ASM Sc. J., 12, Special Issue 1, 2019 for IQRAC2018, 239-245

Estimating The Proportion of Non-Fatality Unreported Traffic Accidents in Malaysia

Nik Sarah Nik Zamri*[1] and Zamira Hasanah Zamzuri[2]

[1,2]*School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor*

**Corresponding author: zamira@ukm.edu.my*

Traffic accidents statistical analysis is important in understanding and estimating the occurrence of the event. However, traffic accident data often exhibit the occurrence of extra zeros, due to the underreporting scenario; in which the accidents are happened but not reported, mainly for non-fatality accidents. Hence, the aims of this paper are to gain insight into the accident reporting system and estimating the proportion of non-fatality unreported accidents in Malaysia. Two sets of data were considered in this study: the survey data and the police data. The data were divided according to the level of severity, which were serious and slight injury for survey data and police data in the year 2013-2015. The survey data collects information on both unreported and reported accidents based on the Malaysian drivers' experiences; whereas the police data is the reported accidents occurred in the year 2013 – 2015. Distributions considered in this study are normal, gamma, beta, and logistic were fitted to the proportion and the log of the proportion of unreported accidents in the survey data. We found that the proportion of unreported serious injury is 0.54 and the proportion of unreported slight injury is 0.58. We also estimated the proportion of unreported accidents per driver, using beta distribution, in which the estimate for the proportion of unreported serious injury is 0.8418 and the proportion of unreported slight injury is 0.8340.
Keywords: Proportion, non-fatality, unreported accidents, distribution, severity

