"Number and severity of BI victims, assuming dependence between vehicles involved in the crash"

Miguel Santolino and Mercedes Ayuso

UBIREA | Institut de Recerca en Economia Aplicada Regional i Pública

Abstract

The number of victims in vehicles in Spanish motor crashes is analyzed by bodily injury (BI) severity level. Generalized linear mixed models (GLMMs) are applied to model the number of non-serious victims, serious victims and fatalities. Dependence between vehicles involved in the same crash is captured including random effects. After comparing between error distributions, the binomial GLMM is selected. The effect of the driver, vehicle and crash characteristics on the number of BI victims by severity level is analyzed, paying special attention to the influence of the age of the driver and the age of the vehicle. We found a nonlinear relationship between driver's age and severity, with young and older drivers being the riskiest groups. On the other hand, the expected severity of the crash linearly increased with the vehicle age until the vehicle was 18 years old and then remained constant at the highest severity level from that age. These results are relevant in countries such as Spain with increasing longevity of drivers and aging of the car fleet.

JEL classification: J11, J14, I10, C5.

Keywords: Motor crashes, Severity, Dependence, Random effects, Driver age, Vehicle age.

Miguel Santolino: Department of Econometrics, Riskcenter-IREA, University of Barcelona
Av. Diagonal 690, 08034 Barcelona. Tel.: +34 93 402 0484. Email: msantolino@ub.edu

Mercedes Ayuso: Department of Econometrics, Riskcenter-IREA, University of Barcelona
Av. Diagonal 690, 08034 Barcelona. Email: mayuso@ub.edu

Acknowledgements

We are grateful to the Spanish Ministry of Science and Innovation for funding received under grant PID2019-105986GB-C21. We are also grateful to the Dirección General de Tráfico for access to their database. We thank Rodrigo Sánchez for his comments and his help in the design of the database.

1. Introduction

This paper analyzes the number of victims in the vehicle in a motor crash by bodily injury (BI) severity level. We investigate those factors related to the vehicle, the driver and the crash that have a significant impact on the expected numbers of occupants with non-serious (slight) injuries, with serious injuries and who die as a result of the crash. We use the official police dataset of motor crashes involving victims on Spanish roads in the year 2016. We apply generalized linear models with random effects, called generalized linear mixed models (GLMMs) to take into account dependence between vehicles involved in the same crash. The inclusion of random effects into fixed linear models allows the analysis of multilevel data when data have more than one source of random variability. As indicated by Mannering et al. (2016), multivariate issues may arise with vehicle crashes that involve multiple occupant injuries from the same accident. In such cases, unobserved factors influencing the injury levels would be correlated, such as the structural characteristics of the vehicles involved, among others (Mannering et al., 2016; Abay et al., 2013; Eluru et al., 2010).

Identifying factors affecting safety and understanding the different impacts of vehicle features will help to develop new safety features and improved transportation safety programs. A vast number of studies have analyzed risk factors that affect the severity of bodily injuries in road traffic crashes. Some studies focus on analyzing the type of vehicle involved and the resulting damage (Wang and Kockelman, 2005; Quin et al., 2013; George at al., 2017, among others). For instance, two-wheeled motor vehicles present a greater risk of serious injury or fatality (Quddus, 2002; Donate-López et al., 2010; Jackson and Mello, 2013; Schneider et al., 2012), whereas heavier vehicles are more protective of their passengers but cause more damage to the other vehicles involved (Fredette et al., 2008). Other researchers have also examined differences between types of crashes and their impact on the severity of victim injury. The injury severity increases when the accident involves frontal impact rather than rear impact (Abu-Zidan and Eid, 2015). In rollover and drop collisions, passengers are more exposed to the likelihood of serious head and cervical spine injuries (Frunt et al., 2012; Ivarsson et al., 2015). Also, analysis shows that driving under conditions of dark or non-optimal road surfaces plays an important role in crash severity (Sullivan and Flannagan, 2002; Wanvik, 2009; Uddin and Huynh, 2017). But as Eluru et al. (2010) indicate, there is strong evidence of the presence of correlated unobserved factors that affect the injury severity levels among vehicle occupants.

We apply regression models with random effects which are a particular case of random parameter models (see, for a discussion, Mannering et al., 2016). A number of previous research studies have used random-parameter models to account for heterogeneity from unobserved factors related to road geometrics, vehicle types and spatial areas (see Washington et al, 2020, Chap.18). Anastasopoulos and Mannering (2009) show that the marginal effects generated by the standard (fixed effect) negative binomial model and the random effect negative binomial model can be quite different. They suggest that ignoring the possibility of random parameters when estimating count data models can result in changes to the magnitude of the effect of factors affecting crash frequency. A similar conclusion is noted in Anastasopoulos and Mannering (2011), where the authors show that random parameter models using less detailed crash-specific data can still provide a reasonable level of accuracy. Osman et al. (2018) indicate that injury severity conditional on crash occurrence can depend on numerous factors, none of which are observed in crash databases. In this sense they indicate that the unobserved heterogeneity derived from these unobserved factors can moderate the influence of other observed covariates in the model, leading to variation in the parameter effects across different observations. The authors properly refer to Mannering et al. (2016) to analyze alternate modeling methods for handling the problem, random parameter methods being the most prominent. Also, Hosseinpour et al. (2018) estimate crash counts for four different multi-vehicle collision types, with dependencies between collision types and spatial correlation between adjacent sites.

In the analysis of factors that affect the expected number of injured occupants by severity level, including random effects, we pay special attention to the age of the driver and the age of the vehicle. We estimate non-parametrically the effect of these regressors to investigate whether these factors are linearly related with the dependent variable in the GLMM framework or whether another relational form is more appropriate. A number of studies suggest that the effect of driver age is not linear with crash severity, as young and old drivers represent the riskiest groups (Alam and Spainhour, 2008; Chin and Zhou, 2018; Regev et al., 2018). Recent research efforts have focused on older drivers, given the increasing longevity of the population, and the effects on road traffic crash injury rates

(Ayuso et al., 2020; Johannsen and Müller, 2013; Clarke et al., 2010). Another relevant risk factor investigated for the shape of its relationship with crash severity is vehicle age. The incorporation of safety technological improvements in newer vehicles has been accelerated in recent decades. The later generations of cars are associated with lower probabilities of injury and fatality in car crashes (Rich et al., 2013; Anderson and Searson, 2015; Høye, 2019; Ayuso et al., 2019; NHTSA, 2013; Blows et al., 2003).

The structure of this paper is as follows. Section 2 defines the GLMM used to model the number of injured victims in a motor crash according to different BI severity levels. Section 3 describes the dataset, and the key descriptive statistics are presented. Results related to the model selection and the binomial generalized linear mixed model estimated are reported in Section 4, where a detailed analysis of the impact of driver age and vehicle age on BI severity level is carried out. Discussion is provided in Section 5, and Section 6 concludes.

