Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Applying Crash Data to Injury Claims

Julian Eggert

2018

Preprint:

Copyright Elsevier 2018. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/

Applying Crash Data to Injury Claims - An Investigation of Determinant Factors in Severe Motor Vehicle Accidents

Darren Shannon[1*]
Finbarr Murphy[1]
Martin Mullins[1]
Julian Eggert[2]

[1] University of Limerick, Ireland
[2] Honda Research Institute Europe GmbH, 63073 Offenbach/Main, Germany
[*] Corresponding author: Darren.Shannon@ul.ie

Abstract:

An extensive number of research studies have attempted to capture the factors that influence the severity of vehicle impacts. The high number of risks facing all traffic participants has led to a gradual increase in sophisticated data collection schemes linking crash characteristics to subsequent severity measures. This study serves as a departure from previous research by relating injuries suffered in road traffic accidents to expected trauma compensation payouts and deriving a quantitative cost function. Data from the National Highway Traffic Safety Administration's (NHTSA) Crash Injury Research (CIREN) database for the years 2005–2014 is combined with the Book of Quantum, an Irish governmental document that offers guidelines on the appropriate compensation to be awarded for injuries sustained in accidents. A multiple linear regression is carried out to identify the crash factors that significantly influence expected compensation costs and compared to ordered and multinomial logit models. The model offers encouraging results given the inherent variation expected in vehicular incidents and the subjectivity influencing compensation payout judgments, attaining an adjusted-R^2 fit of 20.6% when uninfluential factors are removed. It is found that relative speed at time of impact and dark conditions increase the expected costs, while rear-end incidents, incident sustained in van-based trucks and incidents sustained while turning result in lower expected compensations. The number of airbags available in the vehicle is also a significant factor. The scalar-outcome approach used in this research offers an alternative methodology to the discrete-outcome models that dominate traffic safety analyses. The results also raise queries on the future development of claims reserving (capital allocations earmarked for future expected claims payments) as advanced driver assistant systems (ADASs) seek to eradicate the most frequent types of crash factors upon which insurance mathematics base their assumptions.

Keywords: RTAs; Linear Regression; ADAS; Expected Compensation Costs; Claims Reserving.

Funding: This work was supported by the VI-DAS (Vision Inspired Driver Assistance Systems) – a Horizon 2020 research consortium designated by the European Commission [Grant number 690772].

1. Introduction:

Road Traffic Accidents (RTAs) result in an estimated 1.25 million deaths worldwide each year, and are the leading cause of death among young adults aged 15-29 (World Health Organisation 2013). The economic cost of RTAs and their impact on human society in general have led to concerted efforts from governments and corporations alike to enhance traffic safety, largely by way of a continuous stream of safety campaigns and regulatory mandates governing the safety features of vehicles. Road collision data collected by the automotive industry and government agencies have allowed researchers the opportunity to pinpoint the specific characteristics and contextual features that determine the severity of an incident. Statistical analyses using this data have proved critical in understanding the risks facing road users, as large volumes of literature have been dedicated to determining both the likelihood of entering into an incident, and the expected severity of the ensuing impact. Although the statistical techniques that are used vary widely, traditional literature has generally focused on predicting the damage sustained in an impact by means of an ordinal or probabilistic scale. A recent emphasis on improving the quality of data collection has led to a surge of databases with a wealth of information that offer more precise insights into the factors involved in RTAs. This study makes use of such data by generating a risk measure that associates the specific crash characteristics with the expected compensatory costs from the sustained injuries, while also exploring possible practical applications of this risk measure. As a result, the methodology presented in this article serves as a departure from categorical injury severity factors by utilising a proxy variable in order to quantify the expected severity of an incident.

The motivation for this research is prompted by recent developments in prospective ADAS technology, namely predictive systems, and offers further clarity on the influence of unmeasurable contextual factors in determining crash severity. Predictive systems operate by attaching multiple cameras and sensors to a vehicle. The systems then provide warnings of oncoming potential hazards to the driver, who in turn will take corrective actions. In some cases, these corrective actions may be

taken autonomously, as is the case with an Automatic Emergency Braking (AEB) system. Forecasting accurate trajectories are essential for an effective predictive system, and the implementation of forecasting frameworks have attracted significant attention in traffic safety literature. For example, the 'Intelligent Driver Model' is a traffic flow simulation technique used by on-board systems to alert drivers to oncoming perils (An and Harris 1996). A recent extension to this framework, the 'Foresighted Driver Model', has sought to improve its practical application by assuming that the user will attempt to balance risk-averseness with expedient driving. These on-board predictive models have made use of constant streams of continuous data to provide accurate results and have introduced cost functions. These cost functions attempt to capture the risks faced by the driver of the vehicle which arise in form of potential abnormal events (Eggert *et al.* 2015, Klingelschmitt and Eggert 2015, Damerow *et al.* 2016). This study introduces an alternative view of these cost functions by directly linking the risks faced by the vehicle with a compensation cost outcome, which could potentially be adapted to an on-board system as their development continues.

Hindering the widespread introduction of on-board predictive systems are the prevalence of unobserved factors that cannot currently be controlled for, also referred to as unobserved heterogeneity. Examples of this phenomenon facing road users are roadway design and detrition, weather conditions and lighting – all of which have been found to affect the severity of injuries sustained in motor vehicle accidents (Eluru *et al.* 2008, Buddhavarapu *et al.* 2013). Given the unpredictable variations that play a role in each incident, efforts to accurately predict the expected injury severity outcome have been hampered by the vast number of contextual factors that affect driving behaviour, road characteristics and environmental conditions. These studies attempted to categorise the extent of the incident along a five or six-level injury severity scale, resulting in a slight loss of granularity. This study introduces a continuous scalar measure of quantifying the injuries sustained in the accident by way of a 'compensation cost' loss function, found by relating injuries suffered in a road traffic accident to the expected trauma compensation. As a result, by offering a

quantitative insight in to the outcome of RTAs, 'compensation cost' loss functions can potentially capture the large variations in near-identical incidents.

A potential solution for underlying heterogeneity can be found in the anticipated introduction of commercial ADAS-enabled and semi-autonomous vehicles, which is expected to be a significant step toward mitigating and controlling the risks brought about by human driving behaviour. It is surmised that a gradual shift towards full autonomy will substantially reduce the hazards introduced by human error (Kyriakidis *et al.* 2015). A recent designation stated that over 95% of vehicles manufactured for U.S. consumption are to have Autonomous Emergency Braking (AEB) available as standard by 2022, reducing the number of traffic incidents by an estimated 20% or 5 million crashes (US DOT/NHTSA 2016). Meanwhile, the European Commission has encouraged a sustained adoption of emerging Advanced Driver Assistance Systems (ADAS) such as the aforementioned AEB, Collision Avoidance Systems, and adverse-environment Control Systems. These systems are designed to mitigate potential risks that occur throughout journeys and take immediate evasive action if a perilous situation arises. The impact that such evolutionary changes will have on the non-life insurance industry as vehicle safety comes increasingly to the fore is addressed in this study. Although current models function well in quantifying the risk faced by a given road user, the factors that contribute to an accurate derivation of risk pricing may wholly alter as ADAS function become increasingly prevalent. As such, models typically used in underwriting mathematics may need to adapt over time. As well as offering an incisive view of the outcome of RTAs, the framework introduced as part of this research can adapt to shifting RTA statistics as regulatory changes are introduced.

Mannering and Bhat (2014) perform a comprehensive literature review of the current problem space and detail the evolution of the methodological approaches used in accident research. They also pay close attention to the issues that underlie the estimation of accident prediction. Of greatest concern is unobserved heterogeneity, i.e. the lack of unmeasurable variables that significantly affect

the frequency and severity of RTAs. Mannering *et al.* (2016) also target this issue and offer statistical solutions that seek to minimise the impact of unobserved heterogeneity. One commonly used approach is the introduction of random parameter models, which allow for variations in underlying sensitivities (Eluru *et al.* 2008, Kim *et al.* 2008, Ye and Lord 2011, Cerwick *et al.* 2014). Another emerging approach is the use of latent class, or finite mixture, models. These models are used to capture injury severity levels by identifying underlying subgroups that affect known variables of interest and subsequently incorporating the probabilities of their occurrence into the model. Examples of studies adopting this approach are Cerwick *et al.* (2014), Yasmin *et al.* (2014a), and Yasmin *et al.* (2014b).

A further matter addressed by Mannering and Bhat (2014) is that of selective bias, which could allow erroneous conclusions to be drawn. For example, when analysing safety measures that are put in place solely at accident hotspots, the inferred results may underestimate the initiative's impact on the wider domain. Furthermore, the opposite effect could occur – the risk compensation that drivers take under adverse conditions makes accurate estimations of risk difficult (Winston *et al.* 2006). Finally, as is the case with many traffic safety analyses featuring crash data, the under-reporting of less severe accidents is raised (Yamamoto *et al.* 2008, Ye and Lord 2011). Given that the injuries reported on in this paper concern RTAs in which a severe injury is sustained, this is a factor that also affects this research.

