Modelling of Road Traffic Accidents: A Multi-state Markov Approach

Bamidele Mustapha Oseni[*] and Ifeoluwa Hammed Anjorin

Department of Statistics, Federal University of Technology, Akure, Nigeria
[*]Corresponding Author: bmoseni@futa.edu.ng

ABSTRACT

Myriads of statistical techniques have been used to analyze road traffic accident data for causes; consequently serving as a guiding tool for policies made to achieve safer roads. However, relatively little is known about the progression and survival probabilities of road traffic accident victims admitted for treatment in health care institutions. Since the primary goal for safer roads is to save lives, this research work takes this primal one-step further through the analysis of the event history of road accident using the multi-state Markov model. Data of road traffic accident victims in year, 2014 at Akure, Nigeria were collected from both the Federal Road Safety Corps of Nigeria (FRSCN) and State Specialist hospital, Akure. Based on application of multi-state model, it was discovered that progression to injury state is 5 times more likely than death. Also injured victims are 6% more likely to recover from the injury than die. However, the transition probabilities that a victim will dieafter 1, 7 and 14 days of occurrence of the accident were obtained as 0.08, 0.39 and 0.61 respectively. Based on this, it is concluded that more effort should be intensifiedtowards achieving the Decade of action targets of the UN. Also the post-accident treatment of the victims of road accident should be improved as the victims are liable to die, the longer they state in the hospital.

Keywords: Multi-state Markov model, Transition Intensity, Transition probabilities, Road Traffic Accident, Road traffic victims, Event history.

1. Introduction

In 2010, the United Nation General Assembly declared the period 2011-2020 as the Decade of Action for Road Safety, with the intention of stabilizing and reducing the forecast level of road traffic fatalities to the barest minimum around the world (Waeg, 2010). Most agencies responsible for ensuring safer roads have devised various ways of preventing road traffic accidents. In

Nigeria for example, stringent measures are being meted out to motorists who break traffic laws and also large number of road safety officials are seen at motor ways during periods of high traffic intensity such as festive periods. A good number of studies have been conducted into investigating and forecasting road traffic accident based on various causes. While some researcher are interested at devising and examining various vehicle system that may reduce vehicular crashes (Aga & Okada, 2003), others investigate dangerous behaviors, conditions and factors that may lead to road traffic accidents and ways of ameliorating them.

From the methodological perspective, almost all of the statistical tools utilized are used with the aim to understand the causes of road traffic accidents and achieve safer roads. Predominant among the tools are time series, regression and correlation. Lassarre (1986) successfully employed multivariate time series approaches in developing a predictive model of severity of accidents. Jones and Whitfield (1988) modeled severity risk as a function of anthropometric measures, car mass, restraint system use and age of driver using logistic regression. Logistic regression was also utilized by Lui, McGee, Rhodes, and Pallock (1988) in modelling the probability of fatalities conditioned on the occurrence of an accident. Other studies have utilized bivariate probit analysis (Hutchinson, 1986), discriminant analysis (Shao, 1987), double-pair comparison (Evans, 1990), sequential estimation procedure using simple multinomial model (Shankar, Mannering, & Barfield, 1996).

Golob and Recker (2003) studied the relationships among accident types, flow of traffic, weather and lighting conditions using linear and nonlinear multivariate statistical analysis. Kadilar (2014) examined the factors that appear to have potential for serious injury or death of drivers in traffic accident using conditional logistic regression analysis. Given the rich literature in accident data research, especially the development of methods for real-time prediction as a function of current recent traffic and roadway conditions, a systematic review and meta-analysis was carried out by Roshandel, Zheng, and Washington (2015).

Despite the numerous works on road traffic accident, there has not been serious consideration of the survivability of road traffic accident victims. Amongst the few author who investigated survivability of road accident victims is Sevitt (1973). This researcher examined the time to death and their causes for fatal road accidents in Birmingham. This works adopts a multi-

state Markov model approach in monitoring the progression of road accident victims from the time of occurrence to survival or death.

