EXAMINING SAFETY OF ELECTRONIC SIGNS: USING ORDINAL LOGISTIC REGRESSION ON SPEEDING

Z. EBRAHIM & H. NIKRAZ
Department of Civil Engineering, Curtin University, Australia.

ABSTRACT

Speeding continued to be of alarming concern for many countries. This paper aims to focus on highlighting speeders characteristics on 40 km/h on a busy urban road with high pedestrian movement. This case study utilised the ordinal logistic regression model to test four predictors. Three of which were age and gender of the driver and the time of day drivers were detected speeding. Whereas the fourth explanatory variable is the 'period' of the installation of the signs, which tests the usefulness of electronic signs. The study found that the driver's age contributes slightly to risky speeding behaviours, and older drivers speed less. Time of the day was found to be significant in the model, with a higher number of TINs being recorded in the afternoon than in the morning. Although gender was not found to be a significant predictor, it was shown to produce results similar to speeding data recorded in Perth roads with males speeding slightly more than females. This difference was more pronounced when higher speeding levels were compared. The period variable in the model relating to the installation of the signs was significant, with drivers slowing down after the installation of the flashing 40 km/h electronic signs compared with before the installations This may prove the usefulness of such signs in reducing speeding behaviour. Hence, reducing harm by reducing frequency and severity of crashes.
Keywords: Multinomial logistic regression, ordinal logistic regression, speeding.

1 INTRODUCTION

Speeding is a leading cause of fatalities on WA roads. According to the office of road safety in WA, 40–45% of drivers regularly drive above the posted speed limit. Search found that the speeding factor is of concern worldwide, a report by Moran [1] said that there is evidence to suggest that both cyclists and pedestrians are far less likely to be killed for every 10 km/h reduction below 60 km/h, In addition, the risk of an injury crash doubles with every 5 km/h increase in speed above the speed limit in a 60 km/h zone [2]. Reduction in the vehicle speeds would minimise significantly vehicle–pedestrian conflicts in areas of pedestrian presence and activity [3].

Importantly, literature found that there is a link between travelling speed and crashes particularly for KSI (Killed or Seriously Injuries) [4, 5]. Most recent data show that there is a far bigger problem in Perth, involving thousands of drivers who took frequent low-level speeding [6]. All the above indicate a relation between speed limit and travelling speed from one side and safety of road users on the other.

The road under study had a posted speed of 60 km/h limit and road authorities decided a trial to install 40 km/h electronic speed signs along that road. Main roads authority of WA selected this road under study as it has higher pedestrian crashes than any other road in the metropolitan area.

The signs on this road displayed a speed limit of 40 km/h from 7.30 am till 10.00 pm (except on Friday and Saturday, when the time was extended till 1:00 am instead of 10:00 pm). The same signs displayed a speed limit of 60 km/h outside the abovementioned times.

This case study is using speeding data in a road known as road B to determine the improvement in terms of less speeding after the installation of the electronic signs. Earlier in [7], the examination involved the comparison of road B to another road known as road A in terms of
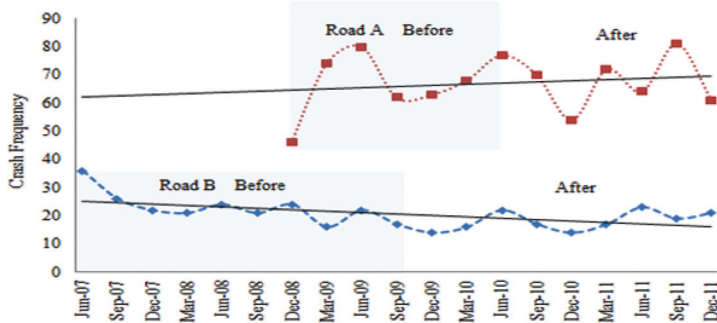
Figure 1: Comparison of road crashes between roads A and B.

road crash data. It was found that road B has witnessed a downward trend in crashes after the installation of the electronic signs, compared with crashes in road A after the installation of the 40 km/h standard signs with standard road markings (see Fig. 1).

Despite this downwards trend in road crashes, concerns were raised due to the high number of speeding fines, the most severe increase came in the number of speeding motorists in 40 km/h zone. There were 96 drivers caught daily in the 40 km/h zone compared with 83 in the previous year [8]. If drivers' speeding behaviours continued to a similar level in that zone, it will have an increase by 4745 speeding fines annually in that stretch of roads if used the same enforcement intensity.