I. Introduction

Police reported traffic accident data are the main source of crash in most authority (Watson et al., 2015). In contrast, the dependency on police data to calculate and analyze road accident injuries may be a problem, as it is clear that not all road accident accidents are reported to the police. This unreported accident can not only impact on the overall size of the road accident injuries, but may introduce a bias associated with a particular road user group. This problem is known as unreported road accident injury, which is usually identified when comparing hospital and traffic data on road accident accidents and may be largely due to differences in the definition of severity of injuries (Abay, 2015). In particular, the comparison reveals that only a limited number of non-fatal injuries in hospitals recorded by the police, although much less known about non-severe (e.g. non-hospitalized) injuries (Watson et al., 2015). Some studies provide evidence that significant road safety injuries are not reported by the police, while the underreporting level may vary between levels of injuries to injuries or different groups of road users (Alsop and Langley, 2001); (Amoros et al., 2006); (Boufous et al., 2008); (Langley et al., 2003). In the area of road accidents and injuries, various data-linking projects have been regulated by Abay (2015); Alsop and Langley (2001); Amoros et al. (2006); Boufous et al. (2008); Cercarelli et al. (1996); Elvik and Mysen (1999); Hauer and Hakkert (1988) and Langley et al. (2003). Abay (2015) combines emergency room data with accident data reported by police in the Denmark region. It has been found that there is a heavy tendency in reporting to the police including the tendency of severity (i.e., higher severity injuries were presumed to be reported). In New Zealand, Alsop and Langley (2001) use the police relations and hospital records. They discovered that less than two-thirds of all road accident victims were recorded in police data. They also find that this varies based on the number of vehicles involved, geographic location, age and injury severity. Amoros et al. (2006) conducted the same study to see underreporting of road accidents in France. They use probabilistic methods to link police accident data with the registration of road trafficking in Rhone County. The result shows that the police reporting rate is about 38%. However, these rates vary accordingly to injury severity, the type of the road user, and the location of the accidents (i.e., urban vs. rural). Another study conducted by Aptel et al. (1998) in France found that after connecting police and hospital data, only 37% of road accident injuries were reported to the police. Similar to other studies, they find that the reporting rate varies according to the place of accident, the type of vehicle involved, and the injury severity. They also pointed out that police reports were likely to account for the excessive level of injury sustained. Langley et al. (2003) conducted links between hospital records and police records to specifically examine the possibility of underreporting of cyclist accidents in New Zealand. The result shows that only 22% of cyclists crashed on public roads could be linked to police records. Of the accidents involving motor vehicles, 54% were registered by the police. They also found that age, ethnicity, and injury severity predicted whether hospital cycle crashed were more likely to be reported in police data. In Australia, Cercarelli et al. (1996) linking police reports, hospital admissions and accident and emergency (A & E) department data from a hospital in Western Australia. The researchers found that about 50% of A & E's attendance was recorded by the police, and about 50% of the cases recorded by the police as hospitalized were actually admitted. Researchers stated that while the differences between data sets does illustrated the reporting case below, it indicates that the differences in the coding system may also result in unlinked cases. Another study conducted in Australia at New South Wales by Boufous et al. (2008) linked hospital data (Patient Statistics Order [ISC]) with Traffic Accident Data System (TADS). Researchers match 56.2% of hospitalization due to road accidents with records in TADS. Researchers also found that the relevance rate varies by age (i.e., lower linkage rate for younger age groups), road user type (e.g., lower linkage rate for cyclists), severity (i.e., higher linkage rates with increased severity) and geographical location. Furthermore, studies under the United Kingdom's national report show a significant difference in underreporting between different levels of injury which is about 20% of the injuries categorized as serious by the Police have been treated and discharged by the hospital (i.e. slight injuries), while those treated by the hospital were serious but appeared in police records as slight were account for about 8% in the country. Study by Elvik and Mysen (1999) deduced that reporting of injuries in official accident statistics is insufficient at all measures of injury severity in a meta-analysis of 49 studies in 13 countries. Specifically, the level of reporting the mean of accidents in the countries included is found to be 95% for fatal injuries, 70% for serious injuries (described as those admitted to hospital), 25% for slight injuries (described as those treated as outpatients), and 10% for very slight injuries (described as those treated outside hospitals). In addition, levels of injury reporting vary among countries, ranging from 21 to 88% for hospital-treated injuries. In Malaysia, Kamaluddin et al. (2018) studied on data linkage of matching police and hospital record in state of Melaka which yield 4.1% matching rate and a 4.7% police reporting rate. the matching rate between police and hospital records is proportional to the level of injury severity. Besides that, Mustaffa and Hokao (2012) developed a database system for the road traffic accidents that can be used to reduce the fatalities of road accidents in the state of Johor Bahru. Since there is still a limited number of references on this issue for Malaysia data, this paper intends to estimate the proportion of unreported accidents in Malaysia. By finding the fitted distributions to the proportion of unreported accidents, perhaps a richer information can be obtained in order to shed some light to the actual number of accidents occured. Futhermore, the distribution fitted to the proportion of unreported accidents data will help in developing a traffic accident model that include underreporting information perhaps to improve the estimation of the actual number of accidents.

II. Methodology

In this paper, we focus on the proportion of the unreported accidents. We will consider two sources of data which are the survey data and the police's data that were obtained from MIROS (Malaysian Institute of Road Safety Research) for the year 2013-2015. MIROS was established in 2007 as an agency under the Ministry of Transport Malaysia to serve as a central repository of knowledge and information on road safety. For the survey data, a survey has been conducted in order to collect the information of unreported traffic accidents in Malaysia. The questionnaires were disseminated to 2000 Malaysia drivers, selected based on the stratified sampling to represent the Malaysian population. In the questionnaire, there are two sections. Section A is 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 survey data were divided according to the level of severity which were serious injury ($n = 73$) and slight injury ($n = 167$) based on police severity score (serious or slight injury) (Yannis et al., 2014).Police classify traffic injuries into three groups based on their severity as follows :(1)fatal; (2)serious injury and (3)slight injury. Fatal has been removed due to the unreported accident is less than 5. As explained previously, this paper intends to estimate the proportion of unreported accidents in Malaysia based on these two sources of data.

There are several studies on matching police and hospital data and this situation is alarming that exist in Malaysian scenario (Kamaluddin et al., 2018); (Mustaffa and Hokao, 2012). Since we are unable to obtain the insurance and hospital data, we aim to measure the proportion of underreporting using survey data. For a fair measurement, we focus only on the accidents occurred in the year of 2013 to 2015. This will filter out the influence of temporal factor and number of years of the drivers' driving experience.