Multi-state Markov model considers several discernable states of a process; a modeling process that allows one to compute transition probabilities by integrating certain functions of the transition intensities, transition rates, mean sojourn time (mean time spent in each state) and assess the dependence of transition intensities on certain covariates. In spite of the many advantages of applying multi-state models, these models are not often applied compared to classical survival analysis. Meira-Machado, deUña-Álvarez, Cadarso-Suárez, and Andersen (2009) state the reasons for this as daunting mathematical theory and lack of available software. As a multi-state process evolves over time, a history is naturally generated. This history contains information on previous states visited, times of entry into previous states and length of stay in states. The simplest and most applied multi-state model assumes that the transition to a future state is only dependent on the present state (Markov property) and that the transition intensities are constant over time (i.e. Time homogenous transition rates). These transition intensities provide the hazards for movement from one state to another and can also be used to determine the mean sojourn time in a given state. Covariates may be incorporated in models through transition intensities to explain differences among individuals in the course of treatments. The complexity of a multi-state model greatly depends on the number of states and also on the possible transitions (Zhao, 2009). A general model for accident data can be constructed in such a way that several intermediate states, like several stages of treatment in the hospital, are incorporated into the model.
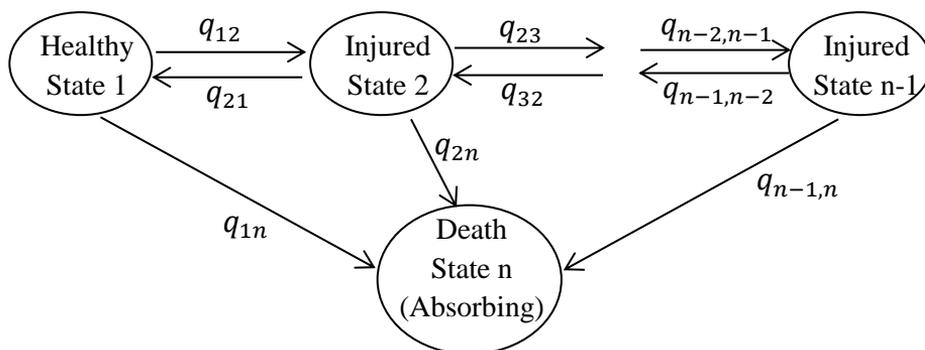

Figure 1: General model for accident victim progression.

There are other multi-state models which are less restrictive but more difficult to implement. The best data source for such comprehensive analysis consists of event histories i.e., records that contain the dates of all events of interest that happen to the victims. Such data are expensive to collect, but as suggested by Ogurtsova (2014) , other data sources, like panel data or cross sectional data can be relied upon. A simple multi-state structure based on three states referring to whether an accident victim is injured or healthy with death as the third and only absorbing state is considered in this work. The study involves following life progression of road accident victims from the time of occurrence of the accident, till they recover or experience the final critical event which is, death, so as to estimate model parameters that would provide insights to the dynamics of the process. These parameters includes the transition intensities and transition probabilities of events, the hazard rate of deaths after accident and deaths during injury based treatment, and the mean sojourn time spent in the injury state. The effect of covariates on the model parameters is studied and the goodness-of-fit of the model is considered.

2. Methodology

A multi-state process is a stochastic process $(X(t), t \in T)$ with finite state space $X = \{1, 2, 3 \ldots N\}$ where $T = [0, \tau]$ is the period of observation (Meira-Machado et al., 2009). As the process evolves over time a history F_{t-} consisting of the observation of the process over the interval $[0, t)$ or $[0, t]$ is generated. F_{t-} can be explained as the history of the process just before time t. As mentioned previously this history contains information on previous states visited, length of stay in the current and previous states, time of transition into previous states and other related information. The two quantities which completely characterize a multi-state process are transition probabilities and transition intensities. The transition probabilities may be defined by:

$$P_{ij}(s,t) = P(X(t) = j \,/\, X(s) = i, F_{t-}) \qquad (1)$$

for $i, \ j \in S \ and \ s, \ t \in T, \ s \leq t$ where $P_{ij}(s,t)$ can be interpreted as the probability of entering state j at time t, conditional on being in state i at time s and the history of the process prior to time s. The transition probabilities in the equation above form a transition probability matrix

$$P(s,t)=\begin{pmatrix} p_{11}(t) & p_{12}(t) & \dots & p_{1n-1}(t) & p_{1n}(t) \\ p_{21}(t) & p_{22}(t) & \dots & p_{2n-1}(t) & p_{2n}(t) \\ \vdots & & \dots & & \vdots \\ \vdots & & \dots & & \vdots \\ p_{n1}(t) & p_{n2}(t) & \dots & p_{nn-1}(t) & p_{nn}(t) \end{pmatrix} \quad (2)$$

whose rows summed up to unity. The transition intensities

$$q_{ij}(t,\ F_t-) = \lim_{\Delta t \to 0} \frac{p_{ij}(t,\ t+\Delta t)}{\Delta t}$$