2. Generalized Linear Mixed Models

Our analysis focuses on the relationship between a set of risk factors and the number of victims in a vehicle involved in a crash, according to injury severity. We proceed to model the number of non-seriously (slightly) injured occupants in the vehicle, the number of seriously injured occupants and the number of fatally injured occupants. Note that injuries are considered non-serious if the victim suffered only minor personal injuries and did not require hospitalization or was hospitalized for less than 24 hrs. Injuries are defined as serious if they required hospitalization for more than 24 hrs. Finally, a victim is defined as fatally injured if death occurred within 30 days following and as a result of the accident. The unit of observation in the analysis is the vehicle involved in the crash.

We deal with three discrete variables: the number of non-seriously injured occupants in the vehicle y^{ns}; the number of seriously injured occupants, y^{s}, and the number of fatally injured occupants, y^{f}. Generalized linear models (GLMs) for discrete variables assume that observations are independent. However, when multiple vehicles are involved in a crash, the number of injured victims of the same severity level in each vehicle will be presumed to correlate. When data present correlated clusters, GLMMs are a more appropriate specification. GLMMs are an extension of GLMs, incorporating random effects for the analysis of multilevel data.

The dependent variable y^j reflecting the number of injured victims in the vehicle according to the severity level *j*=(ns,s,f) follows an exponential family distribution defined as $f\left(y^j|\theta,\phi\right) = \exp\left(\frac{y^j\theta-b(\theta)}{a(\phi)} + c\left(y^j,\phi\right)\right)$, where θ is the canonic parameter and ϕ the dispersion parameter, and $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions. In the case of a Poisson distribution, $y^j \sim P(\mu)$, the canonic parameter is $\theta = \ln(\mu)$. In the case of a binomial distribution, $y^j \sim B(n,\pi)$, the canonic parameter is $\theta = \ln\left(\frac{\pi}{1-\pi}\right)$. Finally, in the case of a negative binomial distribution $y^j \sim NB(k,\mu)$, the canonic parameter is given by $\theta = \ln\left(\frac{k\mu}{1+k\mu}\right)$. The superscript of y is removed to simplify notation.

The number of injured victims is a function of vehicle occupancy. The set of vehicles included in the analysis has different passenger capacities and, even if they have the same capacity, the number of occupants at the time of the crash could differ. In all model specifications the number of occupants of the vehicle at the time of the crash is included as an offset term (exposure to risk). Including this offset, the numbers of injured occupants per severity level are modeled in relative terms, i.e. the proportion of injured victims in relation to the total occupancy of the vehicle. So, vehicles involved in the motor crash with different occupancy can now be compared. When the dependent variable is expressed in relative terms, the binomial regression specification is equivalent to the logit regression model (Milton et al., 2008). In terms of goodness-of-fit, likelihood-based measures can be used to compare the binomial regression with the Poisson and negative binomial regression with dependent variable measured in relative terms.[2]

The GLM relates the conditional mean of the distribution μ and the linear regression through the link function g as follows: $g(\mu_i) = \eta_i = x_i^{\mathrm{T}}\beta$ for the *i*-th vehicle, *i*=1,…, *I*, where η_i is the linear predictor, β is the vector of the regression coefficients and x_i is the vector of regressors. When the canonical link function is selected, then $\theta_i = \eta_i$. Now, we

[2] A more appropriate specification would be a Poisson or negative binomial regression with right censoring in which the censoring value is the number of occupants. Mean parametrization of these models is not available. As a result, the interpretation of coefficient estimates is more complex. For that reason, right censored models with discrete dependent variables are not included.

introduce a Q-dimension vector of cluster-specific parameters $\Theta_n = (\Theta_{n1}, \ldots, \Theta_{nQ})$ and a vector z_{ni} of predictors corresponding to the random effects, for n=1,…,N. In our case, n indicates the crash and only one cluster-specific parameter is considered, so Θ_n and z_{ni} are scalars. In the GLMM with a cluster-specific variable, the conditional mean μ_{ni} is regressed on the predictors as follows: $g(\mu_{ni}) = x_{ni}^{\mathrm{T}}\beta + z_{ni}\Theta_n$. The constant term of the linear predictor is no longer the same for all observations but varies for each group of vehicles involved in the same crash. Thus, unobserved individual-specific heterogeneity associated with the crash in which the vehicle was involved is introduced into the regression modeling (for a review of statistical methods to deal with unobserved heterogeneity in motor data, see Mannering et al., 2016).

Assuming that the random effect is normally distributed with a mean of zero and dispersion matrix depending on unknown variance components, the marginal probability of response $y_n = (y_{n1}, \ldots, y_{nI})$ is given by:

$$p(y_n|\beta) = \int_{-\infty}^{+\infty} \prod_{i=1}^{I} p(y_{ni}|\beta, \Theta_n)\Phi(\Theta_n|0, \sigma^2)d\Theta_n$$
$$= \int_{-\infty}^{+\infty} p(y_n|\beta, \Theta_n)\Phi(\Theta_n|0, \sigma^2)d\Theta_n,$$

where $\Phi(\Theta_n|0, \sigma^2)$ is the normal distribution with a mean equal to zero and variance σ^2. To estimate parameters β and σ^2, we have to maximize the likelihood function, $L(\beta, \sigma^2|y_1, \ldots, y_n) = \prod_{n=1}^{N} L_n(\beta, \sigma^2|y_n) = \prod_{n=1}^{N} p(y_n|\beta)$. There are several methods for maximizing the marginal likelihood. These methods are based on approximation or simulation to obtain an analytic solution. We focus on the Gauss-Hermite quadrature (McCulloch and Searle, 2001; Naylor and Smith, 1982). This non-stochastic numerical approximation is useful when the random effects are assumed to be normally distributed (Golub and Welsh, 1969).

3. Data

The dataset of motor crashes involving victims was provided by the Spanish Traffic Authority (DGT). It contains information recorded by traffic agents who monitored the evolution of victims over the thirty days following an accident. The complete database contains information for 100,494 police-reported motor vehicle crashes with victims from the period January 2016 to December 2016. There are 179,295 vehicles involved in 102,362 crashes, in which 73,611 vehicles did not present any victims and 105,684 vehicles had at least one victim. Only those vehicles with full records according to our research are selected. We analyze 96,472 vehicles involved in 59,040 crashes. Of these, 46.67% of crashes are associated with one vehicle, and 45.88% of them are associated with two vehicles. The remaining 7.45% of crashes are associated with more than two vehicles. In 42.27% of the vehicles, no occupants were injured as a result of the crash, and in 57.73%, at least one occupant was injured.

