Addressing Unobserved Heterogeneity in the Analysis of Bicycle Crash Injuries in Scotland: A Correlated Random Parameters Ordered Probit Approach with Heterogeneity in Means

Grigorios Fountas (Corresponding Author)
Transport Research Institute, School of Engineering and the Built Environment
Edinburgh Napier University, Edinburgh, Scotland, United Kingdom, EH10 5DT
Email: g.fountas@napier.ac.uk

Achille Fonzone
Transport Research Institute, School of Engineering and the Built Environment
Edinburgh Napier University, Edinburgh, Scotland, United Kingdom, EH10 5DT
Email: a.fonzone@napier.ac.uk

Adebola Olowosegun
Transport Research Institute, School of Engineering and the Built Environment
Edinburgh Napier University, Edinburgh, Scotland, United Kingdom, EH10 5DT
Email: a.olowosegun@napier.ac.uk

Clare McTigue
Transport Research Institute, School of Engineering and the Built Environment
Edinburgh Napier University, Edinburgh, Scotland, United Kingdom, EH10 5DT
Email: c.mctigue@napier.ac.uk

Submitted: January 4th, 2021
First Revision: June 12th, 2021
Second Revision: June 19th, 2021

ABSTRACT

This paper investigates the determinants of injury severities in single-bicycle and bicycle-motor vehicle crashes by estimating correlated random parameter ordered probit models with heterogeneity in the means. This modeling approach extends the frontier of the conventional random parameters by capturing the likely correlations among the random parameters and relaxing the fixed nature of the means for the mixing distributions of the random parameters. The empirical analysis was based on a publicly available database of police crash reports in the UK using information from crashes occurred on urban and rural carriageways of Scotland between 2010 and 2018. The model estimation results show that various crash, road, location, weather, and driver or cyclist characteristics affect the injury severities for both categories of crashes. The heterogeneity-in-the-means structure allowed the incorporation of a distinct layer of heterogeneity in the statistical analysis, as the means of the random parameters were found to vary as a function of crash or driver/cyclist characteristics. The correlation of the random parameters enabled the identification of complex interactive effects of the unobserved characteristics captured by road, location and environmental factors. Overall, the determinants of injury severities are found to vary between single-bicycle and bicycle-motor vehicle crashes, whereas a number of common determinants are associated with different effects in terms of magnitude and sign. The comparison of the proposed methodological framework with less sophisticated ordered probit models demonstrated its relative benefits in terms of statistical fit, explanatory power and forecasting accuracy as well as its potential to capture unobserved heterogeneity to a greater extent.

Keywords: Single-bicycle crashes, Bicycle-motor vehicle crashes, injury severity, ordered probit, correlated random parameters; unobserved heterogeneity; Scotland

INTRODUCTION

Besides being popular for recreational and exercise purposes, cycling is a low-cost and valid transportation alternative to motorized modes with benefits on the environment and public health for a large proportion of daily trips (Cycling Scotland, 2019). In Scotland, over the last decade, shares for cycling as a main mode, cycling to work, and cycling to school have all increased. The latest statistics indicate that 4.9% of people cycled to work at least regularly, while 5.2% of primary school pupils and 1.3% of secondary school pupils cycled to school (Cycling Scotland, 2019). However, cyclists are highly susceptible to road crashes, resulting in significant human losses, large medical and productivity costs, and limited public confidence on the safety of cycling. Even though the cycling modal share has consistently been around 1.4% in Scotland (Fountas et al., 2020b) over the last few years (see Figure 1), crashes with serious injuries to cyclists have increased by 34% and the rate of Killed or Seriously Injured (KSI) people per million vehicle KMs by 18%, (see Figure 1). In 2018, 8% of the reported road casualties are cyclists (Young and Whyte, 2020). Future scenarios based on English national survey data show that a potential increase of bicycle trips along with a simultaneous decrease of the passenger trips of public transport (a scenario brought about by the Covid-19 outbreak in several places) would further increase the cyclist casualties (Smith et al., 2019). In this context, the safety level of cyclists constitutes a significant barrier towards the increase of the demand for bicycle trips.

Young and Whyte (2020) point out that, within Scotland, it is difficult to accurately determine the risks associated with cycling due to a lack of reliable data on the number of cyclists, of cycle trips they take, and of miles cycled. This important information is lacking not only in Scotland, but also in various countries across the globe where extensive crash databases are being kept (Poulos et al., 2015; Mannering et al., 2016).

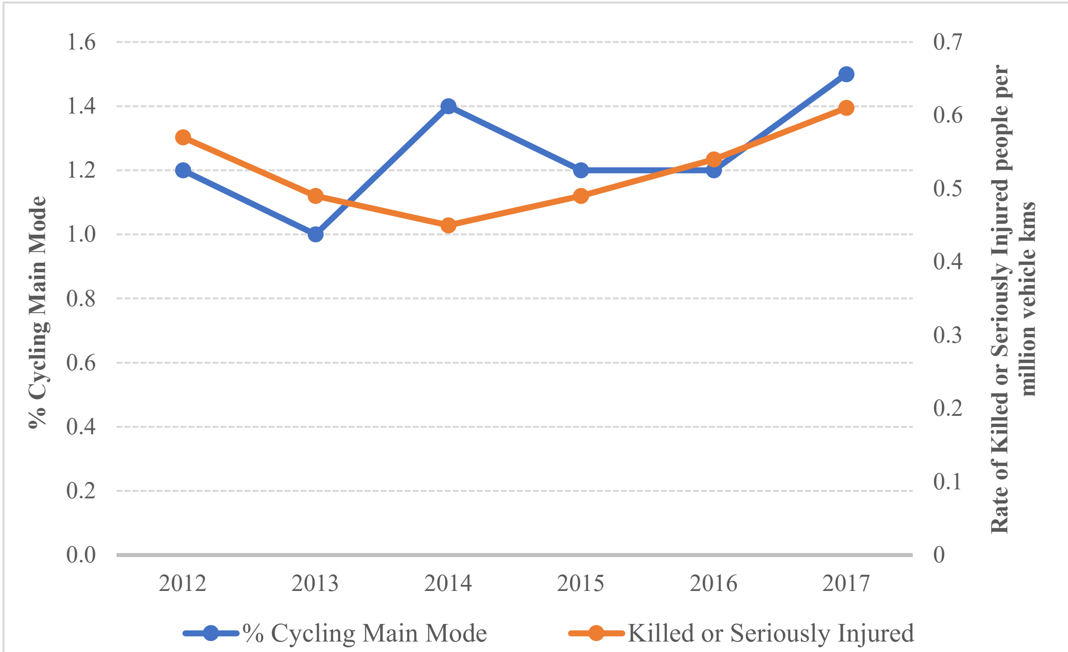

Figure 1 Percentage of journeys with cycling as the main mode and rate of Killed or Seriously Injured cyclists in Scotland between 2012 and 2018. Source: Transport Scotland (2018a)

The vulnerability of cyclists primarily stems from their interactions and conflicts with motor vehicles (e.g., passenger cars, motorcycles, trucks, etc.) on roadway infrastructure not designed for accommodating mixed traffic flows. For example, only 28% of the cycle path network in Scotland is on dedicated infrastructure (Cycling Scotland, 2018), hence free from the risks deriving from the interactions with motor vehicles. 94% of the cyclists' crashes involve at least one passenger car or light goods vehicle. These statistics are in line with previous findings from road safety analyses across the globe, which also show that the bicycle – motor vehicle crashes are more likely to result in casualties (Eluru et al., 2008). The exposure of cyclists to conflicts with motorized traffic is particularly a concern in the UK, where cyclists are often required to share road space with other vehicles in urban settings, thus inducing an increased risk for cyclists and frustration for drivers (Shackel and Parkin, 2014). Evidence also suggests that cycling in a country with a low cycling mode share like Scotland is less safe than cycling in a context where cycling is popular such as the Netherlands and Sweden. Cycling in the Netherlands is reportedly three times safer compared with the UK (Marshall and Ferenchak, 2019), while one-third of the conflicts and critical events that cyclists experienced in 2014 involved a motor vehicle in Sweden (Young and Whyte, 2020).

Despite the potential for severe consequences, the crash-specific circumstances that determine the injury-severity of bicycle-motor vehicle crashes have not been extensively investigated locally or internationally. Recent studies (Kim et al., 2007; Eluru et al., 2008; Wang et al., 2015; Behnood and Mannering, 2017) have shown that various cyclist- or driver-specific traits, roadway and traffic attributes, bicycle or motor vehicle condition and various environmental characteristics (such as weather, visibility, lighting conditions) constitute possible determinants of cyclists' injury severities. Several analyses typically focus on the crash propensity of cyclists using spatially aggregate data (e.g., traffic zone-based analyses – see Cai et al., 2016; Nashad et al., 2016) or crash-level data without accounting for the presence of unobserved factors that may have pronounced effects on the injury-severities of cyclists or motor vehicle occupants. Consider, for example, the instantaneous weather conditions at the time of the bicycle - motor vehicle crash occurrence, which are not typically available in the conventional datasets. Despite their unobserved nature, instantaneous weather or environmental conditions may have a critical impact on driving manoeuvres (Fountas et al., 2018a), which determine –to a large extent– the seriousness of the bicycle-motor vehicle conflicts (Zheng et al., 2021). State-of-practice injury severity models (e.g., using fixed-parameters or fixed-effects formulations) cannot capture this generalised effect. Such a limitation can lead to biased predictors and erroneous inferences, and subsequently, to inappropriate or ineffective countermeasures. Lighting characteristics, pavement surface conditions, roadway infrastructure elements as well as the behavioural responses of road users to all the aforementioned roadway and environmental characteristics may constitute fundamental sources of common unobserved effects (typically referred to as unobserved heterogeneity – see also the extensive discussion provided by Mannering et al., 2016).