(3)

is the instantaneous rate or hazard of making a transition from state i to state j at time t. Thus, the transition intensity matrix is defined as;

$$Q(t) = \begin{pmatrix} q_{11}(t) & q_{12}(t) & \dots & q_{1n-1}(t) & q_{1n}(t) \\ q_{21}(t) & q_{22}(t) & \dots & q_{2n-1}(t) & q_{2n}(t) \\ \vdots & & \dots & & \vdots \\ \vdots & & \dots & & \vdots \\ q_{n1}(t) & q_{n2}(t) & \dots & q_{nn-1}(t) & q_{nn}(t) \end{pmatrix} \quad (4)$$

with the property that $q_{ii} = -\sum_{i \neq j} q_{ij}$ for all $i \in S$.

Some characteristics of the victim such as age group and gender are of particular interest in this study. One may wish to know how susceptible the age groups or gender are to road traffic accident. These explanatory variables are included at each level of the model through a generalized regression as implemented by Jackson, Sharples, Thompson, Duffy, and Couto (2003). A proportional hazards model described by Marshall and Jones (1995), where $q_{rs}(t)$, the transition intensity matrix elements, are related to the covariates $z(t)$ at time t using

$$q_{rs}\big(z(t)\big) = q_{rs}^{(0)} = \exp(\beta_{rs}{}^{T} z(t)) \quad (5)$$

can be used to incorporate the covariates into each level of the model.

2.1 Maximum Likelihood

To simultaneously estimate the progression rate and the effect of mis-measurement and / or covariate, the method of maximum likelihood may be employed. Generally, in a road traffic accident model, it is assumed that the transition times between states and the time of death of the victim are known, though the state of the victim before death may not be known, depending on the model being considered. Suppose the transition is to the absorbing state, i.e. death, say S and the states before death, say X is unknown, then the contribution to the likelihood according to Jackson (2011) is given by

$$L_{i,j} = \sum_{X \neq S} p_{X(t_j),X}(t_{j+1} - t_j))q_{X,S} \quad (6)$$

where the sum is over all possible state which can be visited between $X(t_j)$ and S. Equation (6) contains transition probability which can be obtained using (Cox & Miller, 1965)

$$P(t) = \exp(tQ(t)) \tag{7}$$

where $P(t)$ and $Q(t)$ are defined in equation (2) and (4) respectively. If the exact transition times between the states are $(t_{i1}, t_{i2}, \ldots, t_{in_i})$ with no transition between the times; the contribution to the likelihood is given by

$$L_{i,j} = \exp(q_{X(t_j)X(t_j)}(t_{j+1} - t_j))q_{X(t_j)X(t_{j+1})} \tag{8}$$

Substituting equation (5) for the intensities in equation (8) gives the maximum likelihood estimates of the effect of the covariates.

A set of initial values is required since the maximum likelihood described in equation(8) above can only be obtained using numerical methods. Following the method describe by Jackson (2011), a crude initial estimate q_{ij} can be obtained using

$$q_{ij} = n_{ij}/T_i \tag{9}$$

where n_{ij} is the number of observed transitions from state i to state j and T_i is the number of days spent in state i.

2.2 Application to road traffic accident data

Multi-state model is applied to data obtained from records of road traffic accident victims in Akure, Nigeria from January to December 2014. The data is obtained from both the records of Federal Road Safety Commission of Nigeria (FRSCN) and the State Specialist hospital, Akure where accident victims are referred. The record from FRSCN contains the dates of the accident and the numbers of people involved in, the numbers that died and the numbers that got injured. The records of the victims that got injured are traced to the State Specialist hospital where both the numbers of victims that eventually died and those that eventually recovered after treatment are obtained. The victims that got injured but considers their injuries minor and left, are considered unaffected by the accident. The Demographic information collected about the victims, which includes their gender and age group, are used as the covariates. The age group is classified as above and below 18 years which is referred to in the work as adult and non-adult respectively. The records provide information about the time of arrival and time of departures. While arrivals refer to admissions only, departures refer to discharges which include recovery and/or deaths while being treated. The difference between these time intervals provides the length of stay in the hospital before eventual

