.

[13] Holdridge, J.M., N.Shankar, V., F.Ulfarsson, G., 2005. The Crash Severity Impacts of Fixed Roadside Objects. J. Safety Res. 36 2, 139–147. https://doi.org/10.1016/j.jsr.2004.12.005.

[14] Ibrahim, K., Silcock, D.T., 1992. The Accuracy of Accident Data. Traffic Eng. Control 33 9, 492–497.

[15] Juhra, C., Wieskotter, B., Chu, K., Trost, L., Weiss, U., Messerschmidt, M., Malczyk, A., Heckwolf, M., Raschke, M., 2012. Bicycle accidents – Do we only see the tip of the iceberg? A prospective multi-centre study in a large German city combining medical and police data. Injury 43, 2026–2034. https://doi.org/10.1016/j.injury.2011.10.016.

[16] Kamaluddin, N.A., Abd Rahman, M.F., Varhelyi, A., 2018. Matching of police and hospital road crash casualty records - a data-linkage study in Malaysia. Int. J. Inj. Contr. Saf. Promot. 1–8. https://doi.org/10.1080/17457300.2018.1476385.

[17] Kim, K., Kim, S., Yamashita, E., 2000. Alcohol-Impaired Motorcycle Crashes in Hawaii, 1986 to 1995. Transp. Res. Rec. 1734, 77–85. https://doi.org/10.3141/1734-12.

[18] Kim, K., Nitz, L., Richardson, J., Li, L., 1995. Personal and behavioral predictors of automobile crash and injury severity. Accid. Anal. Prev. 27 4, 469–481. https://doi.org/10.1016/0001-4575(95)00001-G.

[19] Kowtanapanich, W., Tanaboriboon, Y., Charnkol, T., 2007. A prototype of the GIS-based traffic accident database system : Thailand case study, in: Proceedings of the Eastern Asia Society for Transportation Studies.

[20] Lei, L., Kim, K., Nitz, L., 1999. Predictors of Safety Belt Use Among Crash-Involved Drivers and Front Seat passengers: Adjusting for Overreporting. Accid. Anal. Prev. 31, 631–638. https://doi.org/10.1016/S0001-4575(99)00022-6.

[21] Manan, M.M.A., Varhelyi, A., 2012. Motorcycle fatalities in Malaysia. IATSS Res. 36, 30–39. https://doi.org/10.1016/j.iatssr.2012.02.005.

[22] McClintock, C.., Brannon, D., Maynanrd, S., 1974. Applying the logic of sample surveys to qualitative case studies: the case cluster method. Qual. Methodol. 24, 612–629. https://doi.org/10.2307/2392367.

[23] McGinnis, R.G., 1999. Estimating the Influences of Driver, Highway, and Environmental Factors on Run-off-Road Crashes Using Logistic Regression. Transp. Res. Board.

[24] Mercier, C.R., Ii, M.C.S., Rimkus, J.B., Mercier, J.M., 1997. Age and gender s Predictors of Injury Severity in Head-on Highway Vehicular Crashes. Transp. Res. Rec. 1581, 37–46. https://doi.org/10.3141/1581-05.

[25] Mustafa, M.N., 2005. Overview of Current Road Safety Situation in Malaysia 5–9.

[26] Mustaffa, A.A., Hokao, K., 2012. Database development of road traffic accident case study Johor Bahru, Malaysia. J. Soc. Transp. Traffic Stud. 3 1 , 1–8.

[27] Nakahara, S., Wakai, S., 2001. Underreporting of Traffic Injuries Involving Children in Japan. Inj. Prev. 7, 242–244. https://doi.org/10.1136/ip.7.3.242.

[28] Note, F.C., Lanka, S., 2014. Integrated Conference of Better Air Quality (BAQ).

[29] Organization for Economic Cooperation and Development (OECD), 1984.

[30] Patil, S., Geedipally, S.R., Lord, D., 2012. Analysis of crash severities using nested logit model – Accounting for the underreporting of crashes. Accid. Anal. Prev. 45, 646–653. https://doi.org/10.1016/j.aap.2011.09.034.

[31] Peden, M., Scurfield, R., Sleet, D., 2004. World Report on Road Traffic Injury Prevention.

[32] Periyasamy, N., Lynch, C.A., Dharmaratne, S.D., Nugegoda, D., Ostbye, T., 2013. Underreporting of road traffic injuries in the district of Kandy, Sri Lanka. BMJ Open 3. https://doi.org/10.1136/bmjopen-2013-003640.