In terms of empirical studies, Mannering and Bhat (2014) and Mannering *et al.* (2016) highlight the vast body of literature surrounding vehicle crash analysis, and detail a large number of methodological approaches used when modelling crash-related data. Beginning from univariate models finding associations between crash factors and specific injury severities, the complexity of statistical analyses quickly expanded in an attempt to capture the large variations that are inherent in RTAs. Shankar and Mannering (1996) and O'Donnell and Connor (1996), amongst others, developed discrete multivariate model approaches. Research typically focused on ordered probit

models, as exemplified by Kockelman and Kweon (2002) and Eluru *et al.* (2013). O'Donnell and Connor (1996), Kim *et al.* (2008) and Lemp *et al.* (2011) acknowledge the inherent randomness of RTAs by investigating links between attributes of road users and the levels of injury endured in incidents using heteroskedastic ordered logit and probit models. More generally, the research conducted by O'Donnell and Connor (1996) highlight through the use of ordered probit and logit models that, amongst other factors, speed, age of vehicle, and seatbelt use are significant indicators when predicting crash severity. They achieve reasonably high pseudo–R^2 values (36% and 35.6% respectively) in doing so.

Abdel-Aty (2003) and Ye and Lord (2011) examine a number of models in order to identify worthwhile predictors of injury severity. While Ye and Lord (2011) make use of ordered probit, multinomial logit, and mixed logit models, Abdel-Aty (2003) uses ordered probit, multinomial logit and nested logit models to find significant relationships between injury severity and age, gender, seatbelt use, point of impact, vehicle type and speeding. These results were found by examining incidents that occurred in central Florida at certain periods between 1996–2000. The multinomial logit approach was found to produce the poorest goodness-of-fit measure (10.5%), while the nested logit model had the greatest result at 23.9%. However, given the difficulty estimating the optimal nesting structure, the research recommended the use of ordered probit models (which achieved a highest fit of 20%).

Yamamoto *et al.* (2008) identify a useful solution to mitigate the information loss stemming from unknown variables using random parameter ordered and binary probit models. Meanwhile, Ye and Lord (2011) adopt the additional use of a mixed logit model in order to address the issue of underreported data, completed by measuring each model's performance at different estimates of unreported incidents. Recent focus has also shifted to the use of generalised ordered response models in an attempt to account for the unmeasurable variables that greatly influence RTA severity. Eluru *et al.* (2008) and Cerwick *et al.* (2014) detail a generalised logit variant and latent class method

respectively. This generalised alternative (also known as a proportional odds model) relaxes the restrictive assumption of the traditional ordered response model by including the impact of exogenous variables on injury severity levels, and have been show to offer at least as much flexibility as unordered response models (Mooradian *et al.* 2013, Yasmin *et al.* 2014b, Eluru and Yasmin 2015).

This article differs by introducing a scalar response variable instead of an odds-favoured or probabilistic outcome. By introducing the flexibility of a fully continuous output, the extent of a road traffic accident can be explored in minute detail. Furthermore, the expected 'cost' of the crash itself offers a partial view of the economic impact of road traffic accidents. Software developers could also potentially adapt such a model to generate an on-board navigation system that operates by navigating the path of least economic risk. The shifting landscape of the non-life insurance industry also benefits from the precise results offered by a continuous-result model as ADAS technology continually alters the most common risks faced by road users. This study offers a brief discussion on the advancement of ADAS and its relation to insurance prior to concluding. The remainder of the paper is contained within two sections. Section 2 outlines the CIREN database used as part of this research, and describes both the process used to collate the different sources of data and the procedure used to associate the expected medical costs with the injuries sustained in each incident. The formation of the methodological model of choice used in this paper (log-log linear model) is described in the Results section, alongside derivations of the ordered logit and multinomial logit model. This is followed by an examination of the results and a sample case study. This section will also offer a discussion on the significance of the results found, and their implications for the automotive industry as a whole, before the paper's conclusion.

2. Data Description & Extraction:

2.1. Data Source:

The data examined in this paper is derived from the National Highway Traffic Safety Administration's (NHTSA) Crash Injury Research (CIREN) database for the years 2005–2014. The establishment of the

CIREN program allowed for a combination of both medical and engineering research, with the goal of promoting traffic safety and reducing human and economic costs stemming from RTAs. The inclusion requirements for CIREN are quite strict and as a result, data is somewhat limited. Patients included in the sample must have sustained at least one serious or disabling injury based on the Abbreviated Injury Scale (AIS) system. In addition, the injuries must have been sustained in a vehicle no greater than 8 years old at the time of the incident, and the injured party must have been treated at a CIREN-participating trauma centre in the United States.

After combining the crash characteristics with the occupants involved, and collating all injuries sustained by each patient, the dataset totals $N = 1{,}274$ incidents involving 1,439 injured parties. In order to ensure a corroboration between the damage sustained by the vehicle and it's speed at time of impact, each case has an estimate determining the level of confidence in the reconstruction of the incident. This level of confidence is determined by technicians who examine both vehicle and scene data. Only accurate reconstructions are retained, to ensure optimal accuracy in the derived results, which reduces the final sample to $N = 425$ cases. The compensation costs for the injuries sustained are gleaned from the Book of Quantum (Personal Injuries Assessment Board 2016). The Book of Quantum is an Irish governmental document that offers guidelines on the appropriate compensation to be awarded for accidents in which there is medically supported evidence of pain and suffering. The document offers a range of values that are expected based on the most severe injury sustained in the incident, categorised similar to the AIS system. The scale is then adjusted based on the number of secondary injuries and their associated severities. Acting under the assumption that the physiological differences in drivers in both the United States and Ireland are minimal, the severity of RTAs can be taken as constant between the two territories. As such, it is assumed that the Book of Quantum is an appropriate set of guidelines to measure the expected compensatory costs stemming from RTAs.

Full details for combining the data into a single database are outlined as follows:

1. Variables that may serve as predictive indicators of the severity of an incident are gathered. 13 variables are chosen, comprising of Barrier Equivalent Speed (BES) (X_1), Age of Vehicle (X_2), Number of Airbags Available (X_3), Evidence of Alcohol Consumption (X_4), Accident Type Category (X_5), Weather Conditions (X_6), Vehicle Body Type (X_7), Lighting Conditions (X_8), Road Surface Conditions (X_9), Road Alignment (X_{10}), Movement Prior to Crash (X_{11}), Evidence of Speeding (X_{12}), and Evidence of Seatbelt Use (X_{13}). Most categorical variables contain multiple levels; for example, there are 11 varieties of crash type included within X_4. In order to conform to the assumptions that define a regression analysis, each level of a multi-category variable is included as a separate binary variable. Following these transformations, there are 38 x_i independent variables within the sample (Table 2).
2. All injuries associated with each individual in the dataset are collated, using the individual's CIREN ID as a unique identifier. A separate set of case numbers issued for each incident are available alongside the CIREN IDs, which allowed for the patient(s) to be connected to any given incident. As such, injuries can be associated with crash factors directly.
3. The 7-digit AIS codes, which serve as descriptors of the specific injuries suffered by each patient, are separated to allow for an association with the expected compensation costs. Given the level of detail provided by the Book of Quantum, only 3-digit identifiers (AIS-code digits 1, 2 and 7) are included in the analysis. Digit 1 describes the body region affected (Head, Thorax, etc.), digit 2 specifies the anatomical structure that was damaged (Skin, Bone, etc.), and digit 7 reports the severity of the injury (1 = Minor, 2 = Moderate, 3 = Serious, 4 = Severe, 5 = Critical, 6 = Maximum). Special compensations are also taken – for example, AIS codes beginning with 14 (Head, Organ) and 16 (Head, Loss of Consciousness) are assumed to be brain traumas, while fractures in the abdominal region are assumed to be spinal fractures.
4. Guideline compensation cost ranges are extracted from the Book of Quantum and combined with the 3-digit identifier found in step 3. The injury severity categories originally associated

with AIS digit 7 are straightforward to match. Since the Book of Quantum does not offer guidelines for non-survivable and fatal injuries, AIS code '6' injuries are referred to as critical injuries and recoded as AIS code '5'. Therefore, minor injuries are associated with AIS Code 1, Moderate injuries with AIS Code 2, Moderately Severe injuries with AIS Code 3, and Severe and Permanent Conditions are assigned AIS Codes 4/5. Injury costs are found by taking the average of the ranges provided in the Book of Quantum. In cases where there are multiple ranges for similar injuries (for example, broken finger and broken thumb have different ranges of values to consider), the average of the averages are taken, and so on. Furthermore, consideration must be paid to individuals who suffered multiple injuries in an incident. It is stated within the Book of Quantum that when considering the case of multiple secondary injuries, *"it is not appropriate to simply add up values for all the different injuries to determine the amount of compensation. Where additional injuries arise there is likely to be an adjustment within the value range"* (Personal Injuries Assessment Board 2016). In order to ensure conformity to the ranges specified within the Book of Quantum, the final cost was derived by providing a multiplicative weight of 0.6 to the primary injury (to ensure it is on the low scale of its range), and a multiplicative weight of 0.25 was added for each secondary injury. The distributions of the final costs of injury are provided in Figure 1.