This case study is utilising speeding driver's data. It contains four explanatory variables, age, gender, time being detected speeding and the speeding frequency periods before and after the installation of electronic signs. Literature revealed that gender was a prominent explanatory variable in some studies [9–11] and in older female drivers was attributed to causes such as poor attention, cognition and other age-related reasons.

In terms of time of day effect, Lenny *et al.* [12] found that there was an impairment in driving performance in the early afternoon that are similar magnitude to those occurring in the early and late evening. By contrast, a study of industrial crashes in Queensland Australia found that more injuries occur in the morning in Australia than in the afternoon on every working day of the week [13]. In Spain, Camino Lópeza *et al.* [14] concluded that most accidents in construction industry were more towards the afternoon periods attributing it to the "lunch effect"; the study added that these accidents occurred in the hours around the lunch break (from 13:00 to 17:00) and are of greater severity and involve more fatalities than those that occur at other times of day. Other dimension added by Banwell *et al.* [15] in Australia in a survey of the construction workers reported that construction industry involves the consumption of alcohol, which may explain the afternoon accidents involvements.

This current study has included morning and afternoon time to be included as explanatory variable in the model. The paper will focus purely on speeding data rather than crashes.

2 DATA

The speeding data were collected from the WA Police research unit and it contain data for Perth metropolitan city. These TINs issued for speeding over 40 km/h roads in five speeding levels. It is detected for the period July 2007–June 2011. This study classified the speeding categories into four levels (low, medium, high and extreme). These speed limits

Table 1: Speeding levels and outcome.

km above speed limit	Speeding level	Speeding outcome
<9	Low	Low*
10—19	Intermediate	Moderately
20–29	High	Highly
30 km and more	Extreme	Excessively

*Reference category.

were based on the office of road safety of WA. There were originally five levels as follows:

- Not more than 9 km/h.
- More than 9 km/h but not more than 19 km/h.
- More than 19 km/h but not more than 29 km/h.
- More than 29 km/h but not more than 40 km/h.
- More than 40 km/h.

It was decided to join the two highest speeding levels into one category, as shown in Table 1, based on the premise that any speed greater than 30 km/h over the speed limit is considered an extremely unsafe speed. Therefore, the speeding categories were modified into four speeding levels (low, medium, high and extreme).

The case study will apply OLR (ordinal logistic regression) to speeding levels as a dependent variable, and four other explanatory variables, being age, gender, time detected speeding and the frequency of speeding periods before and after the installation of electronic signs. The data for this fourth variable consisted of 2077 TINs (Traffic Infringement Notices) in total; the before data consisted of only 602 TINs (before 4 August 2009, which is considered to be the date of installation of the electronic signs). This compares with 1475 TINs after the installation, until from 30 June 2011. These data are what was available from the authorities. The next section will discuss the selection of the statistical model.

3 METHODOLOGY

The study will deal with three discrete outcomes as the forth outcome the low speeding level will be used as a baseline (reference) level. The speeding outcome will be shown as described in Table 1. The study aims to test the odds probability of the dependent variable speeding level occurring as the values of the IVs change. In other words, the objective is to find out the magnitude of speeding levels due to high speeding levels. This comes from the well-known concerns in the transportation literature about speeding above the limit, which could result in serious and more severe crashes.

Thus, the Y event is very unlikely to occur if f(Y) is close to 0 and that it will be unlikely that the driver will engage in higher speeding behaviour. On the contrary, it is very likely to occur if f(Y) is close to the value of 1.

The study adopting low level speeding 0–9 km above posted speed limits as a reference category compared with other three higher speeding levels. The study attempted initially to examine data fitting to suitable model to fit the data. The first method was OLR and the second was multivariate logistic regression (MLR). Following a comparison of the two methods, it was decided to adopt OLR, for two reasons. First, the dependent variable mentioned above

Table 2: Comparison of the OLR and MLR models.