[33] Rosman, D.L., 2001. The Western Australian Road Injury Database: Ten years of linked police, hospital and death records of road crashes and injuries. Accid. Anal. Prev. 33, 81–88. https://doi.org/10.1016/S0001-4575(00)00018-X.

[34] Rosman, D.L., Knuiman, M.W., 1994. A comparison of hospital and police road injury data. Accid. Anal. Prev. 26, 215–222. https://doi.org/10.1016/0001-4575(94)90091-4.

[35] Ross, H.E., Sicking, D.L., Zimmer, R.A., Michie, J.D., 1993. Recommended Procedures for the Safety Performance Evaluation of Highway Features. Transp. Res. Board 350.
[36] Samuel, J.C., Sankhulani, E., Qureshi, J.S., Baloyi, P., Thupi, C., Lee, C.N., Miller, W.C., Cairns, B.., Charles, A.G., 2012. Underreporting of Road Traffic Mortality in Developing Countries: Application of a Capture-recapture Statistical Model to Refine Ortality Estimates. PLoS One 7. https://doi.org/10.1371/journal.pone.0031091.
[37] Santos, S.R. dos, Davis Jr, C.A., Smarzaro, R., 2017. Analyzing Traffic Accidents based on the Integration of Official and Crowdsourced Data. J. Inf. Data Manag. 8 1, 67–82.
[38] Schacter, D.L., 1999. The Seven Sins of Memory. Insights from Psychology and Cognitive Neuroscience. Psychology 54, 182–203. https://doi.org/10.1037//0003-066X.54.3.182.
[39] Scortiono, S., Vassar, M., Radetsky, M., Knudson, M.M., 2005. San Francisco Pedestrian Injury Surveillance: Mapping, underreporting, and injury severity in police and hospital records. Accid. Anal. Prev. 37, 1102–1113. https://doi.org/10.1016/j.aap.2005.06.010.
[40] Sethi, D., Habibula, S., McGee, K., 2004. World Health Organization.
[41] Shinara, D., Valero-Morab, P., Strijp-Houtenbosc, M. van, Haworthd, N., Schrammd, A., Bruynee, G. de, Cavallof, V., Chliaoutakisg, J., Diash, J., Ferraroi, O.E., Fyhrij, A., Sajatovick, A.H., Kuklanel, K., Ledesmam, R., Mascarelln, O., Morandii, A., Musero, M., Ottep, D., Papadakakiq, M., Sanmartínb, J., Dulfr, D., Saplioglus, M., Tzamaloukag, G., 2016. Underreporting bicycle accidents to police in the COST TU1101 international survey: cross-country comparisons and associated factors. Accid. Anal. Prev. 110, 177–186. https://doi.org/10.1016/j.aap.2017.09.018.
[42] Smith, D.B., Meshkati, N., Robertson, M.., 1993. The older driver and passenger. Automot. Ergon. 453–471.
[43] Stratified Sampling [WWW Document], 2018. . Pennsylvania State Univ.
[44] Steyerberg, E.W., 2001. Internal Validation of predictive models: Efficiency of some procedures for logistic regression analysis. J. Clin. Epidemiol. 54, 774–781. https://doi.org/10.1016/S0895-4356(01)00341-9.
[45] Stuuts, J.C., Williamson, J.E., Whitley, T., Sheldon, F.C., 1990. Bicycle Accidents and Injuries: A Pilot Study Comparing Hospital and Police Reported Data. Accid. Anal. Prev. 22, 67–78. https://doi.org/10.1016/0001-4575(90)90008-9.
[46] Tsuia, K.L., Soa, F.L., Szeb, N.N., Wongb, S.C., Leung, T.F., 2009. Misclassification of Injury Severity among Road Casualties in Police Reports. Accid. Anal. Prev. 41, 84–89. https://doi.org/10.1016/j.aap.2008.09.005.
[47] Umar, R.S.R., 1994. Road Accidents in Malaysia. IATSS Res. 18 2, 38–41.
[48] Ward, H., Lyons, R, Thoreau, R., Underreporting of road casualties, Department for Transport (2006).
[49] World Health Organization (WHO), 2018. Road Traffic Injuries [WWW Document]. URL http://www.who.int/mediacentre/factsheets/fs358/en/.