We will look on the proportion of unreported accidents for overall data based on two approaches which are direct method and statistical distribution.
Direct computation (overall proportion of unreported accidents) :

$$\hat{p} = \frac{u}{u+r} \tag{1}$$

$\hat{p}$ = proportion of unreported accidents
u = number of unreported accidents
r= number of reported accidents

Next, we estimate the proportion of unreported accidents per driver. We compute the proportion of unreported accidents as given in the previous formula for each driver, then we fit the potential distributions to this proportion data. Table 1 listed the potential statistical distributions considered in this study; whereas Table 2 listed the parameter, mean and variance for these distributions. The choice of the distributions are based on Stroup (2015), in which for a proportion data with range 0 to 1, beta distribution can be considered. To offer another option of the beta distribution, we also fit the gamma distribution. Note that for gamma distribution, zero values is not included in gamma probability density function. In order to overcome this problem, we changed the zero value to a very small value which is 1 x 10^{-7}.

We also consider the log of proportion as another measure for the unreported accidents. As explained by Feng et al. (2014), the log transformation is used to deal with the skewed data. The distributions fitted to the log (p) is listed in Table 1.

In this distribution fitting process, the maximum likelihood method is employed to estimates the parameter. Details on this method can be found in Cousineau et al. (2004). Then, the best distributions for each measures are identified by comparing the selection criterias which are Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) as used by Alam et al. (2018) and Vrieze (2012).

Table 1: Potential Distributions for Proportion

Variables	Potential Distributions
Proportion (p)	Beta, Gamma
log (p)	Normal, Logistic,

III. Results and Discussion

This section presents the overall results and related discussions. Section 3.1 provides results on estimating the proportion using direct method and section 3.2 discuss about the results on estimation of proportion using statistical distribution.

A. Direct Method

Table 3 shows the overall proportion of survey data from the year 2013-2015.

Table 3 shows that survey data has been matched with PDRM data based on percentage level of severity. This procedure is being made in order to ensure the proportion of unreported accidents is a reliable estimate.

Table 4 shows by using the formula, we can estimate the proprotion of unreported accidents based on survey data. Then we use this estimate to further predict the unreported accidents in police's data as given in the column below.

We can see based on this crude estimate, it is estimated about 31000 accidents are not reported for the three years period, resulting on around 10000 unreported accidents per year.

B. Statistical Distribution

In order to estimate the proportion of unreported accidents per driver, we find the best distribution by comparing the values of two criterions, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). AIC compares the quality of set of statistical model and used to decide the best approximating models while BIC is a criterion for model selection among a finite set of models. In order to find the best fitted distribution, we compare the AIC and BIC. The distribution is best fitted if their AIC and BIC are smallest in value.

Based on Table 5, beta distribution is the best fitted for proportion of serious injury unreported accidents as the AIC and BIC are the smallest respectively.

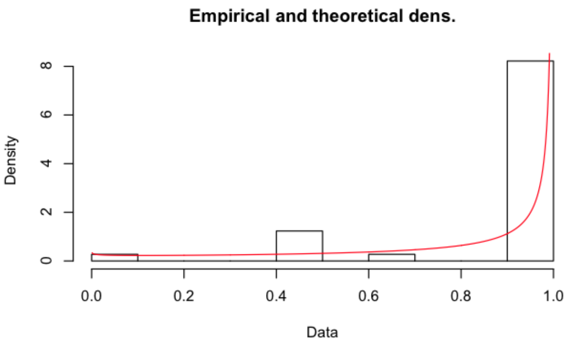

Figure 1: Graph of proportion of serious injury

Figure 1 shows the graph of proportion of serious injury. The graph of proportion lies be-

Table 2: Parameter, mean and variance of each distributions

Distribution	Parameter	Mean	Variance
Normal	μ, σ^2	μ	σ^2
Logistic	μ,s	μ	$\frac{\pi^3 s^2}{3}$
Beta	α,β	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Gamma	k,θ	$k\theta$	$k\theta^2$