5. The costs derived in step 4 are associated with the crash factors, using the incident case numbers as a link variable. Additionally, for incidents in which there are multiple vehicles with serious injuries, each vehicle was treated as a separate case.
6. Finally, in order to ensure optimal accuracy, only accurate recreations of the crash events are included in the analysis. Inaccurate results, as well as incidents where there are no reconstructions or rudimentary reconstructions, are removed. After applying this filter, a sample size of $N = 425$ remained from the initial $N = 1{,}274$.

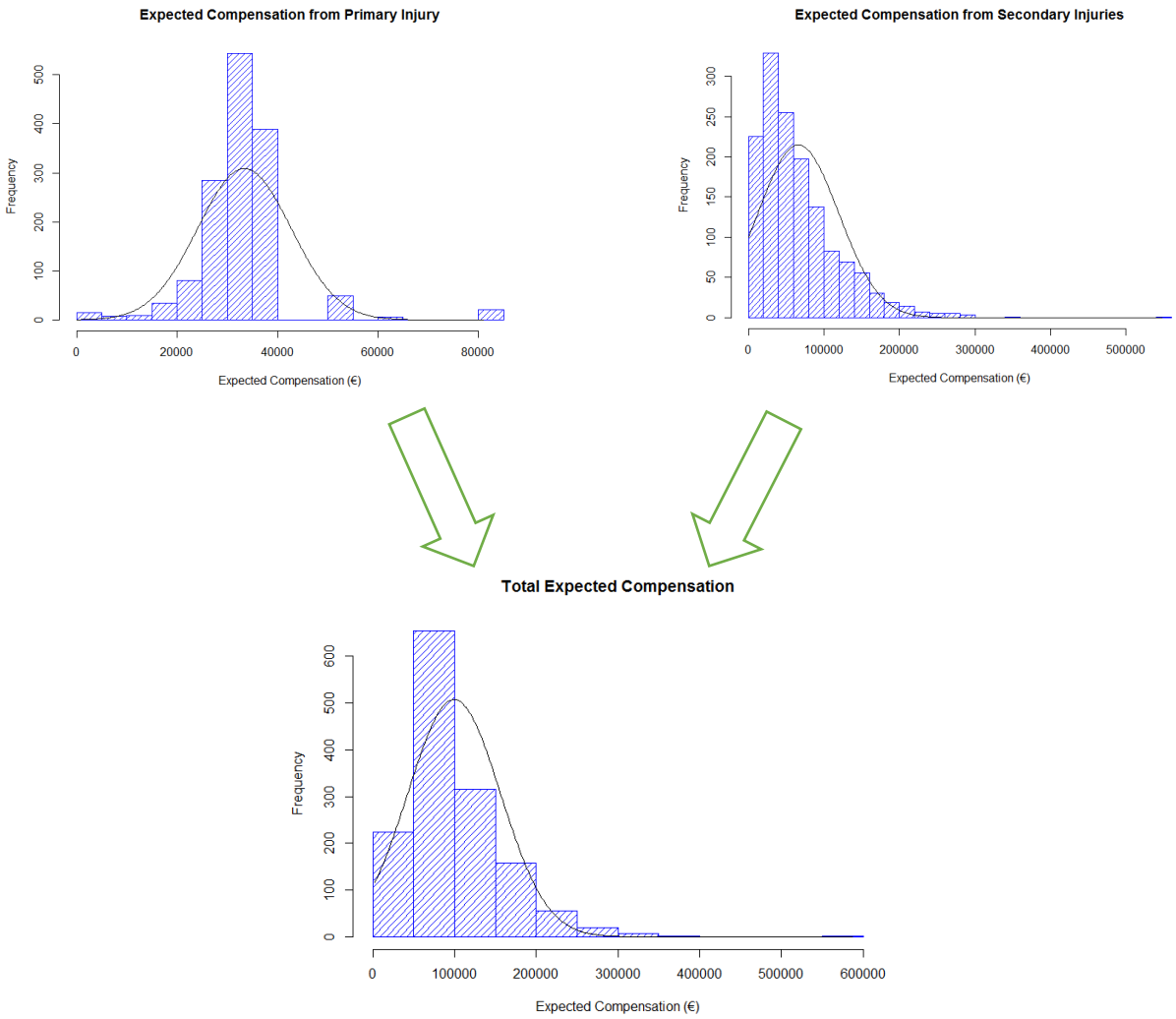


Figure 1 - Distributions of Primary and Secondary Injuries (scaled down by 0.6x and 0.25x multipiers respectively), as well as the combined distribution of the unfiltered dataset $(N = 1{,}274)$.

The unfiltered sample ($N = 1{,}274$) (Figure 1, bottom) discovered in step 4 consisted of a mixture of normal and lognormal distributions (Figure 1, top), and this general shape is retained in the filtered sample (Figure 2). When the distribution of the $N = 425$ data points is broken down further (Figure 2), there is a resemblance to a Poisson-influenced Gamma distribution. The median expected compensation in the filtered sample set is €89,747.33. The final distribution of the expected compensation payouts is highly skewed. In order to carry out a linear examination the costs are transformed logarithmically.

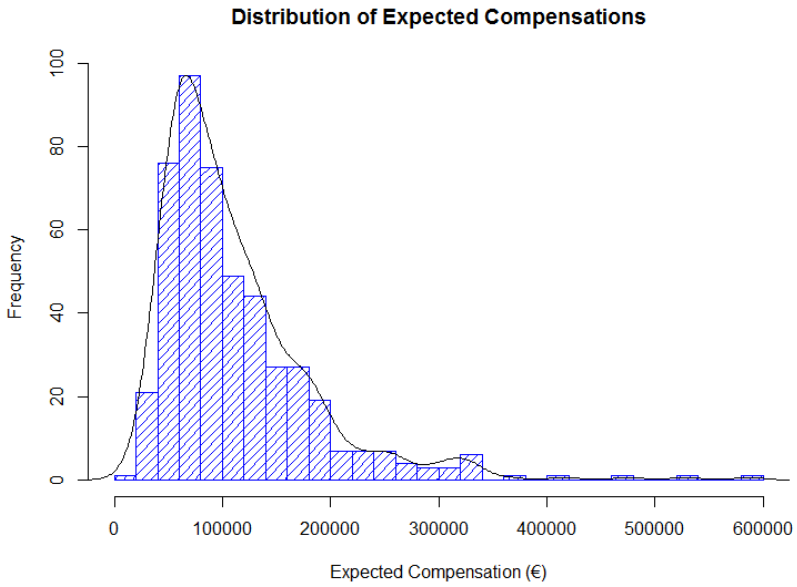


Figure 2 - Distribution of the filtered sample set $(N = 425)$ *describing the total expected compensation cost.*

2.2. Using Expected Compensation Costs as a Proxy for Incident Severity:

Severity measures describing crash impact factors directly, such as 'Barrier Equivalent Speed' (BES) and impact energy (measured in Joules) have previously served as proxies for impact severity, and are estimated using reconstruction studies (Jones and Champion 1989, Smith *et al.* 2005). BES, in the context of Road Traffic Accidents (RTAs), is the impact velocity required between a vehicle and a fixed barrier to achieve the same magnitude of crush energy that was evident in the original incident. Given the relative changes in stiffness between objects in a traffic environment (structures, pedestrian, vehicle, etc.), these measures are subject to exiguous flaws but remain mainstays in impact-severity modelling.

In contrast, throughout this examination, expected payouts stemming from RTAs are used as a proxy for the severity of injuries sustained in an incident. The boxplot in Figure 3 highlights the strong positive link between expected compensation costs and the Maximum Abbreviated Injury Scale score (MAIS) sustained as part of a crash, which is often the injury severity scale of choice in traffic accident analyses. This linear relationship is in accordance with previous research (Finkelstein *et al.* 2006). As such, it is appropriate to infer injury severities based on compensation payouts.

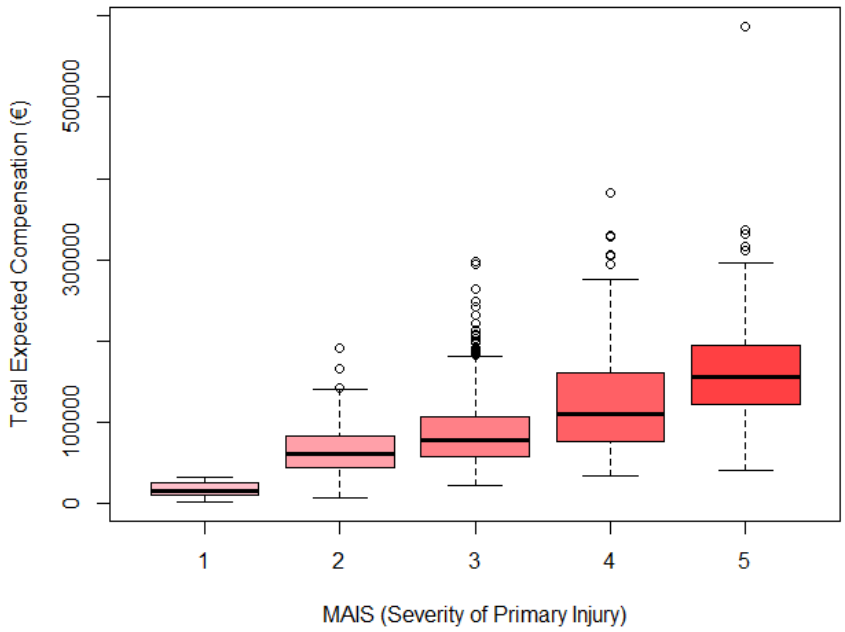


Figure 3 – Boxplot generated by comparing the overall compensation cost expected from an incident with the highest injury score (MAIS) sustained in the incident. It is highly evident that expected compensation costs increase as the severity of the primary injury increases (where '1' denotes minor injury and '5' denotes severe injury).