Single-bicycle crashes constitute another major source of road casualties. In England, more than 80% of cyclist casualties are attributed to single-bicycle crashes (Schepers et al., 2015). Despite the significant burden on public health, this type of crash is not extensively reported in road safety statistics. For example, a comparison of police records and hospital admission data gathered between January 2004 and December 2008 in Victoria, Australia showed that single-vehicle crashes represented only 5.2% of all cyclist road crashes in police data, compared to 55% recorded in hospital data (Boufous et al., 2013). In Finland, data from road traffic crashes recorded in the period 2014-2017 demonstrated that 30% of seriously injured people were cyclists. However, hospital records did not include data on crash characteristics (such as crash types), therefore, the characteristics of these bicycle crashes were unknown (Utriainen, 2020). Some studies suggest that cyclists are reluctant to report crashes to police, particularly when no vehicle was involved (Young and Whyte, 2020), despite the fact that single-bicycle crashes account for the vast majority of bicycle-involved crashes resulting in serious injuries, even in cycling-friendly countries such as Netherlands (Shinar et al., 2018).

Identifying the mechanism underpinning the occurrence of single-bicycle crashes is a quite challenging process. Human errors and cyclists' behavioural patterns as well as interactions with external stimuli (e.g., pedestrians, weather conditions, animals, and physical or built environment components) and road geometric characteristics may affect the injury severities of single-bicycle crashes. Such contributing factors cannot be captured through the conventional statistical analyses of police crash reports or hospital injury data, whereas the state-of-the-art crash models cannot adequately account for the full spectrum of unobserved factors (Mannering et al., 2016; 2020). Hence, very few studies have investigated the mechanism of single-bicycle crashes worldwide (Schepers and Wolt, 2012; Schepers et al., 2015; Boufous et al., 2013; Myhrmann et al., 2020).

This study aims at unveiling underlying dynamics in the generation mechanisms of single-bicycle and bicycle–motor vehicle crashes by statistically analyzing disaggregate crash data, and subsequently, gauging the influence of both unobserved and observed factors on recorded injury severities. As widely documented in earlier research, an abundance of potentially contributing factors are not recorded in the available crash datasets (Mannering et al., 2016; Fountas and Anastasopoulos, 2017), thus giving rise to the issue of unobserved heterogeneity. In this study, we account for unobserved heterogeneity by estimating Correlated Random Parameters Ordered Probit with Heterogeneity in the Means (CRPOPHM) models. These models are structured upon the well-established random parameter framework having also the potential to capture interdependencies between the sources of unobserved heterogeneity through a generalized formulation of the random parameters.

The empirical analysis is targeted on Scotland, where very few studies have explored the safety performance of bicycles (Whyte and Waugh, 2015; Transport Scotland, 2018b), and these primarily at aggregate level, without providing microscopic insights at crash level. This study fills also this gap of empirical knowledge, by analyzing disaggregate injury-severity data of single-bicycle crashes and bicycle–motor vehicle crashes occurred across entire Scotland. The findings shed more light on the quantitative and qualitative effects of the driver, cyclist, crash, roadway and environmental factors on the injury severities of bicycle-involved crashes.

METHODOLOGICAL FRAMEWORK

The Correlated Random Parameters Ordered Probit with Heterogeneity in the Means (CRPOPHM) models extend the correlated random parameter ordered probit framework (Fountas et al., 2018b) by accounting for heterogeneity in the means of the random parameters. In such way, possible correlations between the unobserved factors as well as variations in the effects of unobserved factors are simultaneously captured in the integrated modeling framework. To the best of the authors' knowledge, this is the first time that the specific modeling framework is employed in safety research for the analysis of crash injury severities.

The ordered probability framework has been used extensively for modeling injury severities. The approach accounts for the ordinal discrete nature of the data (Yasmin and Eluru, 2013; Yasmin et al., 2015; Bogue et al., 2017; Marcoux et al., 2018). The conventional ordered probability model is defined on the basis of a latent continuous variable, z, as expressed in the following Equation (Washington et al., 2020)**:**

$$z_i = \boldsymbol{\beta}\mathbf{X_i} + \varepsilon_i, y_i = j, if\ \mu_{j-1} < y_i < \mu_j,\ j = 1,2, \ldots..J \tag{1}$$

where **β** denotes a vector of estimable parameters, the vectors with the potential explanatory variables for crash *i* are denoted by $\mathbf{X}_i$, y_i is an integer standing for the observed injury-severity outcome, *j* stands for the integers representing the injury-severity levels, μ_j represent the threshold parameters, ordered in nature (i.e. such that $\mu_{j-1} < \mu_j$), for the determination of y_i. Finally, ε_i represents a random error component assumed normally distributed in the probit formulation.

Random parameters are introduced to account for the effect of unobserved factors varying systematically across the crash observations (Intini et al., 2020). A major improvement allowed by the random parameters modeling is the estimation of crash-specific parameter vectors for the explanatory variables. To capture the fundamental variations in the effects of observable attributes, crash-specific parameter vectors ($\boldsymbol{\beta}_i$) can be estimated as follows**:**

$$\boldsymbol{\beta}_\mathbf{i} = \boldsymbol{\beta} + \boldsymbol{\lambda}\mathbf{C} + \Gamma\omega_i \tag{2}$$

Where **β** denotes the mean value of the random parameters vector**,** Γ denotes the Cholesky matrix used for the estimation of the covariance matrix for the random parameters (see also Greene, 2016; Fountas et al., 2018a; 2018b; Pantangi et al., 2020 for more details on the process of the Cholesky decomposition), **C** represents a vector of covariates determining the means of the random parameters, i.e., the variables capturing heterogeneity in the means (Eker et al., 2019; Al-Bdairi et al., 2020; Hamed and Al-Eideh, 2020; Ahmed et al., 2020; Yan et al., 2021), **λ** is a vector of coefficients, ω represents a random term, which follows the standard normal distribution (Washington et al., 2020). The incorporation of the **λC** term constitutes an important feature of Equation (2), as it allows the identification of unobserved effects, which differ from those captured by the random parameters but interact with them. Such interactions induce variations in the parametric functions of random parameters, as the distributions of the latter are not fixed, but their means are being determined by the exogenous variables included in **C.**

Following the specification of Washington et al. (2020), the random parameters vary across the observations according to a pre-specified distribution, the mixing distribution. The latter is defined by a density function $q(\boldsymbol{\beta_i}|\boldsymbol{\Delta})$, where **Δ** indicates the vector of the parameters of the distribution**.** In this study, the normal distribution was used to fit the mixing distribution of the random parameters. Mean and standard deviation constitute the defining parameters of the normal distribution[1].

The covariance matrix of the random parameters, V, is calculated as a product of a Cholesky matrix, Γ, and a Cholesky matrix prime, Γ^{ι}, (Greene, 2016; Fountas et al., 2018a, 2018b; Fanyu et al., 2021; Pantangi et al., 2021)**:**

$$V = \Gamma\Gamma^{\iota} \tag{3}$$

In the traditional definition of the covariance matrix of the random parameters, the variances of the random parameters are the diagonal elements of the matrix and the off-diagonal elements of the matrix are set as zero. That is a consequence of the implicit assumption that the random parameters are

[1] To account for possible heterogeneity in the standard deviations of the mixing distributions of the random parameters (i.e., heterogeneity in variances), we also attempted to use an even more generalized formulation for the random parameters, similar to that used in Behnood and Mannering (2017) and Al-Bdairi et al. (2020). However, no statistically significant determinants of the standard deviations were identified throughout the model estimation process, hence the standard deviations of the random parameters were specified as fixed.

uncorrelated. However, previous research has shown that the sources of unobserved heterogeneity may not be independent from each other, thus giving rise to the possibility of interdependencies between the unobserved characteristics captured by the random parameters (Fountas et al., 2019; Eker et al., 2019). To account for the possible correlations between the random parameters, a more generalized formulation of the Cholesky matrix, Γ, is adopted, with the off-diagonal elements taking non-zero values. This unrestricted specification of the Γ matrix allows for the estimation of correlated random parameters, which, in turn, allows capturing the interactive effects of the unobserved characteristics. Through this formulation, the entire set of diagonal and off-diagonal elements of the Cholesky matrix become estimable parameters of the model (Greene, 2016). The standard deviations of the correlated random parameters are computed using the diagonal and off-diagonal values of the covariance matrix (Fountas et al., 2018a, 2018b). The process developed by Fountas et al. (2018a) based on crash-specific parameter estimates, is used for the posterior calculation of the *t*-statistics for the standard deviations.

The correlation coefficient between two random parameters is defined as (Fountas et al., 2018a, 2018b; Jordan et al., 2019):

$$Cor(\chi_{\kappa,}\chi_{\kappa\prime}) = \frac{Cov(\chi_{\kappa},\chi_{\kappa\prime})}{\sigma_{\kappa}\sigma_{\kappa'}} \tag{4}$$

where the covariance between a pair of random parameters produced by the variables χ_{κ} and $\chi_{\kappa\prime}$, is denoted by $Cov(\chi_{\kappa,}\chi_{\kappa})$, while, σ_{κ} and $\sigma_{\kappa'}$ denote the standard deviations of their corresponding mixing distributions.

To specify the correlated random parameters model, a simulation-based Maximum Likelihood Estimation (MLE) approach was adopted. Halton sequences (Halton, 1960) were included in the MLE framework to optimize the efficiency of the numerical integrations throughout the process of the simulation. To ensure the stability of the model parameter estimates, 1,200 Halton draws have been used in the estimation process, in accordance with previous literature (Fountas et al., 2019).

The probability of each crash, i, resulting in an injury-severity outcome j, $P_i(y = j)$ is given as (Washington et al., 2020):

$$P_i(y = j) = \Phi(\mu_j - \boldsymbol{\beta}_i \mathbf{X}_i) - \Phi(\mu_{j+1} - \boldsymbol{\beta}_i \boldsymbol{X}_i) \quad (5)$$

where Φ represents the cumulative function of the standard normal distribution, and all other terms remain as defined earlier.

