PREDICTING ROAD ACCIDENTS: A RARE-EVENTS MODELING APPROACH

Dr. **Athanasios Theofilatos,** National Technical University of Athens (NTUA), Athens, Greece
Prof. **George Yannis,** National Technical University of Athens (NTUA), Athens, Greece
Dr. **Pantelis Kopelias,** University of Thessaly (UTH), Volos, Greece
Mr. **Fanis Papadimitriou,** Attica Tollway Operations Authority – Attikes Diadromes S.A., Paiania, Greece

INTRODUCTION

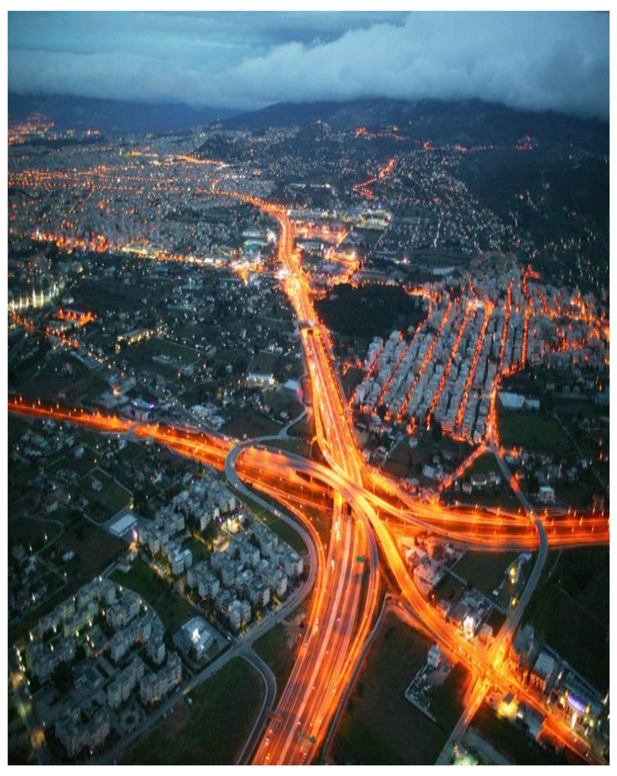
Modeling road accident occurrence has gained increasing attention over the years. Considerable efforts have been made from researchers and policy makers in order to explain road accidents and improve road safety performance of highways. In reality, road accidents are rare events. In such cases, the binary dependent variable is characterized by dozens to thousands of times fewer events (accidents) than non-events (non-accidents).

OBJECTIVE

The objective of this study is to investigate accident likelihood on freeways by utilizing real-time traffic data and by considering accidents as rare events.

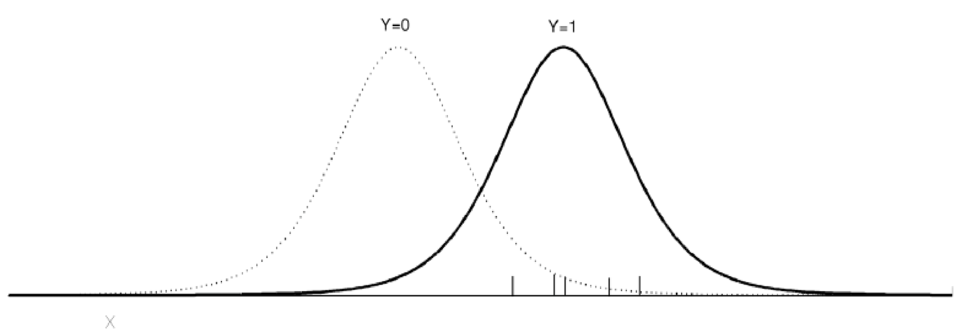
DATA

- ❖ The Attica Tollway (Attiki Odos) was chosen.
- ❖ Attiki Odos is a modern freeway in Greater Athens Area, Greece.
- ❖ Accident and non-accident cases for 2008-2011.
- ❖ 3 random locations
- ❖ Basic Freeway Segments (BFS) were considered.
- ❖ Real-time traffic data
  - ❖ Closest loop detectors
  - ❖ 1-hour intervals
  - ❖ Traffic flow
  - ❖ Speed
  - ❖ Occupancy
  - ❖ Percentage of heavy vehicles in traffic



METHODOLOGY

- ❖ Accidents are considered as rare-events.
- ❖ Traditional logit coefficients are biased.