3. Results:

3.1. Model Formations:

As discussed previously, crash injury analyses utilise a variety of categorical-outcome models in an attempt to predict the severity of traffic incidents. These techniques range from logistic regressions to structured equation models. This study makes use of a linear regression with a mixture of log-linear and log-log transforms as an alternative to these categorical-outcome models. In order to examine the preliminary effectiveness of this model, it will thereafter be compared to ordinal logit and multinomial logit models, both of which feature heavily in crash injury analyses. While the linear regression model will use the expected compensation cost as its dependent variable, the logit regression models will use the MAIS variable as their outcome variable. Given the established relationship between expected cost and MAIS, the results of all models should be comparable.

The objective of the linear regression model is to find an optimally weighted average of the predictor variables to predict the scalar-outcome variable. The log-transforms performed on the skewed

variables have normalised the data and removed traces of heteroskedasticity, so an ordinary least squares (OLS) approach is an appropriate measure to minimise the error terms. This model takes the form

$$\ln(Y) = \beta_0 + \beta_1 \ln(X_{i1}) + \beta_2 (X_{i2}) \dots + \beta_k X_{ik} + \varepsilon_i.$$

β_0 is the default expected compensation when none of the variables describing the incident apply. Each subsequent β_k is the weight applied to each of the variables X_{ik} that are relevant to each observation. An inspection of each independent variable highlighted that Barrier Equivalent Speed (X_1) cases follow a lognormal distribution and so must also be transformed logarithmically in order to conform to the assumptions of the model. ε_i is the error term for individual i, quantifying the difference between the predicted value and the actual outcome. In order to ensure that each variable added to the model holds significant explanatory power, the adjusted-R^2 measure is reported. The adjusted-R^2 measure is an adaptation of the coefficient of determination, and penalises the addition of excess variables.

The standard ordinal logit model follows a similar outcome equation to the linear model, however in this case the observable ordinal injury severity levels y_i are assumed to follow an unobservable continuous latent variable y_i^*. Therefore, the outcome variable becomes

$$y_i^* = \beta_0 + \beta_1 \ln(X_{i1}) + \beta_2 (X_{i2}) \dots + \beta_k X_{ik} + \varepsilon_i$$

Assuming there are $j = 1,2, \dots, J$ injury severity levels and τ_j are the thresholds associated with the severity levels, the latent variable y_i^* is related to the discrete y_i by the threshold τ_j with a response in the form of

$$y_i = j, \;\; if \; \tau_{j-1} < y_i^* < \tau_j, \;\; for \; j = 1,2, \dots, J,$$

Where τ_j is ascendingly ordered and $\in (-\infty, \infty)$. Given these relationships across all response weights β_k, the probability expressions for observation i and alternative j for the ordinal logistic model takes the form

$$\pi_i(j) = \Pr(y_i = j|X_i) = \Phi(\tau_j - \boldsymbol{\beta X_i}) - \Phi(\tau_{j-1} - \boldsymbol{\beta X_i}),$$

Where π_{ij} indicates the injury level probability and $\Phi(\cdot)$ represents the standard logistic cumulative distribution function.

The ordering condition is not necessary for the multinomial logistic model, which is otherwise similar to the ordered logistic model. Although information is lost when disregarding the inherent order of the injury severities, it nevertheless allows the most likely injury category to be chosen for a given observation. This takes the form

$$\pi_i(j) = \Pr(y_{ji} \geq y_{Ji}) \; \forall \, J \neq j,$$

Where y_{ji} signifies the injury level j for observation i, and y_{Ji} signifies all other injury level alternatives for observation i. This is expanded to

$$\pi_i(j) = \Pr(\boldsymbol{\beta_j X_i} + \varepsilon_i \geq \boldsymbol{\beta_J X_i} + \varepsilon_i) \; \forall \, J \neq j,$$

or

$$\pi_i(j) = \Pr(\boldsymbol{\beta_j X_i} - \boldsymbol{\beta_J X_i} \geq \varepsilon_{iJ} - \varepsilon_{ij}) \; \forall \, J \neq j,$$

in which case the probability segment can be assigned based on

$$\pi_i(j) = \frac{\exp(\boldsymbol{\beta_j X_i})}{\sum_J (\boldsymbol{\beta_J X_i})}$$

if the errors are assumed to follow a GEV process. The output of the linear regression model is provided in Table 2, while the full ordinal logit and multinomial logit model results are available in the appendix as Tables A.1 and A.2 respectively.

3.2. Model Results:

Including all 38 variables within the linear regression analysis yields a penalised-fit of 0.186, or 18.6% (Table 2). Control variables are required for each multi-level category so that relative comparisons may be made. These variables are omitted from direct analysis and are determined by their

frequency – the highest number of cases within each category is chosen for omission. After removing non-significant variables through a process of stepwise regression with a $p < 0.1$ cut-off, 6 variables remained – the relative speed measure at the time of the incident (BES), the number of airbags available in the vehicle, evidence of turning at the time of impact, van-based light trucks, dark conditions, and rear-end incidents. These variables were deemed to have a significant influence on the expected compensation costs stemming from traffic incidents, and will be examined in the following subsection. Meanwhile, the overall fit of the model provided in Table 2 proves highly significant ($p < 0.001$) indicating that, collectively, the variables in the model are of practical significance. Both the ordinal logit (Table A.1) and multinomial logit (Table A.2) model also yield significant fits, yielding log-likelihood changes equivalent to $p = 0.011$ and $p < 0.001$ respectively. The variables initially highlighted as influential in these models vary slightly from the variables in the linear model. Both categorical models identified that relative speed at time of impact, the number of airbags available, and the current age of the vehicle are significant predictors of the highest injury level (MAIS). The ordinal model additionally signifies that curved roads, vans, and crashes sustained when turning into established traffic influence the injury severity level, while the nominal model signifies that speeding vehicles, SUVs and cloudy conditions influence the injury severity level. The full results are available in the Appendix.

In order to compare the effectiveness of each model in this analysis, the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Log-Likelihood estimates of the models are presented in Table 1. This comparison is valid between the two categorical models but is indicative only when comparing the linear model to the categorical models because of the different dependent variable. Although the nominal model has a higher Log-Likelihood estimate (–391.61) than the ordinal model (–443.51), the ordinal model outperforms the multinomial model in terms of both AIC and BIC estimates (957.02 and 1098.84 vs. 981.22 and 1382.376). However, both categorical models are outperformed by the linear regression model, which registers a higher Log-Likelihood (-300.26) and lower AIC/BIC estimate (668.53/806.30) than the other models.

	$Dependent$	AIC	BIC	$Log-Likelihood$
Multiple Linear Regression	LnCost	668.53	806.30	–300.26
Ordinal Logistic Regression	MAIS	957.02	1098.84	–443.51
Multinomial Logistic Regression	MAIS	981.22	1382.376	–391.61

Table 1 – Comparison of the models used in the analysis. The ordinal and multinomial models provide mixed results, however the linear model outperforms both. Note: The comparison of the linear regression to the two logistic models must be cognisant of the different dependent variables.