To identify the specific effects of the independent variables on all injury-severity levels, and particularly, on the interior levels, marginal effects are also estimated. Marginal effects demonstrate the change in the outcome probabilities caused by one-unit change of the independent variable (Washington et al., 2020). In this paper, the estimated models include only binary indicators as independent variables, therefore, the change from "0" to "1" in the value of the variables determines the marginal effects, as:

$$\frac{\Delta P_i(y=j)}{\Delta \mathbf{X}} = [\varphi(\mu_{j-1} - \boldsymbol{\beta}\mathbf{X}) - \varphi(\mu_j - \boldsymbol{\beta}\mathbf{X})]\boldsymbol{\beta} \quad (6)$$

Where φ denotes the density function of the normal distribution, and all other terms remain as defined earlier. In this study, the observation-specific coefficients ($\boldsymbol{\beta s}$) of the random parameters have been used for the calculation of the marginal effects.

EMPIRICAL SETTING

For the empirical analysis, we use data from the STATS19, the most comprehensive and publicly available crash database in the UK containing information obtained from the police crash reports (DfT, 2019). The recorded data typically encompass crash conditions and outcomes, road design and classification characteristics, pavement condition, vehicle features, casualties' characteristics (including age, gender, type of residence location and so on), and aggregate environmental conditions observed at the time of the crash (e.g., weather or lighting conditions). Injuries are classified in three different levels of severity: slight, serious and fatal. Crashes resulting in no injuries are not recorded in the dataset

(Fountas and Rye, 2019). In line with a substantial body of safety literature, the dependent variables of the models are derived from the injury outcome of the most severely injured person in the crash (i.e., cyclist for single-bicycle crashes, cyclist or motor vehicle occupant for the bicycle-motor vehicle crashes), as recorded in the police reports.

The dataset we use for the statistical analysis includes crashes occurred in Scotland, UK between 2010 and 2018[2]. Specifically, it encompasses 350 single-bicycle crashes and 6,483 bicycle-motor vehicle crashes (6,319 cases with one bicycle and one vehicle involved; 79 cases with two bicycles and one vehicle involved; 85 cases with one bicycle and two vehicles involved). For both crash categories, slight injury is the most widely observed outcome (approximately 55% for single-bicycle crashes and 81% for bicycle-motor vehicle crashes), whereas the proportions of serious and fatal injuries are considerably higher in single-bicycle crashes (41% and 4%, respectively) than in bicycle-motor vehicle crashes (18% and 1%, respectively). The differences between the proportions of injury outcomes in the two crash categories should be considered with caution, as the possibility of under-reporting in bicycle crashes is important, especially for single-bicycle crashes resulting in less severe outcomes (Schepers et al., 2015).

Table 1 shows the descriptive statistics of the key variables that were identified as statistically significant in the analysis. It should be noted that a wider range of variables were investigated throughout the process of statistical analysis; in the Appendix, Table A1 provides a comprehensive overview of

[2] Given that the crash data span over a period of nine years, we expect that the effect of the determinants of crash injury severities may change over time as a result of the temporal instability in statistical models, which has been consistently identified in recent safety research (Mannering, 2018; Shannon and Fountas, 2021). To address this issue, previous studies (e.g., Islam et al., 2020; Al-Bdairi et al., 2020) suggest to identify periods where the effect of explanatory variables are stable and estimate separate models for these periods. This approach was also considered in this study but the split of the dataset in shorter periods led to misspecification issues, such as non-convergence of some models, especially for the single-bicycle crashes where the number of observations is relatively low. In an effort to identify temporal heterogeneity, we also tried various dummy variables representing individual years or combination of years as potential explanatory factors in the models. However, none of these variables produced statistically significant parameters. We also tested these variables in the heterogeneity-in-the-means function, but they also turned out as statistically insignificant.

these variables along with their summary statistics. Figure 2 and 3 present the spatial distribution of single-bicycle and bicycle-motor vehicle crashes, respectively, per injury-severity level. Both Figures show that the vast majority of bicycle-related crashes occurred in the central belt of Scotland, where the two largest metropolitan areas of the country (Edinburgh and Glasgow) are located. These two areas are associated with high cycling rates, especially for commute and short distance (less than 5 km) trips (Cycling Scotland, 2019). The clustering of bicycle-involved crashes in urban areas is confirmed in Table 1, which shows that 68.6% of single-bicycle and 77.4% of bicycle-motor vehicle crashes occur in areas with urban characteristics.

TABLE 1. Descriptive statistics of key variables

Variable description	**Single-bicycle crashes (N=350)**		**Bicycle-motor vehicle crashes (N=6,483)**	
	Frequency	Percentage (%)	Frequency	Percentage (%)
Urban area indicator (1 if the crash occurred in an urban area, 0 otherwise)	240	68.57	5,018	77.40
Cyclist's gender indicator (1 if male, 0 otherwise)	264	75.43	-	-
Lighting conditions indicator (1 if daylight, 0 otherwise)	280	80.00	-	-
Road surface condition indicator (1 if dry, 0 otherwise)	241	68.86	4,793	73.93
Weather conditions indicator (1 if fine, 0 otherwise)	282	80.57	5,310	82.04
Day of the week indicator (1 if Sunday, 0 otherwise)	41	11.71	-	-
Day indicator (1 if weekend, 0 otherwise)	-	-	1,928	29.74
Time indicator (1 if evening peak hours, 0 otherwise)	-	-	2,069	29.43
Speed limit indicator (1 if speed limit is 30 mph, 0 otherwise)	-	-	5,119	78.96
Carriageway hazard indicator (1 if no hazard was observed on the carriageway, 0 otherwise)	-	-	6,415	98.95
Gender indicator (1 if driver's/cyclist's gender is male, 0 otherwise)	-	-	4,152	64.04

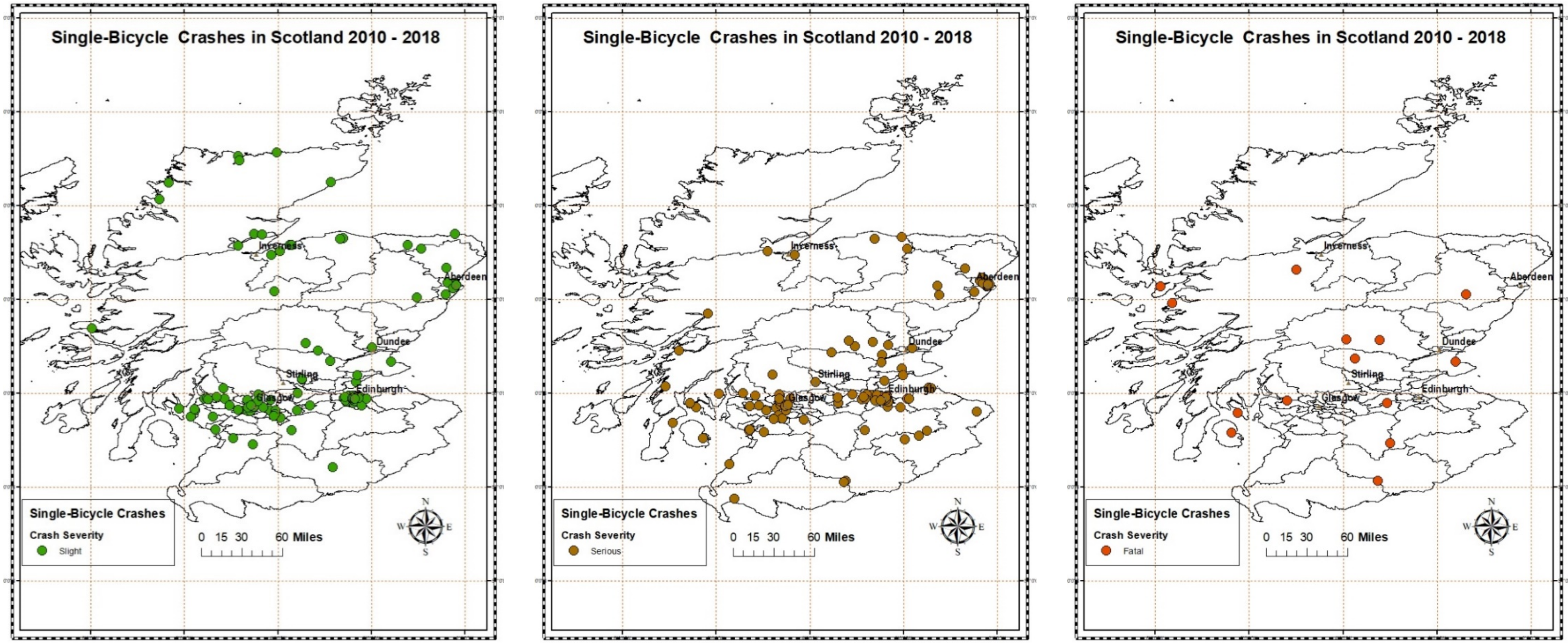

Figure 2. Spatial distribution of single-bicycle crashes per injury-severity level across Scotland