Table 3: Proportion of reported accidents for the Survey Data and PDRM data

	PDRM	**Percentage of PDRM**	Survey	**Percentage of Survey**
Serious Injury	8365	**34.3%**	63	**34.2%**
Slight Injury	15999	**65.7%**	121	**65.8%**
Total	24364	**100%**	184	**100%**

Table 4: Estimation of unreported accidents for police data

	Survey Data			PDRM Data		
	Reported Accidents	Unreported Accidents	$\hat{p}$ **unreported**	Reported Accidents	$\widehat{Unreported}$ **Accidents**	$\widehat{Total}$ **Accidents**
Serious Injury	63(34.2%)	73(29.8%)	**0.54**	8365(34.3%)	**9820(30.8%)**	**18185**
Slight Injury	121(65.8%)	167(68.2%)	**0.58**	15999(65.7%)	**22094(69.2%)**	**38093**

Table 5: Serious injury for unreported accidents

Distribution	p	AIC	BIC
Beta	**0.84**	**-439.80**	**-435.22**
Gamma	0.90	100.03	104.61

tween 0.80 and 1.00 as shown in the Table 5 which the value of proportion is 0.84.

Table 6: Serious injury for log p

Distribution	log p	p	AIC	BIC
Normal	-0.29	0.75	229.73	234.31
Logistic	-0.09	0.91	134.14	138.72

We also fitted the potential distribution to another measure, the log of the proportion of unreported accidents. Based on Table 6, logistic distribution is the best fitted for log p since the AIC and BIC are the smallest.

Based on Table 7, beta distribution is the best fitted for proportion of slight injury unreported accidents as the AIC and BIC are the smallest respectively

Figure 2 shows the graph of proportion of slight injury. The graph of proportion lies between 0.80 and 1.00 as shown in the Table 7 which the value of proportion is 0.83.

Table 7: Serious injury for unreported accidents

Distribution	p	AIC	BIC
Beta	**0.83**	**-1027**	**-1021**
Gamma	0.90	236.59	242.83

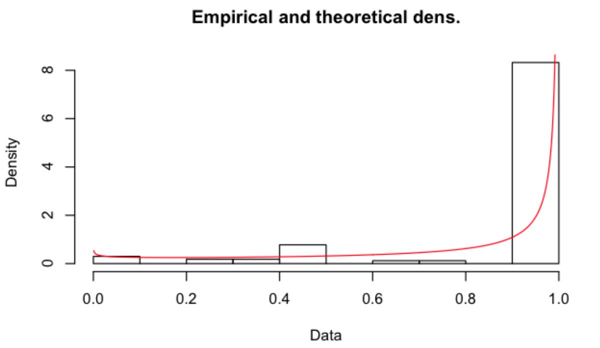

Figure 2: Graph of proportion of slight injury

Based on Table 8, logistic distribution is the best fitted for log p. By using the direct

Table 8: Slight injury for log p

Distribution	log p	p	AIC	BIC
Normal	-0.31	0.73	537.31	543.54
Logistic	-0.10	0.91	338.57	344.80

method, we can estimate the overall proportion of the unreported accidents. The proportion of unreported accidents per driver is estimated by the statistical distribution.

We sample the reported survey data so that the percentage of accidents based on level of severity match with the PDRM data. The reported data also carries the info on unreported traffic accidents. Although the sample is not a lot, it all lead to same conclusion showing there exist unreported accidents scenario and cause serious in predicting the accidents. The duration of this study is focusing on year of accident occured from 2013-2015. Our study also cover limited area therefore, there is a limitation in this study as we also lack data from hospital and insurance records. On the other hand, the results are significant and alarming and should not be taken lightly. In addition, enforcement of laws and citizens' awareness of their responsibility to report crash involvement for all types of crashes should be improved.

IV. Conclusion

Based on the study findings, the underreporting of traffic accidents scenario did occur in Malaysia. The proportion for serious and slight injury per driver are 0.8418 and 0.8340 respectively, estimated by the Beta distribution. It shows that for a slight or serious accident occurred to a Malaysian driver, approximately 0.8 is not reported. It is also estimated around 20000 serious and 40000 slight accidents are not reported for the three years considered (2013-2015), which is an alarming information.