Model Summary: $R^2: 0.243, Adjusted - R^2: 0.182, Std. Error: 0.511$ **Significance of Fit:** $F: 3.94, Sig.: < 0.001$ **Goodness of Fit:** $AIC: 668.53, BIC: 806.30$		Unstandardized Coefficients			Collinearity Statistics
		B	Std. Error	Sig.	VIF
Constant (Y)	LnCost	9.529	.346	.000	
Barrier Equivalent Speed (Log-Transform) (X_1)	LnBES (x_1)***	.533	.070	.000	1.521
Age of Vehicle (X_2)	CurrentVehicleAge (x_2)	.007	.010	.408	1.690
Number of Airbags Available (X_3)	AirbagsAvailable (x_3)*	-.033	.018	.093	1.748
Evidence of alcohol consumption (X_4)	DrinkDriver (x_4)	-.025	.092	.799	1.387
Crash Category (X_5)	RightRoadsideDeparture (x_5)	-.024	.087	.973	1.652
	LeftRoadsideDeparture (x_6)	.057	.091	.398	1.609
	SingleDriverForwardImpact (x_7)	.063	.227	.699	1.181
	RearEnd (x_8)	-.259	.169	.182	1.322
	OppositeDirectionImpact (x_9)	.026	.177	.761	1.318
	Head-On (x_{10})		*(Control Variable)*		
	Angle/Sideswipe (x_{11})	-.089	.167	.597	1.150
	TurnAcrossPath (x_{12})	-.083	.103	.420	1.946
	TurnIntoPath (x_{13})	.008	.113	.944	2.209
	IntersectingPaths (x_{14})	.023	.099	.819	1.778
	BackingUp (x_{15})	.147	.130	.261	1.302
Weather Condition (X_6)	FogSmoke (x_{16})	.341	.334	.307	3.760
	Rain (x_{17})	-.093	.123	.454	1.876
	Snow (x_{18})	-.302	.410	.461	3.809
	Clear (x_{19})		*(Control Variable)*		
	Cloudy (x_{20})	.108	.087	.211	1.180
Vehicle Body Type (X_7)	AutomobileBodyType (x_{21})		*(Control Variable)*		
	SUVBodyType (x_{22})	.004	.072	.955	1.155
	VansLightTrucksBodyType (x_{23})	-.127	.082	.121	1.145
Lighting Condition (X_8)	Daylight (x_{24})		*(Control Variable)*		
	DarkLighting (x_{25})*	.110	.057	.055	1.268
Road Condition (X_9)	DryRoad (x_{26})		*(Control Variable)*		
	WetRoad (x_{27})	-.036	.092	.699	1.923
	SnowIceRoad (x_{28})	.119	.143	.406	1.425
Road Alignment (X_{10})	StraightRoad (x_{29})		*(Control Variable)*		
	CurvedRoad (x_{30})	-.027	.083	.742	2.357
Movement Prior to Crash (X_{11})	MaintainingSpeed (x_{31})		*(Control Variable)*		
	Accelerating (x_{32})	.294	.219	.179	1.083
	Decelerating (x_{33})	-.093	.539	.863	1.109
	Stopped (x_{34})	.007	.383	.986	1.122
	Turning (x_{35})	-.070	.077	.368	2.394
	ChangingLanes (x_{36})	-.311	.540	.566	1.117
Evidence of Speeding (X_{12})	Speeding (x_{37})	.020	.076	.792	1.261
Evidence of Seatbelt Use (X_{13})	SeatbeltUse (x_{38})	.013	.060	.830	1.100

Table 2 – Results of initial model when the full 38 variables are included.

3.3. Significant Linear Regression Variables:

Given that the efficacy of the linear model has now been established, the significant variables identified through stepwise regression are isolated and added to a reduced model (Table 3). The adjusted-R^2 measure of this reformed model is 20.6%, an encouraging result given the limited sample size, the underlying factors in RTAs that are difficult to capture through raw data, and the subjectivity that underpins compensation payments. The variables referenced in Table 3 have previously been identified as having meaningful explanatory power in traffic safety analyses. In a study measuring Barrier Equivalent Speed, acting as a proxy for the impact energy sustained in a crash, Mackay (1968) found conclusive evidence of a strong positive correlation with the severity of injuries. The availability of airbags in a vehicle has been shown to reduce the severity of injuries to a certain point (Savolainen *et al.* 2011). An increasing number of airbags available in the vehicle may also be indicative of a more risk-averse driver or a vehicle with a higher safety rating, further mitigating the risk of sustaining a serious injury or worse. The complexity involved in accounting for these endogenous effects are beyond the scope of this analysis. Eis *et al.* (2005), Kockelman and Kweon (2002), and Duncan *et al.* (1998) supplied evidence of less severe injuries stemming from rear-end collisions, collisions involving light trucks, and incidents occurring in dark conditions respectively. There have been few studies examining the relationship between the number and severity of injuries sustained in incidents where the striking vehicle was turning prior to impact. Overall, the optimal prediction for a compensation pay out using this model is stated as:

$$\ln(Cost) = 9.608 + 0.546\ln(BES) - .035(AirbagsAvailable) - .117(Turning) - .142(VansLightTrucksBodyType) + 0.93(DarkLighting) - .265(RearEnd)$$

The results shows that €16,890.46 is the base-level expected pay out stemming from an accident, i.e. the constant when all of the factors in the equation are set to zero. This value is derived after correcting for the downward bias associated with log-log linear models by including the standard error of the model σ in the estimate, i.e. e^{β_0} becomes $e^{\beta_0+\frac{\sigma^2}{2}}$ or $e^{9.608+\frac{0.503^2}{2}}$.

The cost variable shares a highly significant elastic relationship with the speed-measure used in this analysis – for every percentage increase in the BES, compensatory costs are expected to rise by 0.546%. A strong, positive relationship between these variables is unsurprising given the long-standing evidence that high speeds result in destructive incidents. There is also substantial evidence of a link between the cost function and the number of airbags available in the vehicle. For every unit increase in the number of airbags available in the vehicle, the expected compensation payout decreases by 3.5%. The previously discussed notion of unobserved heterogeneity must be considered when interpreting this result however, as driving propensity and the safety rating of the vehicle are not accounted for when conducting this analysis. This analysis also uncovered evidence of a negative correlation between expected compensation costs and the trajectory of the vehicle – it is found that a vehicle that enters into an incident whilst turning have an almost 12% lower compensation cost relative to accidents in which a straight trajectory is maintained.

This analysis also finds that van-based light trucks have a 14% lower expected compensation pay out than all other vehicles, while accidents occurring under darkness increase the expected injury compensation by 9.3%. Finally, this model finds that a vehicle involved in a rear-end collision has a decreased expected compensation pay out of over 26.5% relative to all other collisions. A possible reason for this phenomenon is the fact that rear-end collisions make up the majority of vehicular incidents, and protective safety measures provided by car manufacturers significantly reduce the severity of rear-end collisions relative to side- and front-impacts.

Model Summary: $R^2: 0.217,\ Adjusted - R^2: 0.206,\ Std. Error: 0.503$ **Significance of Fit:** $F: 19.364, Sig.: < 0.001$ **Goodness of Fit:** $AIC: 630.80, BIC: 663.21$		Unstandardized Coefficients			Collinearity Statistics
		B	Std. Error	Sig.	VIF
Constant (Y)	LnCost	9.608	.216	.000	
Barrier Equivalent Speed (Log-Transform) (X_1)	LnBES (x_1)***	.546	.057	.000	1.009
Number of Airbags Available (X_3)	AirbagsAvailable (x_3)**	-.035	.014	.012	1.041
Movement Prior to Crash (X_{11})	Turning (x_{35})**	-.117	.050	.018	1.011
Vehicle Body Type (X_7)	VansLightTrucksBodyType (x_{23})*	-.142	.077	.064	1.039
Lighting Condition (X_8)	DarkLighting (x_{25})*	.093	.051	.068	1.031
Crash Category (X_5)	RearEnd (x_8)*	-.265	.149	.075	1.020

Table 3 – The significant variables involved in the fit of the final linear model.

3.4. Case Study:

In order to demonstrate the effectiveness of the model in predicting the cost associated with a severe incident, a sample case study was chosen and visualised in Figure 4. This two-vehicle incident involved a 25-year-old female travelling in a 7-year-old 4-door sedan. The point of contact was at 0° indicating a straightforward collision from the rear; however, the collision occurred while completing a turn around a curve with an equivalent speed of 50km/h. There were 2 airbags available in the vehicle and both were deployed. The woman sustained a closed femoral shaft fracture, an open tibia fracture, and an open fibula fracture, all to the right leg. Shoulder and knee bruising were also noted. As detailed in Section 2, the Book of Quantum recommends that the compensation range be based on the most severe injury. The most severe injury sustained in this incident was the femur fracture with AIS level 3, therefore categorized as a moderately severe injury. The range specified for a moderately severe femur fracture with multiple other ruptures is €60,600 – €97,700. Using the methodology put forth in Section 2, the cost of injuries is calculated to be €65,179.88, while the cost predicted by the optimal model following downward bias correction is €78,325.35. Both of these costs are in accordance with the range of expected payouts dictated by the Book of Quantum, and offers evidence of the model's potential as part of an on-board, economic-risk based trajectory system.