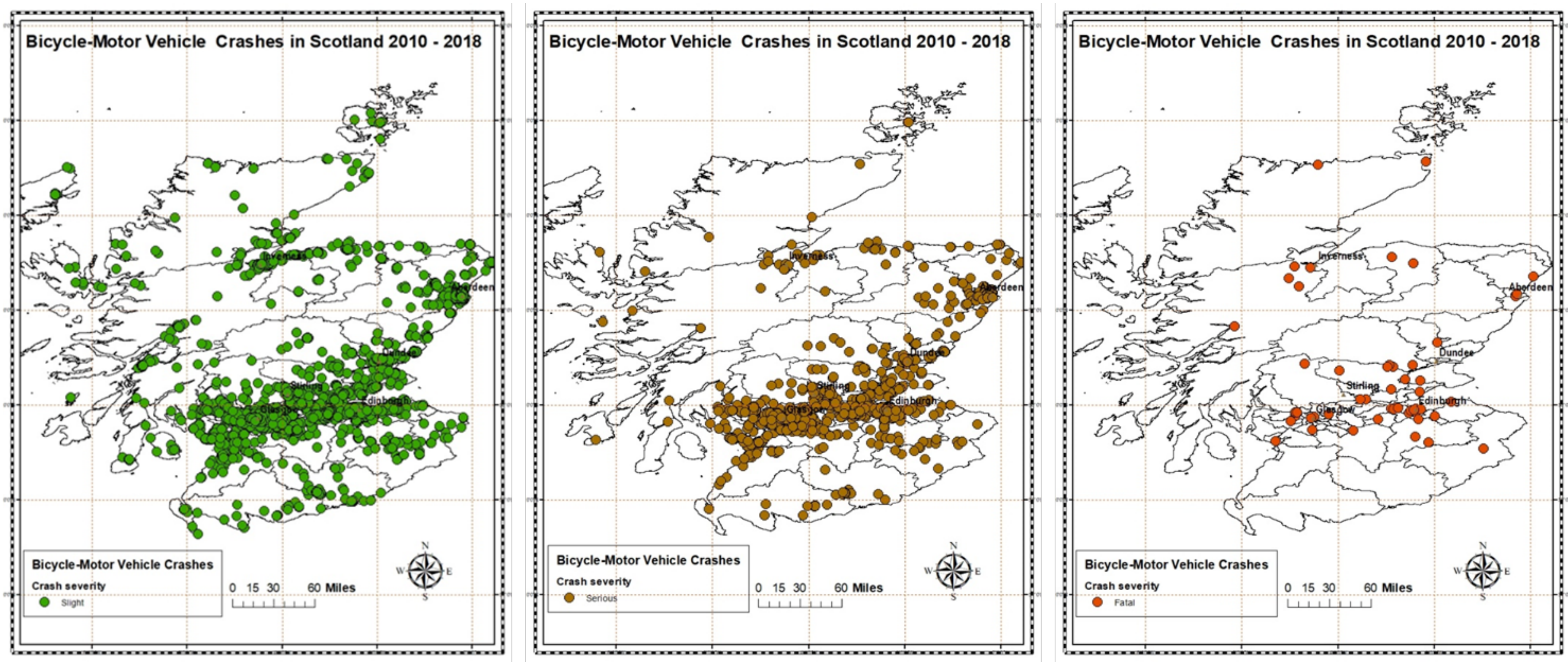

Figure 3. Spatial distribution of bicycle-motor vehicle crashes per injury-severity level across Scotland

MODEL ESTIMATION RESULTS

Overview

The results of the models for crash injury severities of the single-bicycle and bicycle-motor vehicle crashes are presented in Tables 2-7. Table 2 shows three alternative modelling approaches for the single-bicycle crashes: the fixed parameters ordered probit model (FPOP), the uncorrelated random parameters ordered probit with heterogeneity in the means (RPOPHM), and the correlated random parameters ordered probit with heterogeneity in the means (CRPOPHM).[3] Similarly, Table 4 presents the results of the three different approaches for the bicycle-motor vehicle crashes. Overall, positive parameters indicate an increase in the likelihoods of the most severe injury-severity outcome (i.e., fatal injury) and a decrease in the likelihood of the slight injury outcome. Tables 3 and 5 present the diagonal and off-diagonal elements of the Cholesky (Γ) matrix along with the correlation coefficients of the random parameters for the CRPOPHM models estimated for the single-bicycle and bicycle-motor vehicle crashes, respectively. The marginal effects for single-bicycle and bicycle-motor vehicle crashes are given in Tables 6 and 7**.** The results of the CRPOPHM models are discussed in this section, as these were identified as statistically superior to their counterparts (for further details, see the "Model Evaluation" section).

[3] To account for heterogeneity that may arise from the fixed thresholds of the ordered probit models, we have also estimated generalized ordered probit models (or hierarchical ordered probit models according to Greene, 2016), in which the thresholds vary as functions of exogenous variables (Eluru et al., 2008; Eluru and Yasmin, 2015, Fountas and Anastasopoulos, 2017). These models yielded lower statistical fit compared to their random parameter counterparts, as demonstrated by likelihood ratio tests and relevant goodness-of-fit metrics. Hence, the outputs of these models are not provided in the paper, also considering that the slightly different formulation of the generalized ordered probit models does not allow a straightforward comparison with the results from the FPOP, RPOPHM, and CRPOPHM models.

Single-bicycle Crashes

For the single-bicycle crashes, five variables were identified as statistically significant determinants of injury severities. Out of these, four have correlated random parameters, namely, the urban area indicator, the daylight indicator, the dry road surface indicator, and the fine weather indicator (see also Table 2**)**. The randomness of the parameters reveals the heterogeneous patterns of the influence of the corresponding variables on injury severities. In particular, the urban area indicator was found to reduce the likelihood of more severe injuries for more than 70% of the crash observations; for the rest of observations, this variable increases the likelihood of severe injuries. This result may pick up the favorable impact of cycling infrastructure whose presence is more evident in urban areas of Scotland. Note that previous research (Schepers and Wolt, 2012) has identified the lack of cycling infrastructure as a major source of single-bicycle crashes. This result may also reflect the impact of lower speeds of cyclists that are typically observed in urban areas as well as the familiarity of regular cyclists with the dedicated paths; experienced cyclists are more aware of potential hazards and less likely to indulge in risk-taking maneuvers. However, the urban environment may induce external stimuli (e.g., interactions with pedestrians and other cyclists; road surface defects; negotiating curbs or other roadside elements) that could result in hazardous conflicts, and potentially in critical injuries, especially for infrequent cyclists. Therefore, a group of single-bicycle crashed occurred in urban environment may be associated with serious or fatal injuries. Notably, the same variable has the strongest impact on the likelihood of all injury outcomes among the explanatory variables of the model for single-bicycle crashes, as it generates the highest marginal effects – see also Table 6 for the actual values.

Dry road surface has also a non-constant effect on injury severities, as it increases the likelihood of slight injuries for almost 70% of the single-bicycle crashes. Riding on a dry pavement restricts the risk of slipping and falling off the bicycle, which constitutes one of the main types of severe single-bicycle crashes (Schepers and Wolt, 2012; Shinar et al., 2018). The marginal effects in Table 6 show that the same variable also increases the probability of fatal injury (by 0.007), whereas it decreases the probability of the severe injury. This is an intriguing finding, as in all other cases of explanatory variables, an increase in the probability of the slight injury is accompanied by a decrease in the probabilities of both severe and fatal injuries. The association between dry road surface and fatal injuries may be linked with the risk compensating behavior of frequent cyclists, especially on Scottish roads. Due to the local environmental conditions, the surface of roads and cycling routes is often wet, thus prompting the cyclists to adjust their behavior and exercise caution in their cycling maneuvers. Dry road surface may be perceived by experienced cyclists as an additional layer of safety, which is compensated through risky maneuvers (e.g., high speeds in shared spaces, dangerous conflicts with pedestrians or other cyclists, abrupt braking) or traffic violations (e.g., red light violations) with significant potential to result in severe single-bicycle crashes.

On the contrary, fine weather conditions and daylight at the time of the crash are associated prevailingly with higher likelihood of severe injuries for the majority of cases (57.84% and 62.36%, respectively). The behavioral heterogeneity induced by favorable weather and lighting conditions could explain the variations in the effect of these environmental characteristics; a relevant and extensive discussion is provided by Fountas et al. (2020). It should be noted that these two factors represent the prevalent ambient conditions under which the vast majority of cycling trips are made. The latter is also confirmed in Table 1, where both variables reflect proportions equal or greater than 80% of the sample. This underscores the extent of heterogeneity in the effect of these factors given the fact that cycling traffic as a mean of exposure is not controlled for in the model. In addition, previous research has shown that underreporting of cases to police is a quite common pattern in single-bicycle crashes, especially under daylight (Langley et al., 2003), and that the higher is the injury severity of the crash, the more likely is to be reported (Shinar et al., 2018). Hence, to further decompose the heterogeneity stemming from these variables (and the other variables producing random parameters), the impact of other exogenous variables on the distributional effect of random parameters was explored through the heterogeneity-in-the-means structure.

Specifically, the variable indicating whether the crash occurred on Sunday or not was found to affect the means of all random parameters. The indicator of "Sunday" crashes changes the sign of the mean of the random parameter distribution for fine weather, so that on Sundays, fine weather tends to be associated with lower likelihoods of severe injuries. On Sundays, cycling is mainly preferred for leisure trips, which are generally associated with lower speed patterns. The latter in conjunction with fine weather constitute favorable conditions for low-severity injuries. The consideration of cycling trips on Sunday as a proxy of leisure or recreational trip purpose is in line with previous research (Robinson, 2006; Billot-Grasset et al., 2016). The "Sunday" indicator has the opposite effect on the mixing distribution of the dry road indicator, thus resulting in a greater percentage of observations where the dry road surface increases the likelihood of severe injuries. For the other two random parameters (i.e., the daylight indicator and the urban area indicator), the impact of the "Sunday" indicator is consistent with the sign of their means. The latter suggests that for crashes occurred on Sunday, daylight increases the percentage of cyclists sustaining more severe injuries, whereas urban conditions increase the proportion of crashes resulting in slight injuries.

Table 2 also shows that the indicator of male cyclists is a statistically significant variable exerting a uniform effect on injury severities across the crash observations. Male cyclists are more likely to be involved in a crash resulting in a serious or a fatal injury (by 0.011 and 0.00002, as shown in marginal effects in Table 6). The propensity of males to more severe injuries, which has been broadly documented in previous research (Chen and Shen, 2016; Katanalp and Eren, 2020), can be attributed to their risk-taking behavioral patterns, which are more evident in tasks related to driving or physical activities (Byrnes et al., 1999; Hollingworth et al., 2015).