To further understand the traffic safety situation, complementary data (e.g. self-reported accidents) are necessary to enhance the current crash databases, which could help the policymakers and practitioners in improving the planning of preventive road safety measures and road traffic accidents. The parameters used in this study can help to identify the distributions for unreported accidents in Malaysia hence develop the statistical models to estimate the unreported accidents in Malaysia.

Acknowledgements

We would like to express gratitude to Universiti Kebangsaan Malaysia and Ministry of Higher Education (MOHE) for sponsoring this study through grant FRGS/1/2015/ST06/UKM/02/1. The reviewers are also thanked for their constructive comments on improving the paper.

References

[1] K. A. Abay. Investigating the nature and impact of reporting bias in road crash data. *Transportation Research Part A: Policy and Practice*, 71:31–45, 2015.

[2] M. A. Alam, C. Farnharm, and K. Emura. Best-fit probability models for maximum monthly rainfall in bangladesh using gaussian mixture distributions. *Geosciences*, 8(138):1–15, 2018.

[3] J. Alsop and J. Langley. Under-reporting of motor vehicle traffic crash victims in new zealand. *Accident Analysis and Prevention*, 33(3):353–359, 2001.

[4] E. Amoros, J. L. Martin, and B. Laumon. Under-reporting of road crash casualties in france. *Accident Analysis and Prevention*, 38(4):627–635, 2006.

[5] I. Aptel, L. R. Salmi, F. Masson, A. Bourdé, G. Henrion, and P. Erny. Road accident statistics: Discrepancies between police and hospital data in a french island. *Accident Analysis and Prevention*, 31(1-2):101–108, 1998.

[6] S. Boufous, C. Finch, A. Hayen, and A. Williamson. Data linkage of hospital

and police crash datasets in new south wales, australia. *Injury Risk Management Research centre*, pages 1–64, 2008.

[7] L. R. Cercarelli, D. L. Rosman, and G. A Ryan. Comparison of accident and emergency with police road injury data. *Journal of Trauma - Injury, Infection and Critical Care*, 40(5):805–809, 1996.

[8] D. Cousineau, S. Brown, and A. Heathcote. Fitting distributions using maximum likelihood: Methods and packages. *Behaviour Research Methods, Instruments and Computers*, 36(4):742–756, 2004.

[9] R. Elvik and A. Mysen. Incomplete accident reporting: Meta-analysis of studies made in 13 countries. *Transportation Research Record: Journal of the Transportation Research Board*, 1665:133–140, 1999.

[10] C. Feng, H. Wang, N. Lu, T. Chen, H. He, Y. Lu, and X. M. Tu. Log-transformation and its implications for data analysis. *Shanghai Archives of Psychiatry*, 26(2):105–109, 2014.

[11] E. Hauer and A. S. Hakkert. Extent and some implications of incomplete accident reporting. *Transportation Research Record*, 1185:1–10, 1988.

[12] N.A. Kamaluddin, M.F. Abdul Rahman, and A Varhelyi. Matching of police and hospital road crash casualty records - a data-linkage study in malaysia. *International Journal of Injury Control and Safety Promotion*, pages 1–8, 2018.

[13] J. Langley, N. Dow, S. Stephenson, and K. Kypri. Missing cyclists. *Injury Prevention*, 9(4):376–379, 2003.

[14] A.A. Mustaffa and K. Hokao. Database development of road traffic accident case study johor bahru, malaysia. *Journal of Society for Transportation and Traffic Studies*, 3:1–8, 2012.

[15] W.W Stroup. Rethinking the analysis of non-normal data in plant and soil science. *Agronomy Journal*, 107(2):811–827, 2015.

[16] S. I. Vrieze. Model selection and psychological theory: A discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological Methods*, 17 (2):228–243, 2012.

[17] A. Watson, B. Watson, and K. Vallmuur. Estimating under-reporting of road crash injuries to police using multiple linked data collections. *Accident Analysis and Prevention*, 83:18–25, 2015.

[18] G. Yannis, E. Papadimitriou, A. Chaziris, and J Broughton. Modeling road accident injury under-reporting in europe. *European Transport Research Review*, 6(4):425–438, 2014.