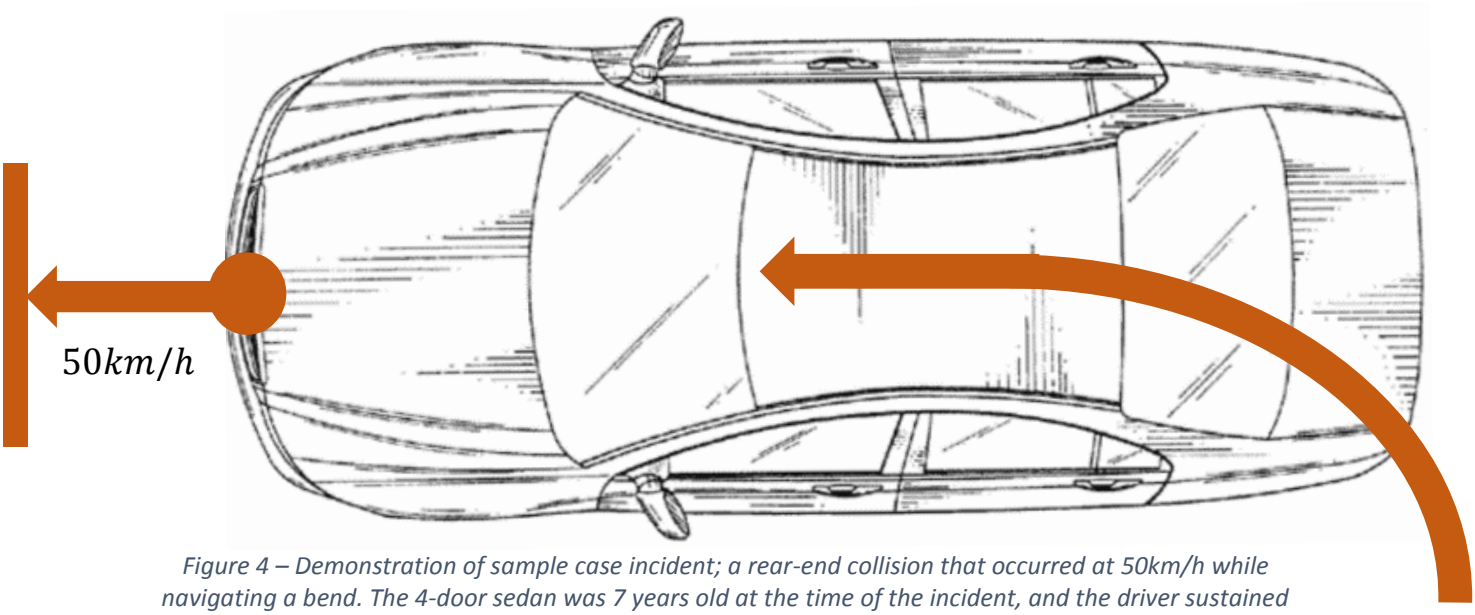


Figure 4 – Demonstration of sample case incident; a rear-end collision that occurred at 50km/h while navigating a bend. The 4-door sedan was 7 years old at the time of the incident, and the driver sustained multiple leg fractures of moderate severity (Case ID: 385003752, CIREN ID: 857085889).

3.5. Implications of Results:

There has been a large surge in traffic safety innovation in recent years with the target of reducing or eliminating collision factors that occur most frequently and are easily avoidable. These innovations will largely affect the variables found to be significant in this research study. For example, particular emphasis has been placed on safety initiatives promoting the purchase of vehicles with higher safety ratings, while reducing speed while driving. The results found as part of this study highlight that an increase in speed dramatically escalates the severity of an incident in terms of expected cost, while there is also a negative relationship found between the availability of airbags and the ensuing cost of injuries.

As part of these safety initiatives, the vast majority of leading automobile manufacturers have signed a commitment with the U.S. Department of Transportation to ensure that over 95% of vehicles intended for sale in the United States will have Automatic Emergency Braking (AEB) as standard by 2022. Legislations have also been introduced in the European Union to ensure that all heavy-duty vehicles have AEB and Lane Departure Warning (LDW) systems from 2015, while all vehicles sold

within the EU have Electronic Stability Control (ESC) from 2014. Furthermore, there are an increasing number of requirements in order to attain a five-star vehicle safety rating within the EU, which in turn will encourage further automotive manufacturers to introduce Advanced Driver Assistance Systems to their vehicles. Of the six variables identified as significant in Table 3, four will undergo significant transformations in the near- to mid-term horizon, including speed, rear-end collisions, traction control (leading to decreased risks faced while turning), and hazards faced in adverse environmental conditions.

A low-speed field study conducted by Cicchino (2016) found appreciable decreases in rear-end crash rates in vehicles which were Forward Collision Warning- (FCW) and AEB-enabled, accomplished by attenuating the speed of the vehicle prior to impact or stopping completely. Results indicated that there were also significant reductions in the number of strike rates with injuries. It remains to be seen if these results can be replicated when the striking vehicle is travelling at a high speed.

ESC systems, meanwhile, are designed to prevent the loss of traction and are projected to prevent 3,000 fatalities and 50,000 injuries per year within the EU, limiting the inherent danger of driving in adverse environments and supplying extra control when navigating turns. This conclusion also applies for the 'turning' variable, as vehicles containing ESC systems have already been proven as a relevant factor in reducing the severity of injuries sustained in incidents while turning (Ferguson 2007).

A combination of FCW, AEB and ESC technologies (Figure 5), as well as many other advanced safety systems not referenced in this study, will result in a notable drop in RTAs. Furthermore, it is expected that a gradual progression toward semi-autonomous and autonomous driving will largely reduce the frequency of vehicle accident rates, by removing the element of human error, and force a departure from existing actuarial models (Sheehan *et al.* 2017). These factors will have profound effects on the automotive industry as a whole. As well as their influence on the nature and severity of injuries incurred, adaptions must be made by insurance underwriters to account for the risks that

will be mitigated. Claims reserving is a classic issue prevalent throughout insurance mathematics, and one such adaption that insurance underwriters must consider is an overhaul of the 'driver risk' profile system that is currently used to estimate the cost of their risk. A distributional model currently commonly accepted as accurately representing expected compensation payouts is the Tweedie Compound Poisson model (Smyth and Jørgensen 2002). This model is underlined by a Gamma distribution that typically has a single peak followed by a smooth decline. However, when compounded with Poisson-derived event rates, random peaks are scattered throughout the distribution. The Tweedie Compound Poisson model shares similar attributes to the distribution found in Figure 2. A shift in the frequency of low-speed impacts may require insurance mathematics models such as the Tweedie Compound Poisson model to be re-configured.

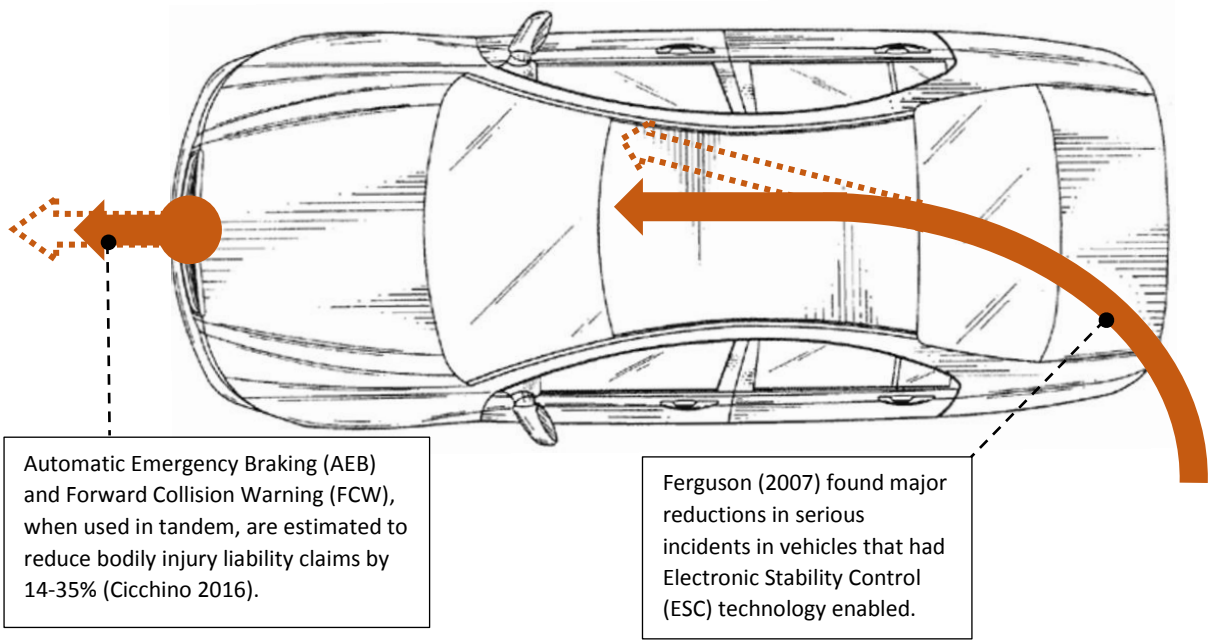


Figure 5 - Advanced Driver Assistance System (ADAS) technologies will largely mitigate the most common and easily avoidable incidents - such as the incident sustained in the case study.

4. Summary & Conclusion:

The most commonly studied method of severity parameterisation in traffic safety analyses is the collection of medical records describing the most severe injury sustained as part of an incident. This

study serves as a departure from previous research by relating injuries suffered in a road traffic accident to expected trauma compensation payouts and deriving a quantitative cost function. Six variables in particular are highlighted as being significantly influential in determining the expected compensation awarded and, by proxy, the level of injuries sustained. The vehicle's relative speed at the time of impact and dark conditions both increase the expected costs. A decrease in expected compensation stems from incidents involving a light truck, the number of airbags available in the vehicle, incidents sustained while turning, and rear-end incidents. As an increasing number of vehicles become available equipped with intelligent Adaptive Driver Assistant functions, all risks faced by the vehicle when navigating the route of maximum utility must be accounted for, including an estimation of the severity of any given incident that may arise. From an econometric perspective, the model proposed in this study offers an avenue into quantifying the expected compensation cost stemming from an incident given a set number of factors. The results that were extracted are encouraging and offer evidence that this system may be a useful tool in plotting a path of least economic risk. However given that only 13 variables were subject to examination, and 6 proved significant, further research is required to expand this model to scenarios that exhibit more sources of variation that could be proven as significant measures of RTA risk.