TABLE 2. Model estimation results for single-bicycle crashes

Variables	*FPOP*		*RPOPHM*		*CRPOPHM*	
	Coeff.	*t-stat*	*Coeff.*	*t-stat*	*Coeff.*	*t-stat*
Characteristics (Non-random parameters)						
Cyclist's gender indicator (1 if male, 0 otherwise)	0.107	0.79	0.184	1.33	0.736	2.05
Characteristics (Random parameters)						
Urban area indicator (1 if the crash occurred in an urban area, 0 otherwise)	-0.678	-5.34	-0.754	-5.38	-4.806	-5.65
*SDPDF**			0.0045	0.05	9.112	7.94
Lighting conditions indicator (1 if daylight, 0 otherwise)	0.132	0.84	0.043	0.24	3.048	4.12
*SDPDF**			0.0162	0.21	9.677	8.42
Road surface condition indicator (1 if dry, 0 otherwise)	0.331	1.77	-0.281	-1.42	-2.692	-3.94
*SDPDF**			0.4403	5.07	5.193	36.16
Weather conditions indicator (1 if fine, 0 otherwise)	-0.143	-0.84	0.468	2.17	1.260	2.49
*SDPDF**			0.4817	5.94	6.369	51.85
Heterogeneity in means: Day of the week indicator (1 if Sunday, 0 otherwise)						
Urban area indicator	-	-	-0.389	-0.08	-2.303	-1.71
Lighting conditions indicator	-	-	1.057	2.26	7.504	4.04
Road surface condition indicator	-	-	0.647	1.16	4.794	2.84
Weather conditions indicator	-	-	-1.048	-1.80	-5.418	-2.98
Threshold parameters						
μ_{1F}	1.746	12.97	2.056	12.17	18.706	5.90
N	350		350		350	
LL (0)	-287.032		-287.032		-287.032	
LL (**β**)	-273.961		-269.005		-263.013	
Forecasting accuracy						
Overall percentage of correct predictions	53.43%		54.00%		56.29%	
Average predicted probability of observed outcome (among all injury-severity levels)	0.475		0.493		0.541	
Goodness-of-fit metrics						
Akaike Information Criterion (AIC)	559.92		568.01		568.02	
Bayesian Information Criterion (BIC)	583.07		625.88		649.04	
Likelihood ratio tests						
	CRPOPHM vs FPOP				***CRPOPHM vs RPOPHM***	
Level of confidence	>90%				>90%	
Resulting χ^2	21.896				11.984	
Degrees of freedom	14				6	
Distributional characteristics of random parameters						
			Above zero	**Below zero**	**Above zero**	**Below zero**
Urban area indicator	-	-	0%	100%	28.89%	70.11%
Lighting conditions indicator	-	-	99.60%	0.40%	62.36%	37.64%
Road surface condition indicator	-	-	26.17%	73.83%	30.21%	69.79%

Variables	***FPOP***		***RPOPHM***		***CRPOPHM***	
Weather conditions indicator	-	-	83.44%	16.56%	57.84%	42.16%

FPOP: Fixed Parameters Ordered Probit model
RPOPHM: Random Parameters Ordered Probit model with Heterogeneity in the Means
CRPOPHM: Correlated Random Parameters Ordered Probit model with Heterogeneity in the Means
**SDPDF:* Standard Deviation of Parameter Density Function

TABLE 3. Diagonal and off-diagonal matrix [*t-stats*], and correlation coefficients (in parenthesis) of random parameters for single-bicycle crashes

	Urban area indicator (1 if the crash occurred in an urban area, 0 otherwise)	Lighting conditions indicator (1 if daylight, 0 otherwise)	Road surface condition indicator (1 if dry, 0 otherwise)	Weather condition indicator (1 if fine, 0 otherwise)
Urban area indicator (1 if the crash occurred in an urban area, 0 otherwise)	9.112 [5.74] (1.000)	-	-	-
Lighting conditions indicator (1 if daylight, 0 otherwise)	9.673 [5.87] (-0.999)	0.260 [0.79] (1.000)	-	-
Road surface condition indicator (1 if dry, 0 otherwise)	-3.280 [-4.49] (-0.817)	-3.530 [-4.50] (0.806)	1.936 [3.47] (1.000)	-
Weather conditions indicator (1 if fine, 0 otherwise)	1.441 [2.53] (0.274)	-4.015 [-4.90] (-0.275)	2.078 [3.85] (0.283)	4.248 [5.75] (1.000)

Bicycle-motor vehicle crashes

Table 4 shows that four variables resulted in correlated random parameters in the CRPOPHM model for the bicycle-motor vehicle crashes: the speed limit indicator, the indicator of hazard on the carriageway, the fine weather indicator and the dry road surface indicator. Carriageways with 30 mph speed limit are associated with higher likelihood of slight injuries for the majority of the crash observations (59.76%, as shown in Table 4). The relatively low speeds observed on these carriageways may lead to bicycle-motor vehicle crashes with limited energy dissipation, and subsequently, to less severe injury severities. However, more than 40% of crashes occurred on 30 mph carriageways are more likely to result in severe casualties. This result could pick up the effect of speeding, which is widespread across motorists on roads with the same or quite close speed limits. This is also in line with previous research (Shackel and Parkin, 2014) showing that the speeds of vehicles overtaking bicycles are significantly greater when the speed limit is exceeded, especially on 30 mph carriageways. Crashes on carriageways without hazards are more likely to result in slight injuries for almost all the observations (92.65%). This is an intuitive result, as carriageways with defects such as potholes, uneven pavement or fixed obstructions increase the risk for loss of bicycle control, and potentially dangerous conflicts with other motor vehicles on the carriageway (Reynolds et al., 2009; Schepers and Wolt, 2012).

In contrast, bicycle-motor vehicle crashes occurred in fine weather are more often associated with severe injuries, as the relevant indicator increases the likelihood of serious and fatal injuries for almost 78% of the observations. Fine weather has been long linked with risk compensating behavior of motor vehicle drivers (Fountas et al., 2020a), which, in turn, results in risk-taking driving maneuvers and ultimately, in more severe crashes with bicycles. In Scotland, poor weather conditions constitute one of the main reasons discouraging people from cycling trips (Cycling Scotland, 2019); as such, the presence of fine weather could also imply a higher number of cycling trips than under poor weather. The higher number of cycling trips in fine weather may also reflect a greater proportion of trips from less skilled or inexperienced cyclists, who are generally more prone to potentially hazardous interactions with cars or other motor vehicles when sharing the circulatory space, and consequently, to serious crashes. However, no information on cycling exposure is available in our dataset, so this point requires further investigation in the future.

Crashes on dry road surface are intuitively associated with low injury severities as confirmed by the fact that the corresponding variable increases the likelihood of slight injuries by 0.08 (as shown in Table 7). As in the model for single-bicycle crashes, this variable resulted in a random parameter, which suggests its heterogeneous effect on the injury severity probabilities. In fact, in more than 72% of cases, the presence of a dry road surface increases the probability of a slight injury; however, for the remaining 28% of cases, more severe outcomes (serious or fatal injury) are more likely to be sustained. The latter proportion may unveil the effect of drivers' risk compensating behavior under seemingly favorable road conditions, which has been extensively identified in previous research (Waseem et al., 2019; Fountas et al., 2020a).

Focusing on the variables of the model yielding fixed parameters, bicycle-motor vehicle crashes in the evening peak hours (16:00 – 18:00) and in urban areas are associated with less severe injuries. During evening peak hours, the majority of cyclists are commuters who are familiarized with their everyday routes. In addition, traffic volumes are typically high during these hours and vehicle speeds are lower, as such bicycle-motor vehicle conflicts in the evening may exhibit a lower injury potential. On the contrary, the likelihoods of severe and fatal injuries are higher for crashes in the weekends, by 0.022 and 0.0002, respectively. This finding may be attributed to cases of driving or cycling under the influence of alcohol, which are widely evidenced during the weekends in Scotland (Fountas et al., 2020a). As earlier mentioned, cycling trips on weekends are usually made for leisure or recreational purposes, hence, the higher likelihood of severe injuries may be also related to the greater proportion of trips made by inexperienced or less frequent cyclists at these days.

The gender of the driver or cyclist contributes to the heterogeneity observed in the means of the correlated random parameters. Specifically, if at least one male driver/cyclist is involved in the crash, the mean of the random parameter's distribution for the weather indicator increases, leading to a larger number of crashes where the fine weather is associated with higher likelihood for serious or fatal injury. The presence of a male driver/cyclist increases the percentage of crashes in which carriageways with no observable hazard are associated with more severe injuries. This suggests that male road users are more likely to undertake driving or cycling risks when no observable hazard is present on the carriageway, which is in line with previous literature (an extensive, relevant discussion is provided in the study of Fountas et al., 2019). For the other two random parameters (speed limit indicator and road surface indicator), the heterogeneity-in-the-means variable is found to enhance the main distributional effect of the random parameters, which – in these cases – is the increase of the likelihood for slight injuries.