Table 1 shows the variables used in the analysis. The dataset contains information on the number of victims in each vehicle by BI severity level, differentiating between non-injury, non-serious or slight injury, serious injury and fatalities. Information related to the driver includes age and gender. Information related to the vehicle involved in the crash includes type, age, and number of occupants (including the driver). Finally, the dataset contains other information related to the accident such as type of crash, type and conditions of the road, and visibility conditions.

Table 1. Description of variables

Name		Categories	Description	Mean*
Victims (dependent variable)				
	Non-injury		Number of non-injury victims in the vehicle	0.65
	Slight		Number of non-serious victims in the vehicle	0.69
	Serious		Number of serious victims in the vehicle	0.06
	Fatalities		Number of fatalities in the vehicle	0.01
Vehicle				
	Vehicle age		Age of the vehicle involved in the crash	10.35
	Vehicle	Car	Cars (category of reference)	69.83
		Van	Vans and minibuses	7.03
		Motorcycle	Motorcycles	12.00
		Moped	Bicycles, Mopeds and ATVs	5.05
		Heavy vehicle	Trucks, tractors and other heavy vehicles	6.09
	Occupants		Number of occupants in the vehicle (including the driver)	1.41
Driver				
	Driver age		Age of the driver involved in the crash (divided by 10)	4.14
	Gender	Female	Driver is female (category of reference)	28.51
		Male	Driver is male	71.49
Crash				

Illumination	Visibility	Driving with visibility (category of reference)	88.53
	No visibility	Driving without appropriate visibility	11.47
Road	Local	City streets and township roads (category of reference)	47.67
	Principal	Highways, freeways and other principal arterials	17.18
	Minor	Minor arterials and collectors	27.98
	Other	Subsidiary roads, unpaved roads, cycling lanes and others	7.17
Condition road	Optimal	Optimal driving conditions of the road surface (category of reference)	84.11
	Non-optimal	Non-optimal driving conditions of the road surface (wet, frozen, muddy)	15.89
Type of crash	Collision	Collision involving another vehicle (category of reference)	61.89
	Pile-up	Multiple vehicle collision	7.82
	Run-over	Collision involving a pedestrian or an animal	8.73
	Rollover	Rollover, drop or collision with an object	14.89
	Other	Other types of crash	6.67

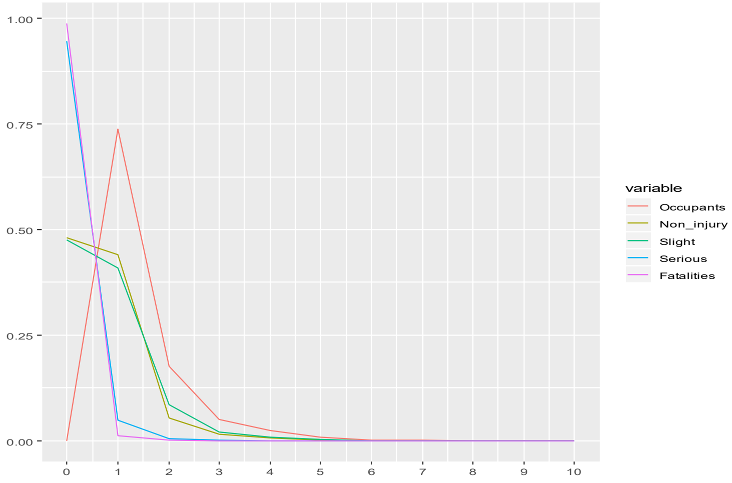
* Relative frequency in % for categorical variables

The average age of the drivers involved in crashes with victims was 41.4 years and the vehicle age was 10.35 years, on average. The number of occupants per vehicle averaged 1.41. Most occupants suffered non-serious injuries (0.69 averaged per vehicle), followed by occupants who did not suffer injuries (0.65 averaged per vehicle), serious injuries (0.06 averaged per vehicle) and fatalities (0.01 averaged per vehicle). Table 2 and Figure 1 show descriptive statistics and empirical density plots for the number of occupants per vehicle (*Occupants*), and the number of victims per vehicle, respectively, for each BI severity level.

Table 2. Descriptive statistics for the number of occupants of the vehicle according to the BI severity level

	Occupants	Non-injury	Slight	Serious	Fatalities
Minimum	1	0	0	0	0
Maximum	61	58	45	21	13
Standard Deviation	1.12	0.92	0.70	0.07	0.01
1st quartile	1	0	0	0	0
3rd quartile	2	1	1	0	0

Figure 1. Empirical densities of the number of victims in the vehicle according to the BI severity level



3.1 Percentage of occupants injured, by BI severity level

A univariate analysis is carried out for classification variables to analyze differences in the distribution of BI severity levels among the occupants of the vehicle. Table 3 shows the percentage of occupants of the vehicle by BI severity level for each category of the classification variables. The Chi-square statistic is significant at 5% for all variables. So, the distribution of the severity level of injury to occupants is statistically different between the categories of the classification variables. In comparison to female drivers, when the driver is male, occupants are more likely to suffer no injuries or severe injuries and fatalities. As expected, most riders of motorcycles and mopeds are injured in motor crashes. On the other hand, 67.34% of occupants of heavy vehicles involved in crashes are not injured.

Driving without good visibility increases the likelihood that occupants will suffer injuries, and crashes on minor arterials and collectors are associated with more severe damage. Roads with non-optimal surface conditions are associated with a higher percentage of slightly injured victims and a lower percentage of serious and fatal victims, as compared to roads with optimal surface conditions. This result may be associated with more careful driving by the driver who perceives that the road surface condition is not good or the weather is bad, or when there are signs indicating this (Mondal et al., 2011). In relation to the type of crash, most run-over crashes do not result in injury to the vehicle occupants

(91.61%). By contrast, in rollovers, drops or collisions with objects, only 16.74% of vehicle occupants escape without injury.

Table 3. Percentage of occupants of the vehicle by BI severity level

		Victim's severity				
		Non-injury	Slight	Serious	Fatal	χ^2
Gender	Female	43.56	53.54	2.55	0.35	
	Male	48.50	44.79	5.47	1.24	1,308.9 (df: 3, p-value< 0.01)
Vehicle	Car	53.83	42.81	2.64	0.72	
	Van	61.35	35.35	2.49	0.81	
	Motorcycle	5.15	76.45	15.96	2.44	
	Moped	5.93	83.66	9.37	1.04	
	Heavy vehicle	67.34	27.4	3.96	1.3	19,372 (df: 12, p-value < 0.01)
Illumination	Visibility	48.45	46.47	4.26	0.82	
	No visibility	36.67	53.57	7.50	2.26	1,273.5 (df: 3, p-value<0.01)
Road	Local	54.88	42.18	2.68	0.26	
	Principal	44.31	50.4	4.23	1.06	
	Minor	36.37	53.18	8.25	2.20	
	Other	43.87	50.78	4.46	0.89	4,974.7 (df: 9, p-value<0.01)
Road condition	Optimal condition	48.49	45.73	4.78	1.00	
	Non-optimal condition	39.70	55.53	3.87	0.90	718.7 (df: 3, p-value<0.01)
Type of crash	Collision	48.92	46.42	3.92	0.74	
	Pile-up	56.22	42.44	1.14	0.20	
	Run-over	91.61	7.43	0.81	0.15	
	Rollover	16.74	70.67	10.21	2.38	
	Other	28.91	61.01	7.94	2.14	17,976 (df: 12, p-value< 0.01)
Total		47.09	47.28	4.63	0.98	

4. Modeling the number of occupants injured by severity level

4.1 Model selection

Three regression models were compared to model, in relative terms, the number of non-seriously injured occupants in the vehicle, the number of seriously injured occupants and the number of fatally injured occupants. In a first stage, GLMs were fitted for three distributions of the exponential family: binomial, Poisson and negative binomial.