Model fitting details	Values		P		Remarks
	OLR	MLR	OLR	MLR	
-2 LL (x^2)	172.56	236.48	0.001	0.001	The change is significant and explains the decrease in unexplained variance, and it is considered a good improvement to the model
Goodness of fit					The predicted values are not significant
Pearson	1241	1268	0.39	0.14	and not different from the observed,
Deviance	979	917	1.00	1.00	thus the fit of the model is good. The Pearson p value for the ordinal model is higher, producing a better fit
Pseudo R^2					Fairly similar values but MLR values are
Cox and Snell	0.08	0.11	–	–	slightly higher. Still OLR representing a
Nagelkerke	0.10	0.13	–	–	good size effect according to and it is a
McFadden	0.05	0.06	–		supplementary test [16–18]

(the speeding level) is ordinal in nature. Secondly, upon comparison of the fitting information for the two models (see Table 2), it was found that the Pearson's *P* value for the ordinal model might produce a better fit. It is important to mention that this comparison was undertaken, because the MLR model did give a good competing fitting information in comparison with the ordinal model when the data were tested.

The explanatory variables are age, as a continuous variable, and time of the day, gender and the period of the electronic sign installation as categorical variables. The dependent variable is the four speeding levels above the 40 km/h speed limit (<9, 10–19, 20–29 and 30 km/h and more).

The study attempted initially to examine data fitting to suitable methods to fit the data according to the above available data. Following the comparison of the two methods, it was decided to adopt OLR, for two reasons. First, the nature of the dependent variable which is here taken as the speeding levels is of an ordinal nature. Secondly, upon comparison of the two models fitting information it was found that the Pearson p value of the ordinal model might produce a better fit (see Table 2).

The study dealt with three discrete outcomes as the low speeding levels will be used as a baseline (reference) level. The speeding outcomes will be discussed based on the OLR fitting information. The study aims to test the probability of the dependent variable speeding level occurring as the values of the independent variable change. In other words, the objective is to find out the magnitude of high speeding levels. This comes from the common concerns in the transportation literature about drivers exceeding the speed limit, with the potential for more severe crashes.

Logistic regression is valued as an important analysis tool to roads safety studies. It is a source of attraction for researchers in targeting the effect to dependent variables on a binary dichotomous outcome or multi-level outcome. For example, researchers in [19] found that in terms of speed and speeding concerns, higher speed limits were contributing to higher risk levels in China. Whereas [20] in Saudi Arabia and using odds ratio concept as an estimate of the logit, found that speed as a cause of accident factors were significantly associated with

accident severity. According to a study by Cameron *et al*. [21] in Australia and using logistic regressions found that there was no interaction between the effect of TINs and speed-related publicity awareness on the frequency of casualty crashes. It further found that the increased level of speed camera ticketing was associated monotonically with road crash reduction.

4 THE MODEL

To build the model, it is essential to construct the model to get a logit score *Y*, and then further check the predicted probabilities and scatter them against the Pearson residuals that the model produces [22]. The logit function is given by eqn (1) which is derived from different predictors multiplied by their corresponding regression coefficients, where *Y* is calculated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n \tag{1}$$

where β_0 is the intercept of *Y* value and when all the predicting variables $X_1, X_2, X_3, \ldots, X_n$ are equal to zero. The variable *Y* (logit) is a measure of the sum of the input of all the independent predictor variables used in the model. The variable *Y* is defined as

$$(Y) \text{ Speeding} = \beta_0 + \beta_1 * \text{Age} + + \beta_2 * \text{Gender} + \beta_3 * \text{Time} + \beta_4 * \text{Period} + \ldots + \beta_n X_n \tag{2}$$

where speeding is the four levels as shown in Table 1, age is the recorded age of speeding driver (continuous variable), gender is the gender of speeding driver (categorical variable, female = 0 and, male = 1), timeis the time of day speeding detected (categorical variable, morning (6.00–11.59 am) = 0, afternoon (12.00–19.00 pm) = 1) and period is the frequency of speeding period before and after the installation of the electronic signs (categorical variable, before installation = 0, after installation = 1).

Then, it is appropriate to calculate the probabilities of that logit, by using eqn (3).

$$P(Y) = f(Y) = \frac{1}{1 + e^{-Y}} \tag{3}$$

where $f(Y)$ is the probability of speeding occurring represented by *Y* and *e* is the base of the natural logarithm. For several predictors, the equation would become

$$P(Y) = f(Y) = \frac{1}{1 + e^{-\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n}} \tag{4}$$

Substituting *Y* of eqn (2) into (4), with all predictors, the final probability calculated will become

$$f(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * \text{Age} + \beta_2 * \text{gender} + \beta_3 * \text{time} + \beta_4 * \text{period})}} \tag{5}$$

Speeding data will therefore be calculated through OLR using speeding levels and their corresponding probabilities. In this study, SPSS software package has been utilised to calculate variables and parameters of the model [23].