Furthermore, the large surge in traffic safety initiatives are expected to markedly reduce the frequency and severity of RTAs, with greatest emphasis placed on mitigating the impact of vehicle speed. These innovations will largely affect the variables found to be significant in this research study and may have profound effects on the automotive insurance industry as a whole. The introduction of ADAS technologies such as FCW, AEB and ESC systems may force a departure from existing capital allocation models as many of the leading causes of driving incidents are mitigated, and may cause insurance underwriters to reconsider the 'driver risk' profile system that is currently used to estimate the cost of their risk.

However, the largest challenge facing the estimation of risk severity facing road users remains unobserved heterogeneity – the unmeasurable unobserved factors that prove to be difficult to control for as part of a statistical analysis. Given the unpredictable variations that play a role in each incident, it is exceedingly difficult to accurately predict and account for the expected injury and risk severity outcome. The categorisations of injuries introduces information loss when parameterising severity as discrete outcomes do not provide all minute details. The alternative research methodology proposed in this study offers an avenue to ratify the slight loss of granularity that is incurred when limiting injuries to five (KABCO) or six (AIS) categories. Although the linear model performed well against the ordinal logit and multinomial logit models, an extensive number of complex discrete-outcome models may produce similar results to those found in this study. Further research is required in this area. In any case, the methods introduced in this research, combined with the increasing prevalence of ADAS functions such as collision warning and stability control systems, can be expanded upon to capture further measures of variation. This in turn could potentially limit the influence of the high number of contextual factors that affect driving behaviour, road characteristics and environmental conditions.

Acknowledgement:

The corresponding author would like to thank the anonymous reviewers for their time and effort taken to provide suggestions of improvement for this paper.

References:

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. Journal of safety research 34 (5), 597-603.

An, P.E., Harris, C.J., 1996. An intelligent driver warning system for vehicle collision avoidance. IEEE transactions on systems, man, and cybernetics-part A: Systems and Humans 26 (2), 254-261.

Buddhavarapu, P., Banerjee, A., Prozzi, J.A., 2013. Influence of pavement condition on horizontal curve safety. Accident Analysis & Prevention 52, 9-18.

Cerwick, D.M., Gkritza, K., Shaheed, M.S., Hans, Z., 2014. A comparison of the mixed logit and latent class methods for crash severity analysis. Analytic Methods in Accident Research 3, 11-27.

Cicchino, J.B., 2016. Effectiveness of forward collision warning systems with and without autonomous emergency braking in reducing police-reported crash rates. Arlington, VA: Insurance Institute for Highway Safety.

Damerow, F., Flade, B., Eggert, J., 2016. Extensions for the foresighted driver model: Tactical lane change, overtaking and continuous lateral control. Intelligent Vehicles Symposium (IV), 2016 IEEE, 186-193.

Duncan, C., Khattak, A., Council, F., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transportation Research Record: Journal of the Transportation Research Board (1635), 63-71.

Eggert, J., Damerow, F., Klingelschmitt, S., 2015. The foresighted driver model. Intelligent Vehicles Symposium (IV), 2015 IEEE, 322-329.

Eis, V., Sferco, R., Fay, P.A., 2005. A detailed analysis of the characteristics of european rear impacts. Proceedings: International Technical Conference on the Enhanced Safety of Vehicles 2005, 9p-9p.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention 40 (3), 1033-1054.

Eluru, N., Chakour, V., Chamberlain, M., Miranda-Moreno, L.F., 2013. Modeling vehicle operating speed on urban roads in montreal: A panel mixed ordered probit fractional split model. Accident Analysis & Prevention 59, 125-134.

Eluru, N., Yasmin, S., 2015. A note on generalized ordered outcome models. Analytic methods in accident research 8, 1-6.

Ferguson, S.A., 2007. The effectiveness of electronic stability control in reducing real-world crashes: A literature review. Traffic injury prevention 8 (4), 329-338.

Finkelstein, E., Corso, P.S., Miller, T.R., 2006. The incidence and economic burden of injuries in the united states Oxford University Press, USA.

Jones, I.S., Champion, H.R., 1989. Trauma triage: Vehicle damage as an estimate of injury severity. Journal of Trauma and Acute Care Surgery 29 (5), 646-653.

Kim, J.-K., Ulfarsson, G.F., Shankar, V.N., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. Accident Analysis & Prevention 40 (5), 1695-1702.

Klingelschmitt, S., Eggert, J., 2015. Using context information and probabilistic classification for making extended long-term trajectory predictions. Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on, 705-711.

Kockelman, K.M., Kweon, Y.-J., 2002. Driver injury severity: An application of ordered probit models. Accident Analysis & Prevention 34 (3), 313-321.

Kyriakidis, M., Happee, R., De Winter, J.C., 2015. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation research part F: traffic psychology and behaviour 32, 127-140.

Lemp, J.D., Kockelman, K.M., Unnikrishnan, A., 2011. Analysis of large truck crash severity using heteroskedastic ordered probit models. Accident Analysis & Prevention 43 (1), 370-380.
Mackay, G.M., 1968. Injury and collision severity. SAE Technical Paper.
Mannering, F.L., Bhat, C.R., 2014. Analytic methods in accident research: Methodological frontier and future directions. Analytic methods in accident research 1, 1-22.
Mannering, F.L., Shankar, V., Bhat, C.R., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic methods in accident research 11, 1-16.
Mooradian, J., Ivan, J.N., Ravishanker, N., Hu, S., 2013. Analysis of driver and passenger crash injury severity using partial proportional odds models. Accident Analysis & Prevention 58, 53-58.
O'donnell, C., Connor, D., 1996. Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis & Prevention 28 (6), 739-753.
Personal Injuries Assessment Board, 2016. Book of quantum.
Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. Accident Analysis & Prevention 43 (5), 1666-1676.
Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. Journal of safety research 27 (3), 183-194.
Sheehan, B., Murphy, F., Ryan, C., Mullins, M., Liu, H.Y., 2017. Semi-autonomous vehicle motor insurance: A bayesian network risk transfer approach. Transportation Research Part C: Emerging Technologies 82, 124-137.
Smith, J.A., Siegel, J.H., Siddiqi, S.Q., 2005. Spine and spinal cord injury in motor vehicle crashes: A function of change in velocity and energy dissipation on impact with respect to the direction of crash. Journal of Trauma and Acute Care Surgery 59 (1), 117-131.
Smyth, G.K., Jørgensen, B., 2002. Fitting tweedie's compound poisson model to insurance claims data: Dispersion modelling. Astin Bulletin 32 (01), 143-157.
Us Dot/Nhtsa, 2016. U.S. Dot/nhtsa - commitments to advancing automatic emergency braking technology (memorandum). In: Administration, N.H.T.S. ed., https://www.regulations.gov/document?D=NHTSA-2015-0101-0005.
Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: The case of airbags and antilock brakes. Journal of Risk and Uncertainty 32 (2), 83-99.
World Health Organisation, 2013. Global status report on road safety 2013: Supporting a decade of action World Health Organization.
Yamamoto, T., Hashiji, J., Shankar, V.N., 2008. Underreporting in traffic accident data, bias in parameters and the structure of injury severity models. Accident Analysis & Prevention 40 (4), 1320-1329.
Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014a. A latent segmentation based generalized ordered logit model to examine factors influencing driver injury severity. Analytic methods in accident research 1, 23-38.
Yasmin, S., Eluru, N., Ukkusuri, S.V., 2014b. Alternative ordered response frameworks for examining pedestrian injury severity in new york city. Journal of Transportation Safety & Security 6 (4), 275-300.
Ye, F., Lord, D., 2011. Investigation of effects of underreporting crash data on three commonly used traffic crash severity models: Multinomial logit, ordered probit, and mixed logit. Transportation Research Record: Journal of the Transportation Research Board (2241), 51-58.