Secondly, to capture the dependence among vehicles involved in the same crash a GLMM with a random effect was fitted to the data. The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for the three GLMs and GLMMs are presented in Table 4.

Table 4. Comparison of binomial, Poisson and negative binomial regressions

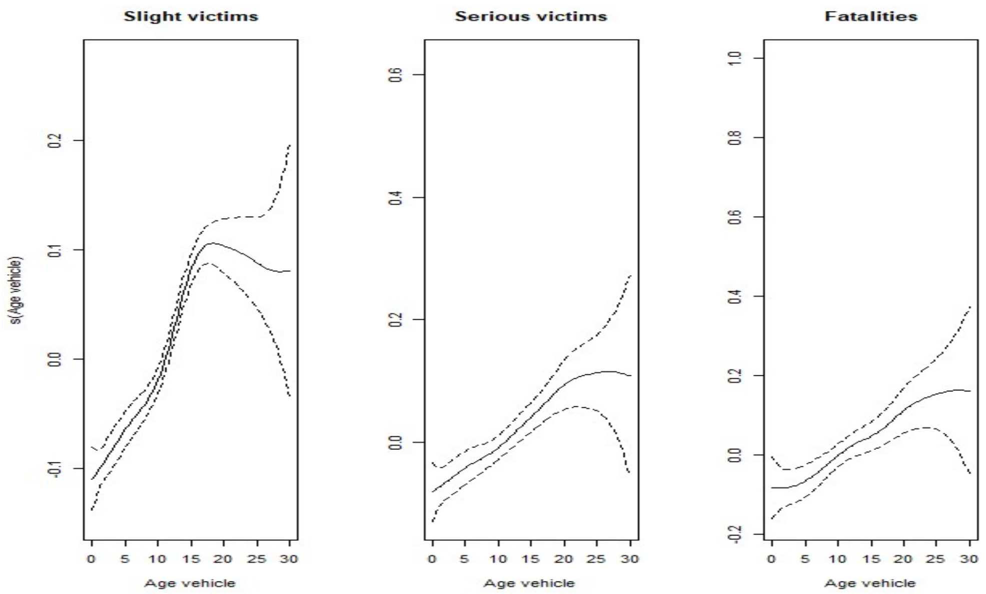
		BINOMIAL		POISSON		NEGATIVE BINOMIAL	
		Generalized Linear Model	Generalized Linear Mixed Model	Generalized Linear Model	Generalized Linear Mixed Model	Generalized Linear Model	Generalized Linear Mixed Model
Slight victims	AIC	147,940.4	146,992.0	176,808.1	176,810.1	176,811.7	176,812.7
	BIC	148,101.6	147,162.6	176,969.2	176,980.7	176,982.3	176,992.7
Serious victims	AIC	38,800.7	37,443.0	38,871.0	38,382.7	38,611.6	38,382.6
	BIC	38,961.8	37,613.6	39,032.1	38,553.2	38,782.1	38,562.7
Fatalities	AIC	11,752.0	11,430.0	11,675.9	11,456.3	11,463.9	11,448.9
	BIC	11,913.2	11,600.6	11,837.0	11,626.9	11,634.5	11,629.0

Analyzing the GLM for slight injury victims, the binomial model has lower AIC and BIC when compared to the Poisson model and the negative binomial model. The same order pertains when we consider the GLMM. In the case of serious injury victims and fatalities, the GLM with the lowest AIC and BIC is the negative binomial. However, the binomial model has the lowest AIC and BIC when GLMMs are considered. In fact, the binomial GLMM always has the lowest AIC and BIC among the set of models considered. These results suggest that when we introduce the random effect, this model is a better fit and captures the correlation between vehicles in the same car crash.

4.2 Relationship between vehicle age, driver age and the BI severity level

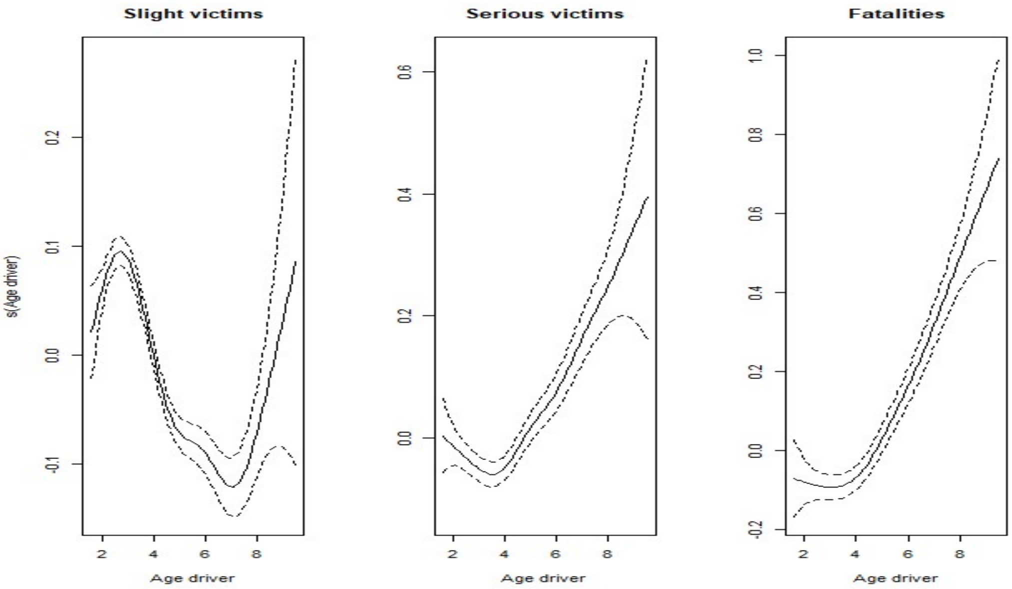
Both the GLMM and GLM frameworks assume that the relationship between continuous variables and the transformed dependent variable is linear. However, this is not always true. Here, we investigate the relationships between vehicle age and driver age, respectively, and the severity of injury sustained by occupants of the vehicle involved in the crash. In this sense, a semiparametric binomial GLMM is fitted to the data. This flexible modeling approach defines the linear predictor of the GLMM as a linear relationship between the categorical variables and smooth functions of the two explanatory continuous variables, *Vehicle age* and *Driver age*. Even though the estimation process frequently has less stability and the coefficients' interpretation is more complex, this flexible modeling is a powerful tool for understanding the effect of the explanatory continuous variables on the dependent variable, in a multivariate context. Figure 2 shows the estimated effect of vehicle age on the linear predictor of the binomial GLMM model. There is an appreciable change in the trend at around 20 years in the case of slight injury victims. This effect is less visible for serious injury and fatality victims.