TABLE 4. Model estimation results for bicycle - motor vehicle crashes

Variable description	*FPOP*		*RPOPHM*		*CRPOPHM*	
	Coeff.	*t-stat*	*Coeff.*	*t-stat*	*Coeff.*	*t-stat*
Characteristics (Non-random parameter)						
Urban area indicator (1 if the crash occurred in an urban area, 0 otherwise)	-0.162	-3.54	-0.170	-3.51	-0.190	-3.71
Time indicator (1 if the crash occurred during evening peak hours, 0 otherwise)	-0.118	-2.95	-0.123	-2.91	-0.134	-3.01
Day indicator (1 if the crash occurred in the weekend, 0 otherwise)	0.093	2.11	0.106	2.29	0.117	2.39
Characteristics (Random parameter)						
Speed Limit indicator (1 if speed limit is 30 mph, 0 otherwise)	-0.393	-8.56	-0.299	-3.77	-0.126	-1.41**
*SDPDF**	-	-	0.0008	0.03	0.510	9.80
Carriageway Hazard indicator (1 if no hazard was observed on the carriageway, 0 otherwise)	-0.457	-8.09	-0.720	-7.08	-0.950	-8.54
*SDPDF**	-	-	0.0864	4.57	0.655	276.63
Weather condition indicator (1 if fine, 0 otherwise)	-0.006	-0.11	0.228	2.61	0.361	3.59
*SDPDF**	-	-	0.0026	0.12	0.474	251.96
Road surface condition indicator (1 if dry, 0 otherwise)	0.018	0.37	-0.226	-2.40	-0.377	-3.67
*SDPDF**	-	-	0.5047	22.19	0.639	154.75
Heterogeneity-in-the means variable: Gender indicator (1 if driver's/cyclist's gender is male, 0 otherwise)						
Speed Limit indicator	-	-	-0.188	-2.10	-0.262	-2.64
Carriageway Hazard indicator	-	-	0.414	3.53	0.514	4.08
Weather condition indicator	-	-	0.158	1.34	0.206	1.64
Road surface condition indicator	-	-	-0.279	-2.59	-0.336	-2.85
Threshold parameters						
μ_1	1.613	29.86	1.790	28.34	2.120	28.10
N	6483		6483		6483	
LL (0)	-3355.551		-3355.551		-3355.551	
LL (**β**)	-3273.152		-3256.993		-3251.137	
Forecasting accuracy						
Overall percentage of correct predictions	81.02%		81.80%		85.07%	
Average predicted probability of observed outcome (among all injury-severity levels)	0.696		0.730		0.778	
Goodness-of-fit metrics						
Akaike Information Criterion (AIC)	6562.30		6545.99		6546.27	

Bayesian Information Criterion (BIC)	6616.52		6654.42		6695.37	
Likelihood ratio tests						
	CRPOPHM vs FPOP				***CRPOPHM vs RPOPHM***	
Level of confidence	>90%				>90%	
Resulting χ^2	44.030				11.712	
Degrees of freedom	14				6	
Distributional characteristics of random parameters						
			Above zero	**Below zero**	**Above zero**	**Below zero**
Speed limit indicator	-	-	0%	100%	40.24%	59.76%
Carriageway hazard indicator	-	-	0%	100%	7.35%	92.65%
Weather conditions indicator	-	-	100%	0%	77.69%	22.31%
Road surface condition indicator	-	-	32.72%	67.28%	27.76%	72.24%

FPOP: Fixed Parameters Ordered Probit model
RPOPHM: Random Parameters Ordered Probit model with Heterogeneity in the Means
CRPOPHM: Correlated Random Parameters Ordered Probit model with Heterogeneity in the Means
**SDPDF:* Standard deviation of Parameter Density Function
** The mean of the random parameter for this variable is statistically insignificant, however the standard deviation is statistically significant at a greater than 95% ***level*** of confidence. To ensure that the inclusion of this random parameter significantly improves the model fit, we conducted a likelihood ratio test between the present model and a model counterpart where a fixed parameter was specified for the specific variable (speed limit indicator). The results of the test demonstrated that the inclusion of this random parameter results in statistically significant model fit improvements, at a greater than 95% level of confidence. Hence, that random parameter was kept in the final model specification.

TABLE 5. Diagonal and off-diagonal matrix [*t-stats*], and correlation coefficients (in parenthesis) of random parameters for bicycle - motor vehicle crashes

	Speed Limit indicator (1 if speed limit is 30 mph, 0 otherwise)	Carriageway Hazard indicator (1 if no hazard was observed on the carriageway, 0 otherwise)	Road surface condition indicator (1 if dry, 0 otherwise)	Weather condition indicator (1 if fine, 0 otherwise)
Speed Limit indicator (1 if speed limit is 30 mph, 0 otherwise)	0.510 [9.80] (1.000)	-	-	-
Carriageway Hazard indicator (1 if no hazard was observed on the carriageway, 0 otherwise)	-0.546 [-9.05] (-0.833)	0.362 [7.71] (1.000)	-	-
Road surface condition indicator (1 if dry, 0 otherwise)	-0.185 [-3.19] (-0.389)	-0.415 [-7.09] (-0.159)	0.137 [2.51] (1.000)	-
Weather conditions indicator (1 if fine, 0 otherwise)	-0.276 [-4.40] (-0.431)	0.320 [5.12] (0.637)	-0.418 [-7.95] (-0.459)	0.234 [10.43] (1.000)

TABLE 6. Marginal effects of the explanatory variables in the model for single-bicycle crashes

Variable description	*FPOP*			*RPOPHM*			*CRPOPHM*		
	Slight injury	Serious Injury	Fatal Injury	Slight injury	Serious injury	Fatal injury	Slight injury	Serious injury	Fatal injury
Characteristics (Non-random parameters)									
Cyclist's gender indicator (1 if male, 0 otherwise)	-0.042	0.035	0.007	-0.072	0.066	0.006	-0.01105	0.01103	0.00002
Characteristics (Random parameters)									
Urban area indicator (1 if it is urban, 0 otherwise)	0.265	-0.202	-0.063	0.294	-0.256	-0.038	0.16256	-0.13382	-0.02874
Light conditions indicator (1 if daylight, 0 otherwise)	-0.052	0.043	0.009	-0.017	0.016	0.001	-0.08116	0.04955	0.03161
Road surface condition indicator (1 if dry, 0 otherwise)	-0.128	0.108	0.02	0.111	-0.1	-0.011	0.00732	-0.01282	0.0055
Weather condition indicator (1 if fine, 0 otherwise)	-0.046	-0.011	-0.177	0.165	0.012	-0.09747	0.06609	0.03137	-0.03137

FPOP: Fixed Parameters Ordered Probit model
RPOPHM: Random Parameters Ordered Probit model with Heterogeneity in the Means
CRPOPHM: Correlated Random Parameters Ordered Probit model with Heterogeneity in the Means

TABLE 7. Marginal effects of the explanatory variables in the model for bicycle-motor vehicle crashes

Variable description	*FPOP*			*RPOPHM*			*CRPOPHM*		
	Slight injury	Serious Injury	Fatal Injury	Slight injury	Serious injury	Fatal injury	Slight injury	Serious injury	Fatal injury
Characteristics (Non-random parameters)									
Urban area (1 if it is urban, 0 otherwise)	0.045	-0.042	-0.0031	0.039	-0.037	-0.0011	0.037	-0.037	-0.00032
Time indicator (1 if the crash occurred during evening peak hours, 0 otherwise)	0.031	-0.029	-0.0019	0.026	-0.025	-0.00069	0.024	-0.024	-0.00017
Day indicator (1 if the crash occurred in the weekend, 0 otherwise)	-0.026	0.024	-0.0017	-0.024	0.023	0.00007	-0.023	0.022	0.00018
Characteristics (Random parameters)									
Speed Limit indicator (1 if speed limit is 30 mph, 0 otherwise)	0.115	-0.106	-0.0091	0.071	-0.068	-0.002	0.024	-0.024	-0.00020
Carriageway Hazard indicator (1 if no hazard was observed on the carriageway, 0 otherwise)	0.145	-0.131	-0.0139	0.214	-0.201	-0.013	0.275	-0.268	-0.00729
Road surface condition indicator (1 if dry, 0 otherwise)	-0.005	0.005	0.0003	0.053	-0.051	-0.002	0.080	-0.079	-0.001
Weather conditions indicator (1 if fine, 0 otherwise)	0.002	-0.002	-0.0001	-0.046	0.045	0.001	-0.060	0.059	0.001

FPOP: Fixed Parameters Ordered Probit model
RPOPHM: Random Parameters Ordered Probit model with Heterogeneity in the Means
CRPOPHM: Correlated Random Parameters Ordered Probit model with Heterogeneity in the Means

ESTIMATION AND INTERPRETATION OF RANDOM PARAMETER CORRELATION

The coefficients of the correlation among the random parameters are computed according to Equation 4 and are presented in Table 3 and Table 5 for the single-bicycle crashes and bicycle-motor vehicle crashes, respectively. Such coefficients reflect the interactions between the unobserved effects captured by the random parameters and differ from the traditional correlation coefficients, which measure the linear correlations between the variables (Fountas et al., 2018b; Jordan et al., 2019; Pantangi et al., 2020).

For the single-bicycle crashes, the negative correlation between the unobserved characteristics underpinning the urban area and the dry road surface indicators (-0.817) reveals their mixed effects on injury severities, i.e. the related unobserved characteristics have opposite effects on the likelihood of injury severities so that when the unobserved characteristics linked to one variable increase the likelihood of slight injury, the unobserved characteristics related to the other one tend to decrease it, and vice versa. The same happens for the daylight and the fine weather indicators as well as for the urban area and the daylight indicators. On the contrary, the positive correlations between the random parameters for the pairs urban area and fine weather, and dry road surface and fine weather (the correlation coefficients are 0.806, 0.274, and 0.283 respectively), show that the interactive effect of the unobserved characteristics captured by these variables is unidirectional, i.e., either positive or negative. Given that the correlation of the random parameters sheds light on the interactive nuances of unobserved heterogeneity, great caution should be exerted in the interpretation of the underlying relationships. Overall, we can infer that the interaction of built environment characteristics (captured by the urban area indicator) and various indicators of environmental conditions (i.e., daylight and dry road surface) are associated with mixed effects on bicycle injury severities. The result is expected as both types of factors offer a wide range of variations with significant potential to trigger heterogeneous behavioral response across the different cohorts of cyclists. For example, in the case of an experienced cyclist, the segregated cycle route provided in an urban environment may decrease the likelihood of striking fixed objects within the route, however, the risk compensating effect of ample visibility under daylight conditions may increase the likelihood of high-speed maneuvers. In the case of an inexperienced cyclist, the dry road surface may reduce the probability of skidding or sliding incidents, but the obstructions posed by other users of the shared and dense urban space (e.g., pedestrians, bicycles, other vehicles) may increase the probability of wobbling and falling off the bicycle. On the contrary, the interaction of unobserved characteristics related to different aspects of environmental conditions (i.e., fine weather, daylight and dry road surface) are found to have more consistent impact on injury severities of single-bicycle crashes. This finding may indicate that the mechanism of behavioral response to combinations of environmental conditions may share similarities across the various cohorts of cyclists. Certainly, that does not mean that the effect of these conditions does not vary across the crash cases. Indeed, the variation has been verified by the generation of random parameters by the relevant variables, not only in the present study, but also in a wide range of previous studies (Fountas and Anastasopoulos, 2017; Behnood and Mannering, 2017).