5 THE RESULTS

Two main evaluation steps are discussed. First, the crucial fitting information that determines the model's good fit which was discussed and, secondly, the parameter estimates testing the coefficients and the odds ratios of the model. Table 2 discusses the log-likelihood, goodness of fit and pseudo R^2. It can be noted that all details are encroaching, showing a good fitting model if the explanatory variables are used.

Table 3: Model parameter estimates.

Predictors	β (SE)	Wald χ²	95 CI limits		Exp(B)	P-value
			Lower	Upper		
Age	−0.01 (0.03)	12.36	−0.019	−0.005	1.00	0.001
Time = 0	−0.70 (0.10)	47.34	−0.90	−0.50	0.50	0.001
Time = 1	0.0					
Gender = 0	−0.02 (0.10)	0.04	−0.21	0.17	0.84	0.84
Gender = 1	0.0					
Period = 0	0.91 (0.11)	67.31	0.69	1.12	2.48	0.001
Period = 1	0.0					

Furthermore, in this section and as shown in Table 3, particular attention needs to be paid to the odds ratios Exp(B) for interpretation of the model values and the decrease and increase of the predictor's effect on speed outcome. The value of Y calculated from eqn (2) is equal to the value of β when all other independent variables are equal to 0. The coefficient is either a positive or a negative one, indicating an increase or a decrease in the probability, respectively. Similarly, if the coefficient is high or low, it means that the risk due to speeding strongly or weakly affects the outcome, respectively [24]. The parameter estimates are as follows:

Age: The coefficient for speeders in terms of age is found to be 0.01 and the odds ratio is >1. Thus, if the speeder's age is increased by one unit (year), the likelihood of a driver speeding decreases by 0.02 units. Older drivers are slightly less likely to speed.

Time of the day: The coefficients for time of detection, where 0 = morning and 1 = afternoon appear to be significant predictors. The coefficient is 0.70 and the odds ratio <1. Hence, drivers are more likely to speed in the afternoon than in the morning.

Gender: Despite the fact that the coefficient of a person's gender (female = 0, male = 1) does not significantly predict the speeding of drivers. It is still important to highlight and discuss this explanatory variable, since gender is frequently related to speeding in the literature. It was found that the coefficient was 0.02 and the odds ratio was <1. Thus, male drivers are slightly more prone to speeding than female drivers. This also supported by the data displayed in Fig. 2.

Period: This variable is of high importance to this study since it determines the usefulness of the flashing electronic 40 km/h sign in reducing speeding. The coefficients for periods (before = 0 and after =1) are 0.91 and the odds ratios >1. This shows that drivers were speeding more before the installation of the signs than afterwards.

6 DISCUSSIONS

Three of the four explanatory variables were found to be significant, with the exception of gender. Despite this, gender is still discussed due to its importance as a variable in speeding and safety on the road. It was found that the age of the driver contributed slightly to speeding. Older drivers were slightly less likely to drive faster than the speed limit compared with younger drivers. It is important to mention that time of the day was found to be significant in the model. Drivers were more prone to being detected speeding during the afternoon than in the morning. When road crash data for the years 2007–2011 was analysed, it was found to be

consistent with the above model findings. The analysis showed that only 33% of pedestrian crashes occurred in the morning period (6:00–11:59 am), compared with 67% in the afternoon period (12:01–5:59 pm). By comparison, 25% of all other crash types occurred in the morning compared with 75% in the afternoon. Therefore, time of day may reveal more information about why drivers speed and how this behaviour may be better monitored, detected and deterred. This may include the development of a strategy for the crucial afternoon hours, such as increasing enforcement at that time of day, at specific locations. The time period and its risk factors to protect vulnerable road users such as the pedestrians was discussed in [25] and also in [26] involving busy shopping strips such the strip under study with high pedestrian activities. In terms of gender, it was found that male drivers were slightly more prone to speeding than female drivers. As can be seen in Fig. 2, the study found that male drivers speed more than female drivers on 40 km/h roads in Perth. The figure also shows that males speed more for each speed level, except for the lowest speed above the limit, for which males and females seem to speed equally.