Figure 2. Estimated effect of age of vehicle in the semiparametric binomial GLMM, by bodily injury severity level



The same analysis is conducted for driver age and results are displayed in Figure 3. In this case we observe a quadratic shape for serious injury victims (b) and fatalities (c). A more complex relationship is observed in the case of the slight injury victims (a), for which a quadratic shape would not capture the initial increase that occurs up to the age of around thirty.

Figure 3. Estimated effect of age of driver in the semiparametric binomial GLMM, by injury severity level



Different transformations of the two explanatory variables were analyzed to capture the relationships shown in Figures 2 and 3, including polynomials and linear approximations. Finally, the transformation associated with the lowest AIC when the model was fitted was selected.[3] In the case of vehicle age, the variable is replaced by two new regressors; *Young vehicle* is defined as a quantitative variable with a continuous part for those vehicles under 18 years old. This regressor takes the value of the vehicle age when it is under 18 years, and 0 otherwise. *Old vehicle* is defined as vehicles of 18 years or older. This regressor is defined as a dichotomous variable that takes the value 1 if the vehicle age is equal to or higher than 18, and 0 otherwise. In relation to driver age, a quadratic form returned the best fit for the three severity levels, including the number of slight injury victims. So, a new regressor is aggregated into the model recording the squared age of the driver (*Squared age*).

4.3 Binomial Generalized Linear Mixed Model

A binomial GLMM is fitted for the number of slight, serious and fatal injury victims in the vehicle, including the new regressors associated with vehicle age and driver age. Table 5 shows the estimated coefficients for the three binomial GLMMs. A negative (positive)

---

[3] Not shown for simplicity.

coefficient indicates a decrease (increase) on the expected number of victims with non-serious, serious or fatal injuries in the vehicle, respectively.

Table 5. Coefficient estimates of the binomial GLMM according to injury severity level of victims

		Slight		Serious		Fatal victims	
		Coef.	Std. Error	Coef.	Std. Error	Coef.	Std. Error
Intercept		0.175***	0.052	-4.840***	0.129	-7.308***	0.266
Gender	Male	-0.494***	0.015	0.213***	0.043	0.613***	0.097
Driver age	Driver age	-0.146***	0.022	-0.180***	0.053	-0.226**	0.098
	Squared age	0.008**	0.002	0.029***	0.006	0.047***	0.010
Vehicle age	Young vehicle	0.022***	0.001	0.021***	0.004	0.023**	0.007
	Old vehicle	0.390***	0.024	0.467***	0.061	0.660***	0.110
Vehicle	Van	-0.180***	0.025	-0.061	0.077	-0.125	0.140
	Motorcycle	1.606***	0.026	2.144***	0.043	1.34***	0.085
	Moped	1.966***	0.042	1.782***	0.065	0.869***	0.162
	Heavy vehicle	-0.694***	0.028	0.080	0.071	0.107	0.123
Illumination	No visibility	0.179***	0.021	0.502***	0.046	0.602***	0.078
Condition road	Non-optimal	0.261***	0.018	-0.358***	0.048	-0.338***	0.090
Road	Principal	0.244***	0.020	0.827***	0.055	1.356***	0.118
	Minor	0.260***	0.017	1.191***	0.043	1.721***	0.100
	Other	0.127***	0.027	0.503***	0.071	0.967***	0.151
Crash	Pile-up	-0.093***	0.025	-0.982***	0.108	-1.28***	0.238
	Run-over	-2.310***	0.039	-1.117***	0.126	-0.9***	0.260
	Rollover	0.698***	0.000	0.654***	0.040	0.811***	0.074
	Other	0.450***	0.026	0.581***	0.056	0.745***	0.099
Stand. Dev. (Random effect)		0.523		1.453		1.551	
AIC		146,930.2		37,417.6		11,403.4	
BIC		147,119.8		37,607.1		11,592.9	

Note: *** p-value<0.001; ** p-value<0.05; * p-value<0.10.

When the driver involved in the crash is male, this increases the expected number of seriously injured occupants and fatalities and decreases the number of slightly injured occupants. The driver's age also has an impact on the expected number of all victims. The expected number of victims decreases with increasing driver age until reaching a minimum, and later increases, and this holds for all types of victim. In the case of slight injury victims, the minimum is reached at the age of 91 years; for serious injury victims the minimum is reached at the age of 31 years, and for fatalities the minimum is reached at the age of 24 years. Driving old vehicles increases the expected number of injured occupants. The expected number of injured occupants increases per year of vehicle age up to 18 years, when the effect on the expected number of injured victims remains stable at the highest level for vehicles of 18 years and older, mainly for seriously injured victims and fatalities.

Compared to cars, the expected number of injured occupants increases for all three severity levels when they are traveling on two wheels: motorcycles, mopeds and bicycles. In the case of vans or heavy vehicles, the number of slightly injured victims is lower than in cars, but no significant differences are found in relation to the number of serious and fatal injury victims. Illumination, road surface conditions and type of road are significant factors in explaining the number of injured occupants. Driving under less than good visibility conditions increases the expected number of all types of injured victims. However, where the road surface is non-optimal, this increases the number of slightly injured victims but decreases the number of seriously injured victims and fatalities.

Principal and minor arterials are associated with a higher expected number of injured occupants than local roads. The estimated coefficients in minor arterials are slightly higher than in principal arterials, regardless of the severity level. So, the expected number of injured victims is higher in minor arterials than in principal arterials. Finally, when the crash is a pile-up or run-over, the expected number of injured victims in the vehicle decreases in comparison to collisions involving other vehicles, collisions with an obstacle, rollovers or drops.

5. Discussion

This study analyzes several factors of risk in road crashes that affect the expected number of BI victims in a vehicle by severity level, taking into account dependence between vehicles involved in the same crash. The observation unit in the analysis is the vehicle. Dependence between vehicles involved in the same crash is considered, including random effects, in the regression. It is shown that the model performance improves when dependence is considered. Even the selected conditional error distribution for serious injuries and fatalities varies when this phenomenon is considered. So, the inclusion of random effects captures, at least partially, the heterogeneity due to the involvement of more than one vehicle in the same motor crash.