For the bicycle-motor vehicle crashes, negative correlation coefficients are observed for the random parameters corresponding to the pairs: 30 mph speed limit and no carriageway hazard, 30 mph speed limit and dry road surface, 30 mph speed limit and fine weather, dry road surface and fine weather. The interactions between the unobserved determinants of the effect of low speed limits and those of road and weather characteristics result in highly heterogeneous patterns. This finding is in line with earlier studies (Anastasopoulos and Mannering, 2016) stating that speed limit is a major source of unobserved heterogeneity with significant implications on the behavioral patterns of all road users. In fact, the 30-mph speed limit may serve as a proxy of built environment characteristics, as in Scotland, this limit is typically used for residential streets and minor or low-standard roads in built-up areas. Hence, the negative correlations of all the pairs including the 30-mph speed limit can be attributed to the disparate interactions between built environment (partially captured by the 30 mph speed limit), road conditions (captured by the no-hazard indicator) and environmental conditions (captured by fine weather and dry road surface), as in the model for single-bicycle crashes. Positive correlation is observed between the random parameters obtained by the variables reflecting no hazard on the carriageway and fine weather. Another interesting finding is that the correlation coefficient of the random parameters of dry road surface and fine weather is negative for the bicycle-motor vehicle crashes and positive for the single-bicycle crashes. Considering that the road-environment interactions have unobserved implications on the behavioral responses of road users (Fountas et al., 2020a), our models prove that such responses vary also across different crash contexts.

MODEL COMPARISON AND EVALUATION

To evaluate the statistical performance of the CRPOPHM models over their lower-order counterparts, FPOP and RPOPHM, Likelihood Ratio Tests (LRTs) were also conducted. Such tests are widely used in safety research to compare the statistical performance of nested modeling approaches. The likelihood ratio test can be defined as (Washington et al., 2020):

$$X^2 = -2[LL(\boldsymbol{\beta}_{\mathrm{c}}) - LL(\boldsymbol{\beta}_{\mathrm{m}})] \tag{7}$$

Where LL($\boldsymbol{\beta}_{\mathrm{m}}$) indicates the value of log-likelihood function for the CRPOPHM, whereas LL($\boldsymbol{\beta}_{\mathrm{c}}$) is the value of the log-likelihood function for the model counterpart in question (i.e., FPOP or RPOPHM). The test metric follows a chi-square distribution having as many degrees of freedom as the difference in the number of parameters included in the models that are evaluated. Two LRTs were conducted (CRPOPHM *vs* FPOP; CRPORHM *vs* RPOPHM) for each category of crashes. The results of the LRTs are reported in Tables 2 and 4 and suggest that the CRPOPHM models outperform statistically the FPOP and RPOPHM models. Apart from the LRTs, we also computed goodness-of-fit metrics, namely the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to further assess the statistical performance of the compared models. The AIC and BIC statistics can be defined as:

$$\text{AIC} = 2K - 2LL(\boldsymbol{\beta}) \tag{8}$$

$$\text{BIC} = -2LL(\boldsymbol{\beta}) + K\ln(N) \tag{9}$$

where, K is the number of estimable parameters in the model, N is the number of crash observations used for model estimation, and all other terms as previously defined. The AIC and BIC statistics for both single-bicycle and bicycle – motor vehicle crashes are included in Table 2 and Table 4, respectively. Lower values of these metrics generally imply better statistical fit. While AIC values are almost equal for RPOPHM and CRPOPHM models in both crash types (almost equal to 568 for single-bicycle crashes and 6546 for bicycle-motor vehicle crashes), BIC sees higher values in the CRPOPHM models (approximately 649 for single-bicycle crashes and 6695 for bicycle – motor vehicle crashes) relative to the other counterparts. In fact, BIC imposes stricter penalties on models with more estimable parameters compared to other goodness-of-fit metrics, thus favoring models with fewer parameters (Shriner and Yi, 2009). That seems to be the case with the CRPOPHM models, which feature higher numbers of estimable parameters, mainly due to the relaxation of the assumption for the off-diagonal elements of the Cholesky matrix; in fact, these elements constitute additional estimable parameters only in the CRPOPHM models, while not being present in the competing approaches (FPOP and RPOPHM). The sensitivity of correlated random parameter approaches to relative goodness-of-fit metrics that penalize models based on the number of estimable parameters has been also acknowledged in previous research (Fountas et al., 2018a; 2018b). Despite that, the proposed CRPOPHM approach offers relative merits in the specific analysis context, not only due to the previously discussed outcomes of the LRTs, but also because the competing FPOP and RPOPHM models yield several statistically insignificant variables, which may induce considerable bias on model estimation, thus hampering the robustness of their statistical outputs.

Apart from the statistical performance, another dimension of comparison stems from the marginal effects for the single-bicycle and bicycle-motor vehicle models. Focusing on the former, the strongest (in magnitude) effect originates from the variable representing urban areas, which causes the greatest increase in the likelihood for slight injuries (by 0.163) and the most pronounced decrease in the likelihood of severe injuries (by 0.134) across all the variables included in the model, as shown in Table 6. The urban area indicator is also present as an influential factor in the model of bicycle – motor vehicle crashes, however its impact is weaker – the likelihood of slight injuries increases by 0.037, whereas the corresponding likelihood for the serious injuries decreases by almost the same amount, as its impact on fatal injuries is negligible, as shown in Table 7. Other common variables across the two models encompass the dry road surface and fine weather. The latter constitutes a source of varying effects across the two models, not only because it results in random parameters in both cases. Specifically, fine weather increases the likelihood of slight injuries by 0.067 (and expectedly, reduces the likelihood for fatal injuries by 0.031) for single-bicycle crashes. Conversely, in bicycle-motor vehicle crashes, the same variable decreases the likelihood of slight injuries by 0.060 and increases the likelihood of fatal injuries by 0.001. As previously mentioned, favorable weather conditions translate into better visibility and sight distances that can enhance risk tolerance, especially for some groups of drivers. On the contrary, the propensity of cyclists to sustain slight injuries in single-bicycle crashes may be also attributed to the mitigating impact of ambient conditions in cases of hazardous situations or interactions with other road users bearing potential to result in serious single-bicycle crashes. It should be also noted that the negative effect of fine weather on fatal injuries is among the strongest (in magnitude) impacts within the single-bicycle model. Dry road surface is found to have consistent effects on slight and serious injuries in both models, increasing the likelihood of slight injuries (by 0.007 and 0.08 in single-bicycle and bicycle-motor vehicles, respectively) and decreasing the likelihood of serious injuries (by 0.013 and 0.079 in single-bicycle and bicycle-motor vehicles, respectively). The relative comparison of the marginal effect values demonstrates the more pronounced impact of this variable on the injury severities for bicycle-motor vehicle crashes. Another interesting finding stems for the observed, yet contradictory effects of the same variable on fatal injuries. Dry road surface slightly increases (by 0.006) the specific likelihood in single-bicycle crashes, whereas it does reduce this (by 0.001) in bicycle-motor vehicle crashes. The association between dry roads and fatal injuries in single-bicycle crashes may warrant further investigation in the future, as it may be related to the risk compensation exhibited by some cyclists in the presence of dry road surface, which may be critical when other road hazards (e.g., steep curves, pavement material reducing friction, uneven surface, and so on) are present on dry surfaces (Prati et al., 2017). In the same context, it should be mentioned that the strongest effect in the model for bicycle-motor vehicle crashes is exerted by the variable indicating the presence of a hazard on the road, which increases the likelihood of slight injuries by 0.275, and decreases the likelihood of serious and fatal injuries by 0.268 and 0.007, respectively. The major influence of observed roadway hazards on bicycle-related crashes has been well documented in previous research (Reynolds et al., 2009; Prati et al., 2017), being also confirmed in the current model for bicycle-motor vehicle crashes. The same variable was also tested in the model for bicycle-motor vehicle crashes, but it turned out as statistically insignificant.

The comparative evaluation of the models also extends to their forecasting accuracy. In line with previous studies based on sophisticated variants of ordered models (e.g., Fountas and Anastasopoulos, 2018; Balusu et al., 2018), we computed the overall percentage of correct predictions per model as well as the percentage of correct predictions per injury severity outcome. A prediction is considered as correct if the outcome with the highest predicted probability for a specific observation coincides with the observed outcome. For the calculation of the predicted probabilities, we used the parameter estimates of all three models. The overall percentages of correct predictions are provided in Tables 2 and 4, whereas comparative overviews for the percentages of correct predictions per injury severity outcome are illustrated in Figures 4 and 5. For both single-bicycle and bicycle-motor vehicle crashes, CROPHM models are found to offer higher overall percentages of correct predictions. Focusing on the slight injury outcome, the CRPOPHM approach tends to have similar prediction performance with the FPOP and RPOPHM for both categories of crashes. Notably, in the case of bicycle-motor vehicle crashes, the accuracy of all approaches is almost 100%. For the severe injury outcomes (serious and fatal injury), the CRPOPHM outperforms its counterparts, as it provides significantly higher percentages of correct predictions, as shown in Figures 4 and 5. In fact, the CRPOPHM is the only approach that results in correct predictions at the highest injury level, even though these predictions are quite limited. The low number of correct predictions at the highest level can be attributed to the low number of fatal injuries in the dataset as well as to the lack of information for potentially significant variables, such as cycling exposure or behavioral characteristics of drivers or cyclists.