death/ full recovery per individual and the unit of time is in days. Thus, the model considered is a three-state model with the states $X = 1, 2, 3$ referred to respectively as healthy, Injury and death as shown in the Fig. 2 below.
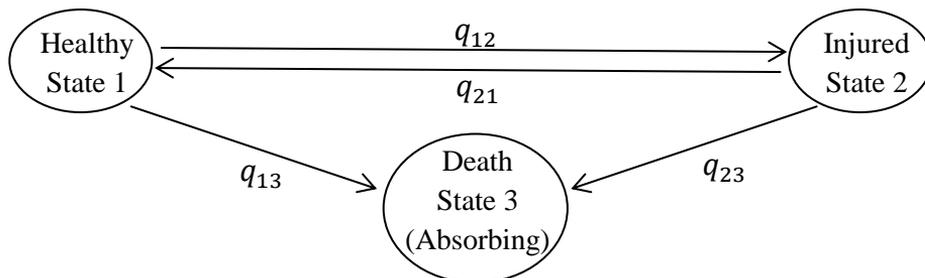

Figure 2: Three-state model for accident victims

It is assumed that all the states are transient apart from state 3 which is absorbing, thus a plausible transition intensity matrix for the model is

$$Q(t) = \begin{pmatrix} q_{11} & q_{12} & q_{12} \\ q_{21} & q_{22} & q_{23} \\ 0 & 0 & 0 \end{pmatrix} \tag{10}$$

The data contains 2077 individuals who are completely monitored from the time that the accident occurred till they were discharged. The observed state transition is depicted in the table 1 which is considered as a "snapshots" of the process. The table indicates that 891 individuals were injured, 185 individuals died without treatment and 1001 individuals who are unaffected after the accident are classed as healthy. 458 injured individuals recovered while 433 died while being treated. Individuals who were injured before and after the observation period are not included.

Table 1: The observe state transition

		To State		
		1	2	3
From State	1	1001	891	185
	2	458	0	433
	3	0	0	0

The simplest model for the data which is the time homogenous Markov model is fitted using the property of intensity matrix defined in equation(4). Thus, the modify intensity matrix is given in equation (11).

$$Q(t) = \begin{pmatrix} -(q_{12} + q_{13}) & q_{12} & q_{13} \\ q_{21} & -(q_{21} + q_{23}) & q_{23} \\ 0 & 0 & 0 \end{pmatrix} \qquad (11)$$

3. Results and Discussion

The transition intensity matrix is calculated by assuming exact transition times. The initial estimate of the transition rate matrix is given by

$$\begin{pmatrix} 0.10 & 0.20 & 0.48 \\ 0.20 & 0.00 & 0.50 \\ 0.00 & 0.00 & 0.00 \end{pmatrix} \qquad (12)$$

This estimate is used to compute maximum likelihood estimates of q_{ij} where $i \neq j$ and $i, j = 1$, 2, 3 which are given in Table 2. From the table, the progression to injury state is 5 times more likely than death (that is q_{12}/q_{13}). This implies that for every accident were 5 people sustained injury, one person is likely to die. In other words, a high percentage of the accidents that occur during the period under study is fatal. Also victims are 6% more likely to recover from the injury sustained than die from the injury (that is $100(q_{21} - q_{23})/q_{23}$). An average of 8 days (that is $-1/q_{22}$) is spent by injured victims in the hospital before discharge or death.

Table 2: Time Homogeneous Markov Model Parameter Estimates

q_{ij}	Estimate	95% CL
q11	-0.5180	(-0.5499, -0.4880)
q12	0.4290	(0.4017, 0.4580)
q13	0.0891	(0.0771, 0.1029)
q21	0.0644	(0.0588, 0.0706)
q22	-0.1253	(-0.1338, -0.1174)
q23	0.0609	(0.0554, 0.0669)

-2 * log-likelihood: 11272.5

Something, accident victims may begin to feel the effect of the crash few days after the occurrence of the accident. A suitable tool to measure a probable change of state by a victim after some days, is the transition probability p_{ij} which gives the probability that a victim in state i will move to state j. Table 3 shows the transition probabilities after day 1, 7 and 14. From the table, transition to state 3 from state 1 after 1, 7 and 14 days are given as 0.08, 0.39 and 0.61 respectively. This implies that an accident victim has a

probability of 0.08 of being dead a day after the occurrence of the accident, 0.39 a week after and 0.61 after two weeks. The probability that the victim is classified as injured after day 1, 7, and 14 are 0.31, 0.52 and 0.34 respectively. This implies that the probability that the effect of the crash is established a day after the accident is 0.31, 0.52 a week after and 0.34 after two weeks. After two weeks, the probability that a victim is confirmed dead is greater than both the probability that the victim is healthy and the probability that the victim is established as injured.