When two or more vehicles are involved in the same crash one could expect to derive a relationship between their respective resulting damage and injury. Various papers inform us about the incidence injury among different types of vehicles affected (for example a car and a heavy vehicle) or the positions of the occupants inside the vehicles. Dependence between the BI severity levels of people involved in the same crash can be very relevant if we want to predict the expected number of victims and their injury severity; for example, as a consequence of a safety policy or, more specifically, in the insurance context, when we want to calculate provisions for automobile claims coverage. Methodologically speaking, this objective accords with previous studies that suggest that ignoring the possibility of including random parameters when estimating count-data models could affect the magnitude of the coefficients (Anastasopoulos and Mannering, 2009, 2011; Osman et al., 2018; Hosseinpour et al., 2018).

Our results confirm conclusions previously drawn in the literature. Male drivers are associated with more seriously injured victims (see a review, for example, in Regev et al., 2018). Two-wheeled motor vehicles are more likely to be associated with serious or fatal injuries than four-wheeled or heavier vehicles (Donate-López et al., 2010; Schneider et al., 2012). This result is expected since two-wheeled vehicles offer less protection to riders. For instance, previous literature suggests that heavy vehicles (pickup trucks, minivans and sport utility vehicles -SUVs-) are safer for their own occupants and cause more damage to the other vehicles involved in a crash (Fredette et al., 2008; George et al., 2017). A number of studies have found that driving in dark conditions increases expected accident severity (Sullivan and Flannagan, 2002; Wanvik, 2009; Uddin and Huynh, 2017). Sullivan and Flannagan (2002) concluded that the risk of fatally injury to pedestrians involved in crashes is between 3 and 6.75 times higher in the dark than in daylight. Wanvik (2009) found that the risk of injury from crashes in darkness increases on average by 17% on lit rural roadways and by 145% on unlit rural roadways. Uddin and Huynh (2017) also confirm the importance of examining crashes based on lighting conditions on rural and urban roadways. Here, we have found that the expected numbers of slight, serious and fatal injuries to victims increase when there is less than good visibility.

We found that non-optimal road surface conditions increase the expected number of slightly injured occupants, but they reduce the expected number of serious and fatal injury victims. Although we could expect an increase in crash BI severity because of bad road conditions (bad weather, poor surface, ...), we argue that unobserved factors may have an opposite effect, such as increased attention to driving, higher traffic density or higher signaling rates. Different studies have shown that the influence of good road conditions on traffic crashes and levels of injury is not clear, with positive effects in some cases and negative in others (see, for example, Mondal et al., 2011).

The expected number of injured victims is higher on principal and minor arterials than on local roads. While the number of crashes in local areas is usually higher than on arterials and collectors, they are associated with a lower severity (DGT, 2017). The type of crash analyzed has an explicit influence on the expected number of occupants injured. When a vehicle is involved in a multiple collision the expected number of injured occupants falls as compared to a two-vehicle collision. Collisions involving multiple vehicles (pile-up) are more frequently rear impact crashes, which are associated with less severe BI outcomes (Abu-Zidan and Eid, 2015; DGT, 2017). Abu-Zidan and Eid (2015) indicated that injury severity among those involved in front and side impacts was double that of rear impacts. Also, the expected number of injured occupants falls when the type of crash is a run-over as compared to a two-vehicle collision. Note that in a crash collision involving a pedestrian, the pedestrian is expected to sustain the highest BI damage (Pour-Rouholamin and Zhou, 2016; Islam and Jones, 2014), while the occupants of the vehicle (whom we are analyzing in this study) are more protected road users. Finally, when the vehicle is involved in a rollover, drop or collision with an object, an increased number of injured occupants is expected for all levels of severity. Note that our results are based on motor crashes involving injured victims. In the literature, when crashes with victims are analyzed, single-vehicle crashes are frequently associated with more severe BI damages than collisions involving two or more vehicles (Daniels et al., 2010; Abu-Zidan and Eid, 2015; DGT, 2017).

Our analysis pays special attention to the age of the driver and the vehicle age as factors explaining the number of occupants injured with different severity levels as a result of a crash. We demonstrate that the relationship between these factors and the (transformed) dependent variable is not linear. Both factors were redefined to reflect their association with the expected number of injured occupants.

For the age of the driver, we found a quadratic relationship with occupants' injury severity. Indeed, young drivers and old drivers were the riskiest groups. Previous studies have identified these two high risk groups of drivers (Rakotonirainy et al., 2012; Zhou et al., 2015; Regev et al., 2018). Here, we found that young drivers presented a high risk of accidents resulting in slightly injured occupants, but old drivers were the riskiest age group in the case of seriously and fatally injured crash victims. This does not mean that older drivers are necessarily more dangerous drivers since older drivers (and likely their old passengers) are inherently more likely to be seriously injured in crashes due to their physical fragility (Regev et al., 2018; Noh & Yoon, 2017). Previous studies have suggested the need for the effect of an increasing proportion of elderly road users to be very present in road safety policies (Loughran et al., 2007; Boot et al., 2014; Ayuso et al., 2020). The number of older drivers is increasing in many countries as a result of the general population aging. As the number of older drivers is becoming more significant, researchers have access to an increasing amount of data about this group of drivers, opening an important area of future research.

The age of vehicles is also gaining attention in road safety research. In fact, previous studies have suggested that vehicle age is positively associated with driver age (Ayuso et al., 2019; Eby et al., 2016; Simões y Pereira, 2009, among others). New vehicles are safer as a result of technological and safety advances implemented in the new generation of automobiles. Here, we found that the expected number of occupants injured by severity level increases with the vehicle age up to 18 years and then remains constant at the highest level. This finding is very relevant in countries with old fleets of automobiles, such as Spain, where the average age of automobiles has increased from 7.65 years in 2002 to 12.42 years in 2018 (ANFAC, 2019). In the EU, passenger cars were on average 10.8 years old in 2018 (ACEA, 2019).

The high level of significance of most parameter estimates provides a better understanding of the effect of automobile and crash characteristics on the expected number of occupants injured by severity level. Nevertheless, this study has limitations. We control heterogeneity due to multiple vehicles being involved in the same crash, but other sources of unobserved heterogeneity were not controlled in the study. For example, count BI severity models were estimated separately for the different levels of injury severity experienced by occupants, but some unobserved factors could impact all levels of severity, simultaneously. In addition, relevant information to explain the severity of the crash was not available in the dataset. For instance, the age and position of passengers in the vehicle, the use of safety measures or the place where the crash occurred have been extensively investigated as factors influencing crash severity (Abdel-Aty, 2003; Smith and Cummings, 2006; Abay et al., 2013). All these unavailable factors constitute unobserved heterogeneity as well as a lack of information related to driving behavior. Research in the context of telemetry shows a close relationship between driving behavior and crash severity (Ayuso et al., 2019; Perez-Marín et al., 2019; Paetgen et al., 2013, 2104, among others). The incorporation of driving behavior information into the model could differentiate aspects that would allow a deeper knowledge of the influence of traditional risk factors. For example, the driving behavior of old drivers may help to distinguish the proportion of the higher crash severity risk attributable to declining skill and the proportion associated with increased physical frailty.