- ❖ Rare-events logistic regression (King and Zeng, 2001a and 2001b).
- ❖ Case-control sampling design based on stratified sampling.
  - ❖ All events and a random selection of non-events.
  - ❖ A proportion of 1:10 for the ratio of events (accidents) to non-events (non-accidents) was used in each sample.

- ❖ A number of corrections are applied:
  - ❖ $\alpha_0 = \hat{\alpha} - \ln[(\frac{1-\tau}{\tau})*(\frac{1-\gamma}{\gamma})]$
  - ❖ α_0 is the new corrected constant term, α is the uncorrected constant term, τ is the proportion of events in the population and γ is the proportion of events in the sample.

  - ❖ The corrected logit function:
  - ❖ $\log itp_i = \ln(\frac{1-p_i}{p_i}) = a_0 + \sum \hat{\beta}_i x_i$
  - ❖ $p_i' = p_i + C_i$
  - ❖ C_i is the correction factor
  - ❖ $C_i = (0.5 - p_i) * p_i * (1 - p_i) * x_0 * V(\beta) * x_0'$
  - ❖ p_i is the probability of an event estimated using the corrected estimated coefficient a_0, x_0 is the 1*(m+1) vector of values for each independent variable, V(β) is the variance-covariance matrix, and lastly x_0' is the x_0 transposed.