The variable relating to the period before and after installation of the signs is also significant. Drivers were found to have slowed down after the installation of the flashing electronic 40 km/h signs. This has recently been supported by Ebrahim and Nikraz [7] who found that the incidence of crashes decreased after the installation of the 40 km/h electronic speed signs. The signs are therefore contributing to sustainable safe speeds.

7 CONCLUSION

To conclude, this case study utilised the OLR model to test four predictors, three of which were age, gender of the driver and the time of day the driver was detected speeding. These are explanatory predictors, whereas the fourth one which is the 'period' tested the usefulness of electronic signs and complemented the other three predictors.

The study found that:

- The driver's age contributes slightly to risky speeding behaviours, and older drivers speed less.
- Time of the day was found to be significant in the model, with a higher number of TINs being recorded in the afternoon than in the morning.
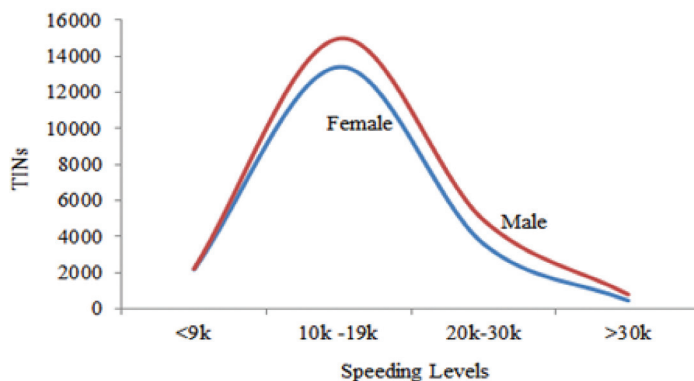
Figure 2: TINs issued for speeding over 40 km/h roads in Perth, by gender at four speeding levels, July 2007–June 2011.

- Although gender was not found to be a significant predictor, it was shown to be produce results similar to those for speeding behaviour with males speeding slightly more than females. This was more pronounced with higher speeding levels.
- The period variable in the model relating to the installation of the signs was significant, with drivers slowing down after the installation of the flashing electronic 40 km/h signs. This proves the usefulness of such signs in reducing speeding.

There are certain limitations of the model and are worth mentioning such as the period of the study needs to be expanded in order to obtain reliable data. Also, the before/after data in particular needs to be unified on the basis of the same number of hours. Importantly, the length of the time a driver has been licensed. In other words, the amount of experience they have had of driving. Finally, extra information on driver history and repeat offending needs to be included among the variables.

REFERENCES
[1] Moran, A., Report lower speed limits by as much as 20km/h Toronto. http://digitaljournal.com/article/323583#ixzz1vPsSSIOa, 24 April 2012.
[2] Vic Roads, Official web page, available at https://www.vicroads.vic.gov.au/safety-and-road-rules/driver-safety/speeding/speeding-and-safety, 2012.
[3] Gitelman, V., Balasha, D., Carmel, R., Hendel, L. & Pesahov, F., Characterization of pedestrian accidents and an examination of infrastructure measures to improve pedestrian safety in Israel. *Accident Analysis and Prevention*, pp. 1–11, 2010. doi: http://dx.doi.org/10.1016/j.aap.2010.11.017
[4] Nilsson, G., *Speed Accident Rates and Personal Injury Consequences for Different Road Types*. Rapport 277, Swedish National Road Transport Research Institute (VTI): Sweden, 1984.
[5] Kloeden, C.N., McLean, A.J., Moore, V.M. & Ponte, G., *Travelling speed and the risk of crashes involvement* (CR172). Federal Office of Road Safety: Canberra, 1997.
[6] O'Leary, C., Good drivers a danger. *The Weekend West*, May 14–15, 2011.
[7] Ebrahim, Z. & Nikraz, H., Before and after studies to reduce the gap between road users and authorities. *Nineteenth International Conference on Urban Transport and the Environment*, Greece, pp. 663–672, 2013. doi: http://dx.doi.org/10.2495/ut130531
[8] Robinson, C., Safety concerns as 96 drivers caught daily in 40 km/h zones. *The Sunday Time*, 23 October 2010.
[9] Yan, X., Radwan, E. & Abdel-Aty. M., Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model. *Accident Analysis and Prevention*, **37(6)**, pp. 983–995, 2005. doi: http://dx.doi.org/10.1016/j.aap.2005.05.001
[10] Yan, X., Radwan, E. & Mannila, K.K., Analysis of truck-involved rear-end crashes using multinomial logistic regression. *Advances in Transportation Studies: an International Journal*, **17**, pp. 39–52, 2009.