6. Conclusions

Modeling the number of slight, serious and fatal injury victims in a vehicle involved in a crash should include the dependence between all vehicles involved in the same crash. The inclusion of random effects in the regression to capture this phenomenon significatively improves the quality of fit. The gender of the driver, the type of road, the type of vehicle involved in the crash, the visibility and road conditions, and the type of crash are factors with explanatory capacity for the expected number of occupants injured in each vehicle according to BI severity level. The age of the driver and the age of the vehicle have a nonlinear influence on severity. The expected number of victims increases for both young and old drivers. The age of the vehicle increases the expected number of injured occupants with the highest impact being for cars of 18 years or older. Accurate modeling of the number of injured occupants by severity level that takes into account the dependence between vehicles involved in the same crash is relevant for traffic authorities in every country as well as for motor insurance companies who cover damages for victims involved in motor crashes. The premium design could be improved if the expected number of victims by severity level is included in the estimation of the crash severity.

Acknowledgments

We are grateful to the Spanish Ministry of Science and Innovation for funding received under grant PID2019-105986GB-C21. We are also grateful to the *Dirección General de Tráfico* for access to their database. We thank Rodrigo Sánchez for his comments and his help in the design of the database.

References

Abay, K. A., Paleti, R., Bhat, Ch. R. (2013). The joint analysis of injury severity of drivers in two-vehicle crashes accommodating seat belt use endogeneity. Transportation Research Part B, 50, 74–89.

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

Abu-Zidan, F. M., Eid, H. O. (2015). Factors affecting injury severity of vehicle occupants following road traffic collisions. Injury, 46, 1, 136–141.

ACEA (2019) ACEA Report Vehicles in use Europe 2019. European Automobiles Manufacturers Association.

Alam, B. M., Spainhour, L. K. (2008). Contribution of behavioral aspects of older drivers to fatal traffic crashes in Florida. Transportation Research Record: Journal of the Transportation Research Board, 2078, 49–56.

Anastasopoulos, P. C., Mannering, F. L. (2009). A note on modeling vehicle accident frequencies with random-parameters count models. Accident Analysis and Prevention, 41, 1, 153–159.

Anastasopoulos, P. Ch., Mannering, F. L. (2011). An empirical assessment of fixed and random parameter logit models using crash and non-crash-specific injury data. Accident Analysis and Prevention, 43, 1140-1147.

ANFAC (2019) Informe Anual 2018. Asociación Española de Fabricantes de Automóviles y Camiones.

Anderson, R. W. G., Searson, D. J. (2015). Use of age–period–cohort models to estimate effects of vehicle age, year of crash and year of vehicle manufacture on driver injury and fatality rates in single vehicle crashes in New South Wales, 2003–2010. Accident Analysis and Prevention, 75, 202–210.

Ayuso, M., Guillen, M., Nielsen, J. P. (2019). Improving automobile insurance ratemaking using telematics: incorporating mileage and driver behaviour data. Transportation, 46, 3, 735-752.

Ayuso, M., Sánchez, R., Santolino, M. (2020). Does longevity impact the severity of traffic crashes? A comparative study of young-older and old-older drivers. Journal of Safety research, 73, 37-46.

Ayuso, M., Sanchez, R., Santolino, M. (2019). Longevidad de los conductores y antigüedad de los vehículos: impacto en la severidad de los accidentes. Anales del Instituto de Actuarios Españoles, 4ª época, 25, 33-53.

Blows, S., Ivers, R. Q., Woodward, M., Connor, J., Ameratunga, S., Norton, R. (2003). Vehicle year and the risk of car crash injury. Injury Prevention, 9, 353-356.

Boot, W. R., Stothart, C., Charness, N. (2014). Improving the Safety of Aging Road Users: A Mini-Review. Gerontology, 60, 90-96.

Chin, H. C., Zhou, M. (2018). A study of at-fault older drivers in light-vehicle crashes in Singapore. Accident Analysis and Prevention, 112, 50–55.

Clarke, D. D., Ward, P., Bartle, C., Truman, W. (2010). Older drivers' road traffic crashes in the UK. Accident Analysis and Prevention, 42, 1018-1024.

Daniels, S., Brijs, T., Nuyts, E., Wets, G. (2010) Externality of risk and crash severity at roundabouts, Accident Analysis & Prevention, 42, 6, 1966-1973.

DGT (2017). Las principales cifras de la siniestralidad vial, España 2016. Dirección General de Tráfico, Ministerio del Interior, Gobierno de España.

Donate-López, C., Espigares-Rodríguez, E., Jiménez-Moleón, J. J., de Dios Luna-del-Castillo, J., Bueno-Cavanillas, A., Lardelli-Claret, P. (2010). The association of age, sex and helmet use with the risk of death for occupants of two-wheeled motor vehicles involved in traffic crashes in Spain. Accident Analysis and Prevention, 42, 297–306.

Eby, D. W., Molnar, L. J., Zhang, L., St. Louis, R. M., Zanier, N., Kostyniuk, L. P., Stanciu, S. (2016). Use, perceptions, and benefits of automotive technologies among aging drivers. Injury Epidemiology, 3, 28, 1–20.

Eluru, N., Paleti, R., Pendyala, R., Bhat, C. R. (2010). Modeling injury severity of multiple occupants of vehicles: Copula-based multivariate approach. Transportation Research Record, 2165, 1–11.

Fredette, M., Mambu, L. S., Chouinard, A., Bellavance, F. (2008). Safety impacts due to the incompatibility of SUVs, minivans, and pickup trucks in two-vehicle collisions. Accident Analysis and Prevention, 40, 1987-1995.

Funk, J. R., Cormier, J. M., Manoogian, S. J. (2012). Comparison of risk factors for cervical spine, head, serious, and fatal injury in rollover crashes. Accident Analysis and Prevention, 45, 67-74.

George, Y., Athanasios, T., George, P. (2017). Investigation of road accident severity per vehicle type. Transportation Research Procedia, 25, 2076–2083.

Golub, G. H., Welsch, J. H. (1969). Calculation of Gauss quadrature rules. Mathematics of Computation, 23, 106, 221–230.