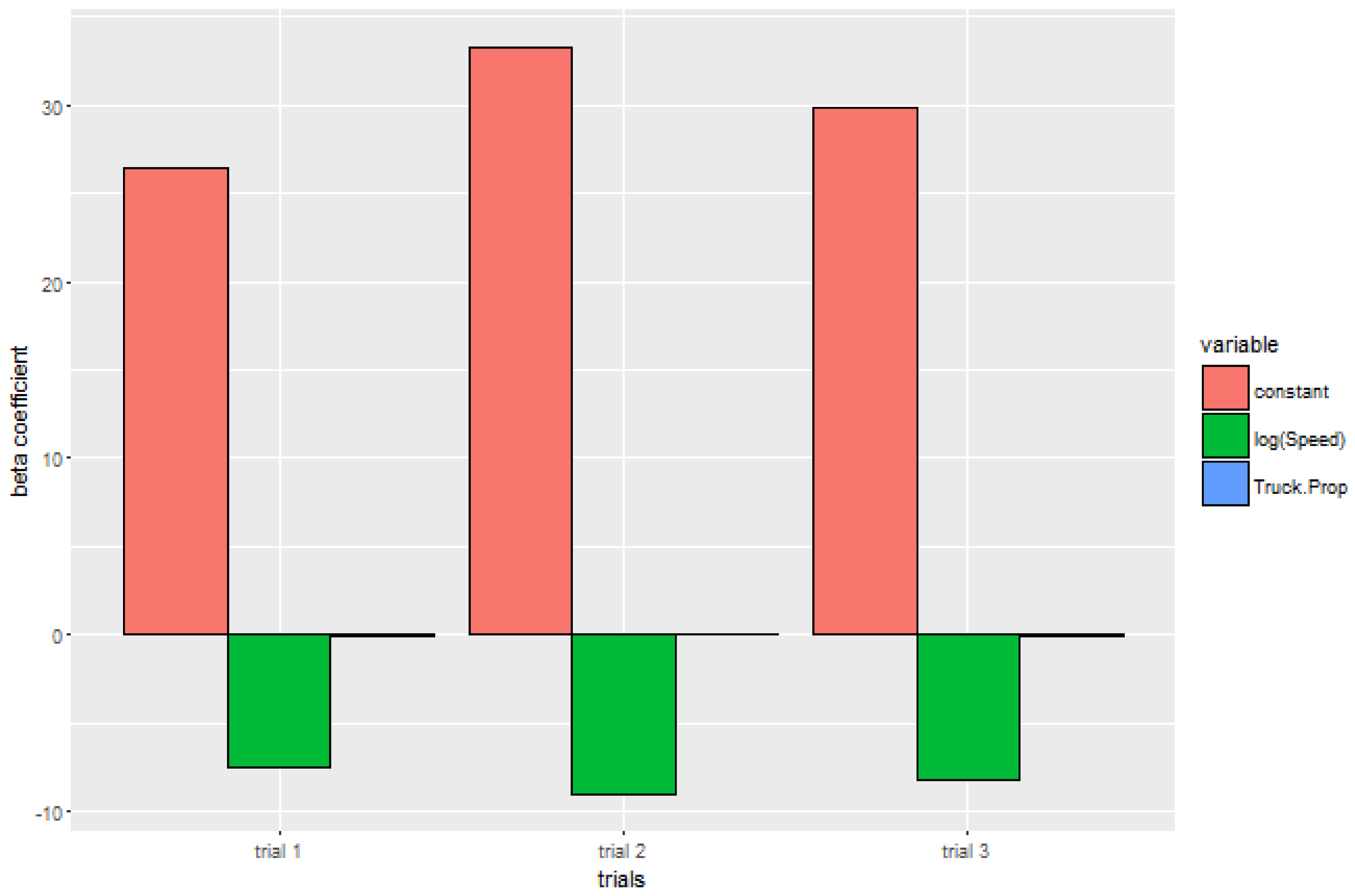
RESULTS

Trial 1	β	S.E.	z value	p-value
Constant	26.4158	11.3706	2.3232	0.0212
Truck.Prop.	-0.0394	0.1072	-0.3684	0.7129
log(Speed)	-7.4700	2.4369	-3.0653	0.0025
Log-likelihood at zero			-113.9	
Final log-likelihood			-100.9	
Likelihood ratio test			26.0	
AIC			106.9	
McFadden R^2			0.1141	

Trial 2	β	S.E.	z value	p value
Constant	33.2999	14.3741	2.3117	0.0216
Truck.Prop.	0.0157	0.0981	0.1597	0.8733
log(Speed)	-9.0004	3.0874	-2.9152	0.0039
Log-likelihood at zero			-113.9	
Final log-likelihood			-100.6	
Likelihood ratio test			26.6	
AIC			106.6	
McFadden R^2			0.1168	

Trial 3	β	S.E.	z value	p value
Constant	29.8363	12.6321	2.3619	0.0192
Truck.Prop.	-0.0444	0.0964	-0.4600	0.6460
log(Speed)	-8.2035	2.7063	-3.0311	0.0028
Log-likelihood at zero			-113.9	
Final log-likelihood			-100.8	
Likelihood ratio test			26.2	
AIC			106.8	
McFadden R^2			0.1150	

- ❖ Stratified sampling.
- ❖ 3 trials.
- ❖ All accident cases and a random sample of non-accident cases were included in each trial.
- ❖ The best models are presented.
- ❖ Adequate model fit.
- ❖ The risk factors were identified.
- ❖ Consistent effect of speed and truck proportion in the different trials.
- ❖ The logarithm of speed significantly affects accident occurrence.
- ❖ Consistent findings with past literature.



CONCLUSIONS

- ❖ First time application of rare-events model in road safety.
- ❖ Accident risk factors identified.
- ❖ Speed (logarithm) is found as the main risk factor.
- ❖ Confirmed that lower speeds increase accident likelihood.
- ❖ Model application in other case studies will assist real-time prediction of accident occurrence especially in locations with low number of accidents.

REFERENCES

Ahmed, M., Huang, H., Abdel-Aty, M., Guevara, B., 2011. Exploring a Bayesian hierarchical approach for developing safety performance functions for a mountainous freeway. Accident Analysis & Prevention 43, 1581-1589.
Ahmed, M., Abdel-Aty, M., Yu, R., 2012. Assessment of the interaction between crash occurrence, mountainous freeway geometry, real-time weather and AVI traffic data. Transportation Research Record 2280, 51-59.
King, G., Zeng, L., 2001a. Explaining rare events in international relations. International Organization 55(3), 693-715.
King, G., Zeng, L., 2001b. Logistic regression in rare events data. Political Analysis 9(2), 137-163.

ORGANISERS:
MINISTRY OF INFRASTRUCTURE AND CONSTRUCTION
European Commission
Instytut Badawczy Dróg i Mostów
Road and Bridge Research Institute
CEDR Conférence Européenne des Directeurs des Routes
Conference of European Directors of Roads
ERTRAC European Road Transport Research Advisory Council
ERRAC The European Rail Research Advisory Council
WATERBORNE
alice Alliance for Logistics Innovation through Collaboration in Europe
SUPPORTED BY:
ETRA
ECTP INNOVATIVE BUILT ENVIRONMENT
The National Centre for Research and Development