Table 3: Transition Probabilities at a specified day.

Day 1 transition probability			
	STATE 1	STATE 2	STATE 3
STATE 1	0.6051	0.3140	0.0805
STATE 2	0.0472	0.8930	0.0598
STATE 3	0	0	1
Day 7 transition probability			
STATE 1	0.0907	0.5170	0.3930
STATE 2	0.0776	0.5640	0.3590
STATE 3	0	0	1
Day 14 transition probability			
STATE 1	0.0483	0.3380	0.6140
STATE 2	0.0508	0.3580	0.5910
STATE 3	0	0	1

The transition intensities with the effect of the covariates are shown in table 4. There is 124% and 178% increments in transition intensity q_{12} (from healthy to injury state) when the road traffic accident is modeled in terms of age-group and gender as covariates respectively. Also, there is 1586% increment in transition intensity q_{23} when age-group is considered as a covariate in the model while there is 1693% increment when gender is considered as a covariate. These actually showed that the two covariates considered in the model have a considerable impact in the model.

Table 4: Comparison of the transition intensities modeled in terms of the covariates

q_{ij}	Estimate (No Covariates)	Estimates (Age-Group as Covariates)	Estimates (Gender as Covariates)
q_{12}	0.4284 (0.4012, 0.4575)	0.9576 (0.7257, 1.2640)	1.1946 (1.0382, 1.3750)
q_{13}	0.0876 (0.0756, 0.1014)	1.4702 (0.8628, 2.5050)	0.7222 (0.5131, 1.0160)
q_{21}	0.0644 (0.0588, 0.0706)	0.7225 (0.4735, 1.1020)	0.9732 (0.7984, 1.1860)
q_{23}	0.0607 (0.0552, 0.0668)	1.0232 (0.7074, 1.4800)	1.0882 (0.8913, 1.3290)

2 * log-likelihood: 11258

The expected and observed number of victims in each state is compared using the prevalence plot in fig. 1 below. This is a rough indication of goodness of fit of the model. From the plot, it is clear that the model overestimate the number of victims in state 1 from about day two onwards, though it becomes better as the days increases. Also the model underestimates the number of victims in state 2 from about day 2 to day 10. The expected number of victims that died is almost correctly estimated by the model till about day 10 when it begins to underestimate it.

4. Conclusion

An investigation has been carried out into the progression of accident victims from the moment when the accident occurred to the period when they fully recovered or died using data collected from 2077 road accident victims in 2014 at Akure, Nigeria. The data contained 891 injured individuals, 185 that died without treatments and 1001 that are seemly unaffected by the accident. 433 out of the 891 died while being treated and 458 eventually recovered. Using multi-state Markov model approach, it was observed that the probability of a victim progressing from injury state to dead state increases as the days progressed. This simply indicates that more efforts should be placed in the post-accident care. Also, the estimated results show that greater number of the accidents that occurred in the period under consideration were found to be fatal; one out of every 6 victims died. In other to achieve the Decade of Action, several preventive measures should be put in place to reduce the incidence of road traffic accidents.
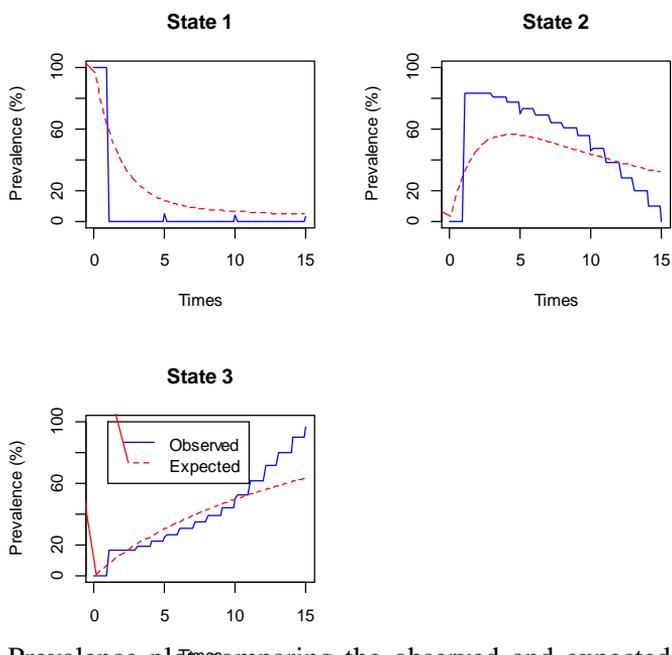
Figure 3: Prevalence plots comparing the observed and expected number of victims in each state.