Hosseinpour, M., Sahebi, S., Zamzuri, Z. H., Yahaya, A. S., Ismail, N. (2018). Predicting crash frequency for multi-vehicle collision types using multivariate Poisson-lognormal spatial model: A comparative analysis. Accident Analysis and Prevention, 118, 277–288.

Høye, A. (2019). Vehicle registration year, age, and weight – Untangling the effects on crash risk. Accident Analysis and Prevention, 123, 1–11.

Islam, S., Jones, S. L. (2014). Pedestrian at-fault crashes on rural and urban roadways in Alabama. Accident Analysis and Prevention, 72, 267–276.

Ivarsson, J., Poplin, G., McMurry, T., Crandall, J., Kerrigan, J. (2015). Occupant injury in rollover crashes – contribution of planar impacts with objects and other vehicles. Accident Analysis and Prevention, 85, 111–117.

Jackson, T. L., Mello, M. J. (2013). Injury patterns and severity among motorcyclists treated in US emergency departments, 2001–2008: a comparison of younger and older riders. Injury Prevention, 19, 297-302.

Johannsen, H., Müller, G. (2013). Accident and injury risks of elderly car occupants. In 23rd Conference on the Enhanced Safety of Vehicles, 13-0223, 1–10.

Loughran, D. S., Seabury, S. A., Zakaras, L. (2007). Regulating Older Drivers Are New Policies Needed? Rand Institute for Civil Justice, Occasional Paper.

Mannering, F. L., Shankar, V., Bhat, Ch. R. (2016) Unobserved heterogeneity and the statistical analysis of highway accident data. Analytic Methods in Accident Research, 11, 1-16.

McCulloch, C., Searle, S. R. (2001). Generalized, Linear, and Mixed Models. John Wiley & Sons, Inc., New York.

Milton, J., Shankar, V., Mannering, F. (2008). Highway accident severities and the mixed logit model: An exploratory empirical analysis. Accident Analysis and Prevention 40(1), 260-266.

Mondal, P., Sharma, N., Kumar, A., Bhangale, U. et al.. (2011). Effect of rainfall and wet road condition on road crashes: A critical analysis, SAE Technical Paper 2011-26-0104.

Naylor, J. C., Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. Applied Statistics, 31, 3, 214-225.

NHTSA (2013). How vehicle age and model year relate to driver injury severity in fatal crashes. National Highway Traffic Safety Administration, NHTSA's National Center for Statistics and Analysis, DOT HS 811 825.

Noh, Y., & Yoon, Y. (2017). Elderly road collision injury outcomes associated with seat positions and seatbelt use in a rapidly aging society—A case study in South Korea. PLoS ONE 12(8): e0183043. https://doi.org/10.1371/journal.pone.0183043

Osman, M., Mishra, S., Paleti, R. (2018). Injury severity analysis of commercially-licensed drivers in single-vehicle crashes: Accounting for unobserved heterogeneity and age group differences. Accident Analysis and Prevention, 118, 289-300.

Paefgen, J., Staake, T., Fleisch, E. (2014). Multivariate exposure modelling of accident risk: insights from pay-as you-drive insurance data. Transportation Research Part A: Policy and Practice, 61, 27–40

Paefgen, J., Staake, T., Thiesse, F. (2013). Evaluation and aggregation of pay-as-you-drive insurance rate factors: a classification analysis approach. Decision Support Systems, 56, 192–201.

Pérez-Marín, A. M., Ayuso, M., Guillen, M. (2019). Do young insured drivers slow down after suffering and accident? Transportation Research Part F: Traffic Psychology and Behaviour, 62, 1, 690-699.

Pour-Rouholamin, M., Zhou, H. (2016). Investigating the risk factors associated with pedestrian injury severity in Illinois. Journal of Safety Research, 57, 9–17.

Qin, X., Wang, K., Cutler, Ch. E. (2013). Analysis of crash severity based on vehicle damage and occupant Injuries. Transportation Research Record: Journal of the Transportation Research Board, No. 2386, Transportation Research Board of the National Academies, Washington, 95-102.

Quddus, M. A., Noland, R. B., Chin, H. C. (2002). An analysis of motorcycle injury and vehicle damage severity using ordered probit models. Journal of Safety Research, 33, 445–462.

Rakotonirainy, A., Steinhardt, D., Delhomme, P., Darvell, M., Schramm, A. (2012). Older drivers' crashes in Queensland, Australia. Accident Analysis and Prevention, 48, 423–429.

Regev, S., Rolison, J. J., Moutari, S. (2018). Crash risk by driver age, gender, and time of day using a new exposure methodology. Journal of Safety Research, 66, 131-140.

Rich, J., Prato, C-G., Hels, T., Lyckegaard, A., Kristensen, N.B. (2013). Analyzing the relationship between car generation and severity of motor-vehicle crashes in Denmark. Accident Analysis and Prevention, 54, 81–89.

Schneider, W. H., Savolainen, P. T., Boxel, D. V., Beverley, R. (2012). Examination of factors determining fault in two-vehicle motorcycle crashes. Accident Analysis and Prevention, 45, 1, 669–676.

Simões, A., Pereira, M. (2009). Older drivers and new in-vehicle technologies: adaptation and long-term effects. Human Centered Design, LNCS 5619, 552-561.

Smith, K. M., Cummings, P. (2006). Passenger seating position and the risk of passenger death in traffic crashes: a matched cohort study. Injury Prevention, 12, 83-86.

Sullivan, J. M., Flannagan, M. J. (2002). The role of ambient light level in fatal crashes: Inferences from daylight saving time transitions. Accident Analysis and Prevention, 34, 4, 487–498.

Uddin, M., Huynh, N. (2017). Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accident Analysis and Prevention, 108, 44–55.

Wanvik, P.O. (2009). Effects of road lighting: an analysis based on Dutch accident statistics 1987–2006. Accident Analysis and Prevention, 41, 1, 123–128.

Wang, X., Kockelman, K. M. (2005). Use of heteroscedastic ordered logit model to study severity of occupant injury: distinguishing effects of vehicle weight and type. Transportation Research Record: Journal of the Transportation Research Board, No. 1908, Transportation Research Board of the National Academies, Washington, 195–204.

Washington, S., Karlaftis, M., Mannering, F., Anastasopoulos, P. (2020). Statistical and econometric methods for transportation data analysis. CRC Press, Taylor and Francis Group, New York, NY.

Zhou, H., Zhao, J., Pour-Rouholamin, M., Tobias, P. A. (2015). Statistical characteristics of wrong-way driving crashes on Illinois freeways. Traffic Injury Prevention, 16, 8, 760–767.

UBIREA
Institut de Recerca en Economia Aplicada Regional i Pública
Research Institute of Applied Economics
Universitat de Barcelona
Av. Diagonal, 690 • 08034 Barcelona

WEBSITE: www.ub.edu/irea/ • **CONTACT:** irea@ub.edu