Appendix:

Model Summary (Pseudo-R^2): $Cox-Snell: 0.117, Nagelkerke: 0.132, McFadden: 0.056$ **Significance of Fit (-2LL):** $Final\ (Intercept\ Only): 887.02\ (940.01), Sig.: 0.011$ **Goodness of Fit (Final):** $AIC: 957.02, BIC: 1098.84$ **Test of Parallel Lines (-2LL):** $Null: 887.02, General: 820.27, Sig.: 0.383$		Exp(B) – Odds Ratios	Sig.	95% Confidence Interval	
				Lower Bound	Upper Bound
Thresholds (Reference: Critical Injury)	Moderate Injury	.038	.486	0.000	380.713
	Serious Injury	.893	.981	0.000	9046.206
	Severe Injury	5.073	.730	0.001	51349.733
Barrier Equivalent Speed (Log-Transform) (X_1)	LnBES (x_1)**	2.017	.012	1.170	3.476
Age of Vehicle (X_2)	CurrentVehicleAge (x_2)**	1.079	.041	1.003	1.160
Number of Airbags Available (X_3)	AirbagsAvailable (x_3)**	.860	.034	.748	.988
Evidence of alcohol consumption (X_4)	DrinkDriver (x_4)	1.140	.710	.571	2.277
Crash Category (X_5)	RightRoadsideDeparture (x_5)	.768	.448	.389	1.518
	LeftRoadsideDeparture (x_6)	.564	.116	0.277	1.151
	SingleDriverForwardImpact (x_7)	.617	.581	0.111	3.431
	RearEnd (x_8)	1.676	.447	.443	6.333
	OppositeDirectionImpact (x_9)	.449	.232	0.121	1.668
	Head-On (x_{10})		*(Control Variable)*		
	Angle/Sideswipe (x_{11})	1.568	.499	.425	5.779
	TurnAcrossPath (x_{12})	1.292	.526	.586	2.849
	TurnIntoPath (x_{13})*	.429	.053	0.182	1.011
	IntersectingPaths (x_{14})	.607	.194	0.286	1.289
	BackingUp (x_{15})	1.242	.672	.456	3.379
Weather Condition (X_6)	FogSmoke (x_{16})	1.012	.993	0.080	12.805
	Rain (x_{17})	.676	.417	0.263	1.740
	Snow (x_{18})	1.026	.987	0.045	23.612
	Clear (x_{19})		*(Control Variable)*		
	Cloudy (x_{20})*	.565	.086	0.295	1.084
Vehicle Body Type (X_7)	AutomobileBodyType (x_{21})		*(Control Variable)*		
	SUVBodyType (x_{22})	1.570	.111	.902	2.731
	VansLightTrucksBodyType (x_{23})*	1.750	.080	.935	3.275
Lighting Condition (X_8)	Daylight (x_{24})		*(Control Variable)*		
	DarkLighting (x_{25})	.961	.856	.622	1.483
Road Condition (X_9)	DryRoad (x_{26})		*(Control Variable)*		
	WetRoad (x_{27})	1.794	.106	.883	3.646
	SnowIceRoad (x_{28})	.732	.562	0.255	2.103
Road Alignment (X_{10})	StraightRoad (x_{29})		*(Control Variable)*		
	CurvedRoad (x_{30})*	1.700	.099	.905	3.193
Movement Prior to Crash (X_{11})	MaintainingSpeed (x_{31})		*(Control Variable)*		
	Accelerating (x_{32})	.746	.728	0.144	3.882
	Decelerating (x_{33})	.147	.329	0.003	6.897
	Stopped (x_{34})	1.674	.724	0.096	29.171
	Turning (x_{35})	.617	.109	0.342	1.113
	ChangingLanes (x_{36})	3.147	.597	0.045	219.536
Evidence of Speeding (X_{12})	Speeding (x_{37})	1.159	.526	.734	1.831
Evidence of Seatbelt Use (X_{13})	SeatbeltUse (x_{38})	.943	.840	.534	1.667

Table A.1 – Results of ordinal logit regression model when all 38 variables are included.

Model Summary (Pseudo-R^2): $Cox-Snell: 0.309, Nagelkerke: 0.346, McFadden: 0.167$ **Significance of Fit (-2LL):** $Final\ (Intercept\ Only): 783.22\ (940.01), Sig.: < 0.001$ **Goodness of Fit (Final):** $AIC: 981.22, BIC: 1382.376$			Exp(B) – Odds Ratios	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Thresholds (Reference: Critical Injury 5)	Moderate Injury	2	.	.996	.	.
	Serious Injury	3	.	.941	.	.
	Severe Injury	4	.	.956	.	.
Barrier Equivalent Speed (Log-Transform) (X_1)	LnBES (x_1)**	2	.237	.048**	.057	.987
		3	.267	.030**	.081	.877
		4	.492	.273	.139	1.749
Age of Vehicle (X_2)	CurrentVehicleAge (x_2)***	2	.741	.004***	.603	.910
		3	.964	.623	.835	1.114
		4	.926	.335	.793	1.082
Number of Airbags Available (X_3)	AirbagsAvailable (x_3)**	2	1.531	.029**	1.045	2.244
		3	1.091	.563	.812	1.465
		4	1.080	.631	.790	1.475
Evidence of alcohol consumption (X_4)	DrinkDriver (x_4)	2	.658	.655	.105	4.134
		3	.814	.766	.210	3.151
		4	.711	.640	.170	2.977
Crash Category (X_5)	RightRoadsideDeparture (x_5)	2	1.446	.632	.320	6.530
		3	1.763	.342	.547	5.684
		4	1.236	.751	.333	4.586
	LeftRoadsideDeparture (x_6)	2	3.668	.151	.622	21.640
		3	1.585	.475	.448	5.611
		4	1.162	.833	.289	4.679
	SingleDriverForwardImpact (x_7)	2	1.676	.998		
		3	.	.962	.	.
		4	.	.960		
	RearEnd (x_8)	2	.540	.679	.029	10.010
		3	1.129	.919	.109	11.715
		4	2.389	.570	.119	48.150
	OppositeDirectionImpact (x_9)	2	.878	.937	.034	22.391
		3	1.913	.609	.159	22.989
		4	.434	.504	.037	5.035
	Head-On (x_{10})		*(Control Variable)*			
	Angle/Sideswipe (x_{11})	2	.899	.941	.055	14.716
		3	1.091	.942	.107	11.136
		4	1.945	.660	.101	37.617
	TurnAcrossPath (x_{12})	2	.	.889		
		3	.	.880	.	.
		4	.	.882		
	TurnIntoPath (x_{13})	2	3.042	.354	.289	32.048
		3	1.260	.802	.207	7.671
		4	.397	.338	.060	2.628
	IntersectingPaths (x_{14})	2	2.159	.420	.333	14.020
		3	1.149	.851	.268	4.921
		4	.620	.552	.129	2.991
	BackingUp (x_{15})	2	5.281	.273	.270	103.199
		3	.803	.806	.139	4.638
		4	4.333	.276	.309	60.706
Weather Condition (X_6)	FogSmoke (x_{16})	2		.981		
		3	.	.981	.	.
		4		.978		
	Rain (x_{17})	2	1.052	.973	.054	20.504
		3	.573	.665	.046	7.145
		4	.327	.411	.023	4.712

	Snow (x_{18})	2	25272.25	.982	.	.
		3	.108	.996		
		4	15664.45	.983		
	Clear (x_{19})	*(Control Variable)*				
	Cloudy (x_{20})*	2	3.948	.086*	.824	18.913
		3	2.116	.204	.665	6.727
		4	1.264	.709	.369	4.333
Vehicle Body Type (X_7)	AutomobileBodyType (x_{21})	*(Control Variable)*				
	SUVBodyType (x_{22})**	2	.257	.041**	.070	.947
		3	1.192	.749	.408	3.479
		4	.895	.850	.285	2.817
	VansLightTrucksBodyType (x_{23})	2	.329	.186	.063	1.713
		3	.736	.630	.212	2.559
		4	1.613	.510	.388	6.702
Lighting Condition (X_8)	Daylight (x_{24})	*(Control Variable)*				
	DarkLighting (x_{25})	2	1.308	.642	.421	4.066
		3	.791	.600	.329	1.901
		4	.936	.891	.364	2.408
Road Condition (X_9)	DryRoad (x_{26})	*(Control Variable)*				
	WetRoad (x_{27})	2	.278	.213	.037	2.080
		3	.462	.341	.094	2.267
		4	.914	.920	.157	5.304
	SnowIceRoad (x_{28})	2	1.602	.722	.119	21.510
		3	2.075	.374	.416	10.366
		4	2.169	.422	.327	14.371
Road Alignment (X_{10})	StraightRoad (x_{29})	*(Control Variable)*				
	CurvedRoad (x_{30})	2	.337	.210	.061	1.846
		3	.471	.296	.115	1.933
		4	.733	.684	.165	3.262
Movement Prior to Crash (X_{11})	MaintainingSpeed (x_{31})	*(Control Variable)*				
	Accelerating (x_{32})	2	1.343	.999	.	.
		3	.	.960		
		4	.	.959		
	Decelerating (x_{33})	2	.353	.999	.	.
		3	1.276	.999		
		4	.	.999		
	Stopped (x_{34})	2	2.230	.999	.	.
		3	.	.980		
		4	1.270	.999		
	Turning (x_{35})	2	3.544	.134	.678	18.532
		3	1.438	.604	.364	5.677
		4	1.158	.842	.275	4.877
	ChangingLanes (x_{36})	2	1.320	.999	.	.
		3	.	.986		
		4	1.988	.999		
Evidence of Speeding (X_{12})	Speeding (x_{37})*	2	1.067	.923	.287	3.965
		3	1.558	.415	.537	4.520
		4	2.650	.086*	.872	8.048
Evidence of Seatbelt Use (X_{13})	SeatbeltUse (x_{38})	2	.660	.534	.179	2.438
		3	1.389	.520	.510	3.782
		4	1.078	.890	.374	3.108

Table A.2 – Results of multinomial logit regression model when all 38 variables are included.