References

1. Aga, M., & Okada, A. (2003). *Analysis of Vehicle Stability Control (VSC)'s effectiveness from accident data*. Paper presented at the 18th ESV Conference, Nagoya.

2. Cox, D. R., & Miller, H. D. (1965). *The Theory of Stochastic Processes*. London: Chapman and Hall.

3. Evans, L. (1990). The fraction of traffic fratalities attributable to alchohol. *Accident Analysis and Prevention, 22,* 587-602.DOI:10.1016/0001-4575(90)90030-O

4. Golob, T. F., & Recker, W. W. (2003). Relationships among urban freeway accidents, traffic flow, whether and lightning conditions. *Journal of Transport Engineering, 129*(4), 342-353. DOI:10.1061/(ASCE)0733-947X(2003)129:4(342)

5. Hutchinson, T. P. (1986). Statistical modelling of injury severity, with special reference to driver and front seat passenger in single-vehicle

crashes. *Accident Analysis and Prevention, 18*, 157-167. DOI: 10.1016/0001-4575(86)90060-6

6. Jackson, C. H. (2011). Multi-State Models for Panel Data: The msm Package for R. *Journal of Statistical Software, 38*(8), 1-29. DOI:10.18637/jss.v038.i08

7. Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., & Couto, E. (2003). Multistate Markov models for disease progression with classification error. *The Statistician, 52*(2), 193-209. DOI:10.1111/1467-9884.00351

8. Jones, I. S., & Whitfield, R. A. (1988). Predicting injury risk with "new car assessment program" crashworthiness ratings. *Accident Analysis and Prevention, 20*, 411-419. DOI:10.1016/0001-4575(88)90039-5

9. Kadilar, G. O. (2016). Effect of driver, roadway, collision, and vehicle characteristics on crash severity: a conditional logistic regression approach. *International Journal of Injury Control and Safety Promotion*, 23(2), 135-144. http://dx.doi.org/10.1080/17457300.2014.942323

10. Lassarre, S. (1986). The introduction of the variables "traffic volume," "speed" and "belt wearing" into a predictive model of the severity of accidents. *Accident Analysis and Prevention, 18*(129-134).DOI:10.1016/0001-4575(86)90057-6

11. Lui, K. J., McGee, D., Rhodes, P., & Pallock, D. (1988). Application of conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fratalities. *Journal of Safety Research, 19*, 197-203. DOI:10.1016/0022-4375(88)90024-2

12. Meira-Machado, L., deUña-Álvarez, J., Cadarso-Suárez, C., & Andersen, P. K. (2009). Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research, 18*(2), 195-222. DOI:10.1177/0962280208092301

13. Ogurtsova, E. (2014). Estimating transition rates for multistate models from panel data and repeated Cross-Sections. University Of Groningen.

14. Roshandel, S., Zheng, Z., & Washington, S. (2015). Impact of real-time traffic characteristic on freeway crash occurrence: Systematic review and meta-analysis. *Accident Analysis and Prevention, 79*, 198-211. DOI:10.1016/j.aap.2015.03.013

15. Sevitt, S. (1973). Fatal road accidents in Birmingham: Times to death and their causes. *Injury, 4*(4), 281-283. DOI:10.1016/0020-1383(73)90001-6

16. Shankar, V., Mannering, F., & Barfield, W. (1996). Statistical analysis of accident severity on rural freeways. *Accident Analysis and Prevention, 28*(3), 391-401. DOI:10.1016/0001-4575(96)00009-7

17. Shao, S. P. (1987). Estimating car driver injury severity in car/tractor-trailer collisions. *Accident Analysis and Prevention, 19*, 207-217. DOI:10.1016/0001-4575(87)90004-2

18. Waeg, G. (2010). A decade of action for road saftey [Press release]. Retrieved from www.who.int/roadsafety/decade_of_action

19. Zhao, L. (2009). Multi-state processes with duration-dependent transition intensities: statistical methods and applications. Simon Fraser University.