[11] Oxley, J., Charlton, J., Fildes, B., Koppel, S., Scully, J., Congiu, M. & Moore, K., *Crash Risk of Older Female Drivers*. Report No. 245. Monash University Accident Research Centre: Victoria, Australia, 2005.

[12] Lenny, M.G., Triggs, T.J. & Redman, J.R., Time of day variations in driving performance. *Accident Analysis and Prevention*, **29(4)**, pp. 431–437, 1997. doi: http://dx.doi.org/10.1016/s0001-4575(97)00022-5

[13] Wigglesworth, E., Occupational injuries by hour of day of week: a 20-year study. *Australian and New Zealand Journal of Public Health*, **30(6)**, pp. 505–508, 2006. doi: http://dx.doi.org/10.1111/j.1467-842x.2006.tb00776.x

[14] Camino Lópeza, M.A., Fontanedab, I., González Alcántarab, O.J. & Ritzel, D.O., The special severity of occupational accidents in the afternoon: "The lunch effect." *Accident Analysis & Prevention*, **43(3)**, pp. 1104–1116, 2011. doi: http://dx.doi.org/10.1016/j.aap.2010.12.019

[15] Banwell, C., Dance, P., Quinn, C., Davies, R. & Hall, D., Alcohol, other drug use and gambling among Australian Capital Territory (ACT) workers in the building and related industries. *Drugs: Education, Prevention and Policy*, **13(2)**, pp.167–178, 2006. doi: http://dx.doi.org/10.1080/09687630600577550

[16] Field, A., *Discovering Statistics Using SPSS*, 3rd edn., SAGE Publications Ltd.: London, UK, pp. 264–315, 2009. doi: http://dx.doi.org/10.1002/bjs.7040

[17] Peng, C.Y. & Nichols, R.N., Using multinomial logistic models to predict adolescent behavioural risk. *Journal of Modern Applied Statistical Methods* **2(1)**, pp. 1–13, 2003.

[18] Menard, S. Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, **54(1)**, pp. 17–24, 2000. doi: http://dx.doi.org/10.2307/2685605

[19] Konga, C. & Yanga, J., Logistic regression analysis of pedestrian casualty risk in passenger vehicle collisions in China. *Accident Analysis and Prevention*, **42**, pp. 987–993, 2010. doi: http://dx.doi.org/10.1016/j.aap.2009.11.006

[20] Al-Ghamdi, A.S., Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis and Prevention*, **34**, pp. 729–774, 2001. doi: http://dx.doi.org/10.1016/s0001-4575(01)00073-2

[21] Cameron, M., Newstead, S., Diamantopoulou, K. & Oxley, P., *The Interaction between Speed Camera Enforcement and Speed Related Mass Media Publicity in Victoria*. Report No: 201, Monash University Accident research centre: Melbourne, 2003.

[22] Washington, S.P., Karlaftis, M.G. & Mannering, F.L., *Statistical and Econometric Methods for Transportation Data Analysis*. Chapman & Hall/CRC, A CRC Press Company: Boca Raton, FL, pp. 257–295, 2003. doi: http://dx.doi.org/10.1201/9780203497111

[23] SPSS 18.0., SPSS Inc.: Chicago, IL 60606-6412, 2008. doi: http://dx.doi.org/10.1177/106480460100900407

[24] Petrucci, C.J., A primer for social worker researchers on how to conduct a multinomial logistic regression. *Journal of Social Service Research*, **35(2)**, pp. 193–205, 2009. doi: http://dx.doi.org/10.1080/01488370802678983

[25] Mountain, L.J., Hirst, W.M. & Maher, M.J., Are speed enforcement cameras more effective than other speed management measures? The impact of speed management schemes on 30 mph roads. *Accident Analysis and Prevention*, **37**, pp. 742–754, 2005. doi: http://dx.doi.org/10.1016/j.aap.2005.03.017

[26] Corben, B., Logan, D.B., Johnston, I. & Vulcan. P., *Development of a road safety strategy for Western Australia 2008–2020*. Report No. 282, available at http://www.monash.edu.au/miri/research/reports/muarc282.pdf, 2008.