Beyond the percentage of correct predictions per injury severity outcome, we expand the evaluation of the forecasting accuracy for the estimated models by computing the predicted probabilities of the observed outcomes. This metric has been widely used in previous research (e.g., Yasmin et al., 2014; Fountas and Anastasopoulos, 2018) due to its potential to show how closely the model-predicted outcome aligns with the observed outcome for each crash observation. To calculate this metric, we use the predicted probability of the observed injury-severity outcome for each crash record. For example, if the observed outcome of a specific crash is serious injury, this metric is informed by the model-predicted probability of the serious injury outcome for the same crash. The closer to 1 is the predicted probability of the observed outcome, the higher is the forecasting accuracy offered by the model for the crash outcome. These probabilities are provided at both aggregate and disaggregate level. In particular, the average predicted probabilities of the observed outcome, aggregated among all injury-severity levels, are shown in Table 2 and 4, for single-bicycle and bicycle-motor vehicle crashes, respectively. The predicted probabilities of the observed outcome, disaggregated by injury-severity level for single-bicycle and bicycle-motor vehicle crashes are graphically illustrated in Figure 6 and 7, respectively. To provide an example on how the information displayed in these Figures can be interpreted, Figure 6 shows that the CRPOPHM model, for all the single-bicycle crash cases that resulted in slight injuries (i.e., slight injury is the observed outcome), offers an average predicted probability of slight injury equal to 0.71 approximately; the FPOP and RPOPHM counterparts yield lower predicted probabilities, equal to 0.56 and 0.58, respectively. Overall, the CRPOPHM approach is found to offer the highest predicted probabilities of the observed outcome across all competing approaches, at both aggregate (average probabilities across all injury severities) and disaggregate (average predicted probabilities by injury severity) level. That provides further evidence on the robust forecasting accuracy performance that can be achieved with the simultaneous consideration of correlated random parameters and heterogeneity in the means, as featured by the CRPOPHM approach (Ahmed et al., 2021).

Overall, the CRPOPHM approach yields significantly higher proportions of correct predictions and predicted probabilities of observed outcomes in bicycle-motor vehicle crashes relative to single-bicycle crashes. This could be attributed to the smaller sample size of the latter as well as to the lack of information related to cycling infrastructure (e.g., cycle lanes/segregated paths), cycling exposure and use of passive safety devices (e.g., helmet), which may be influential in determining the severity of single-bicycle crashes (Myhrmann et al., 2020).

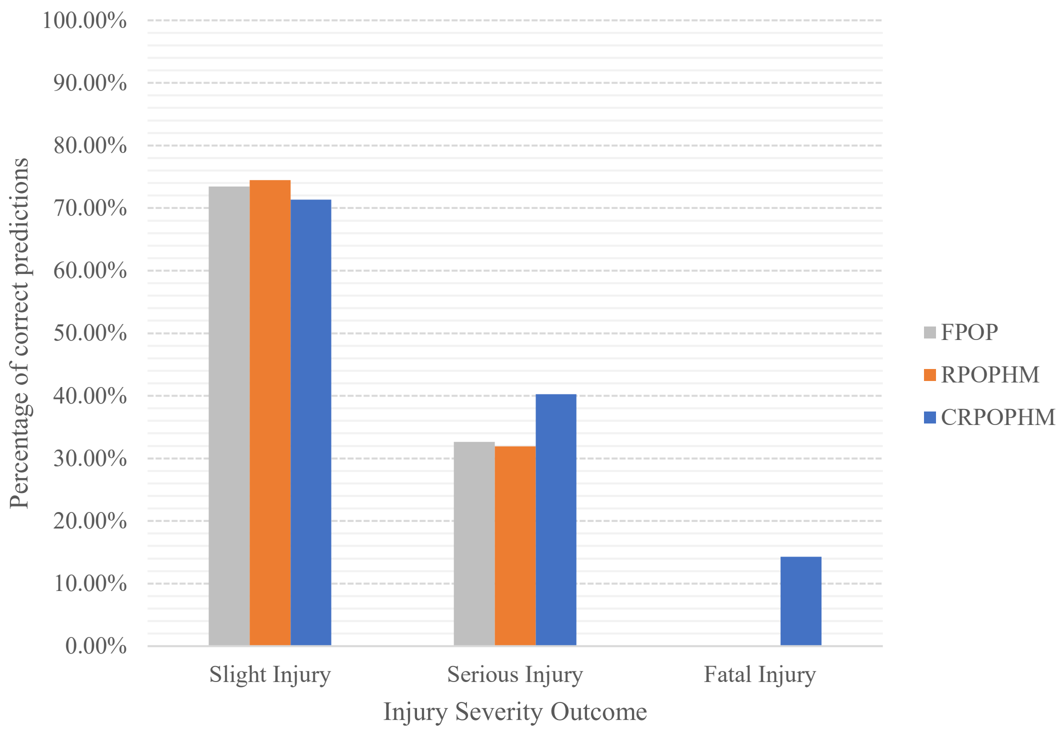

Figure 4. Percentage of correct predictions per injury severity outcome for single-bicycle crashes

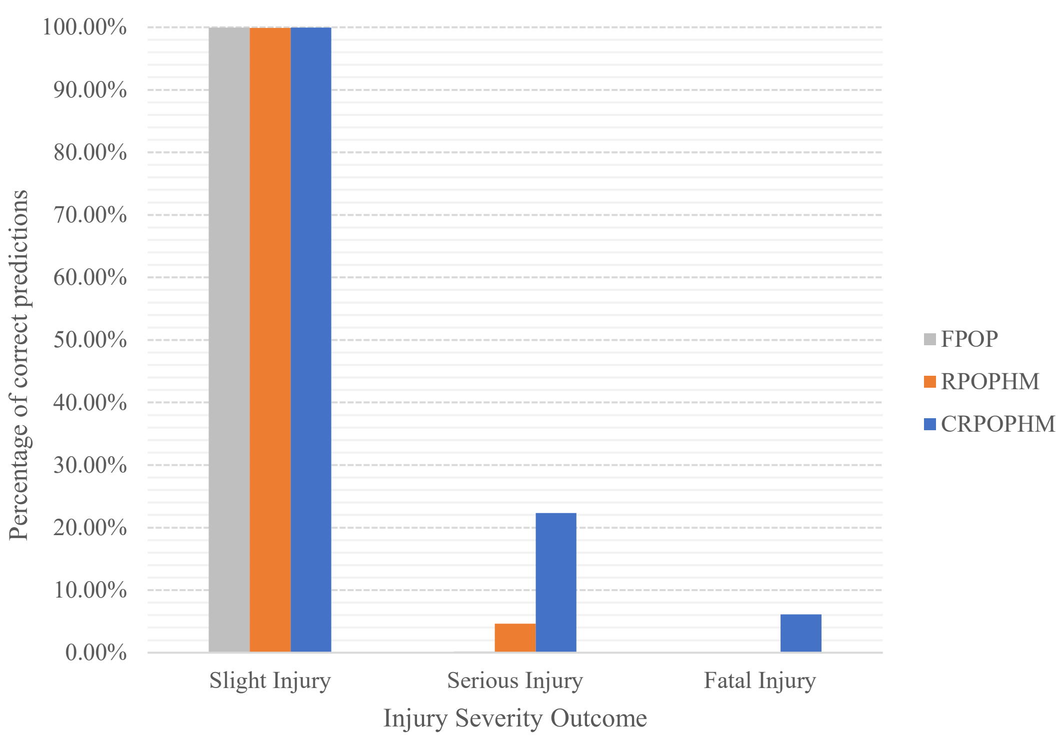

Figure 5. Percentage of correct predictions per injury severity outcome for bicycle-motor vehicle crashes

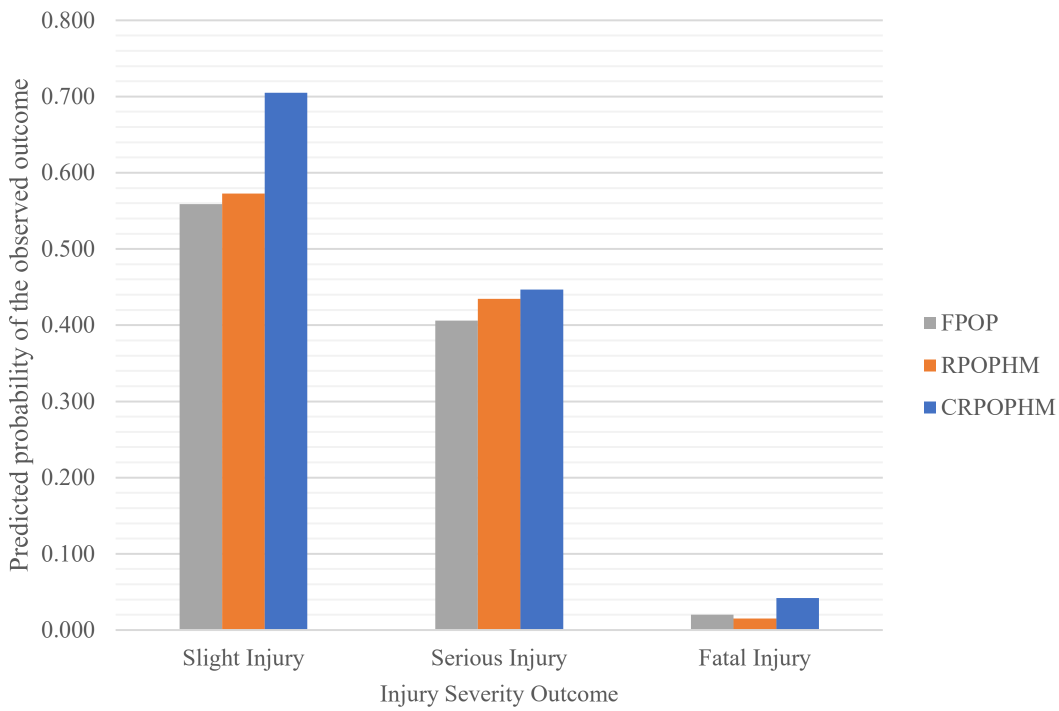

Figure 6. Predicted probability of the observed outcome by injury severity level for single-bicycle crashes

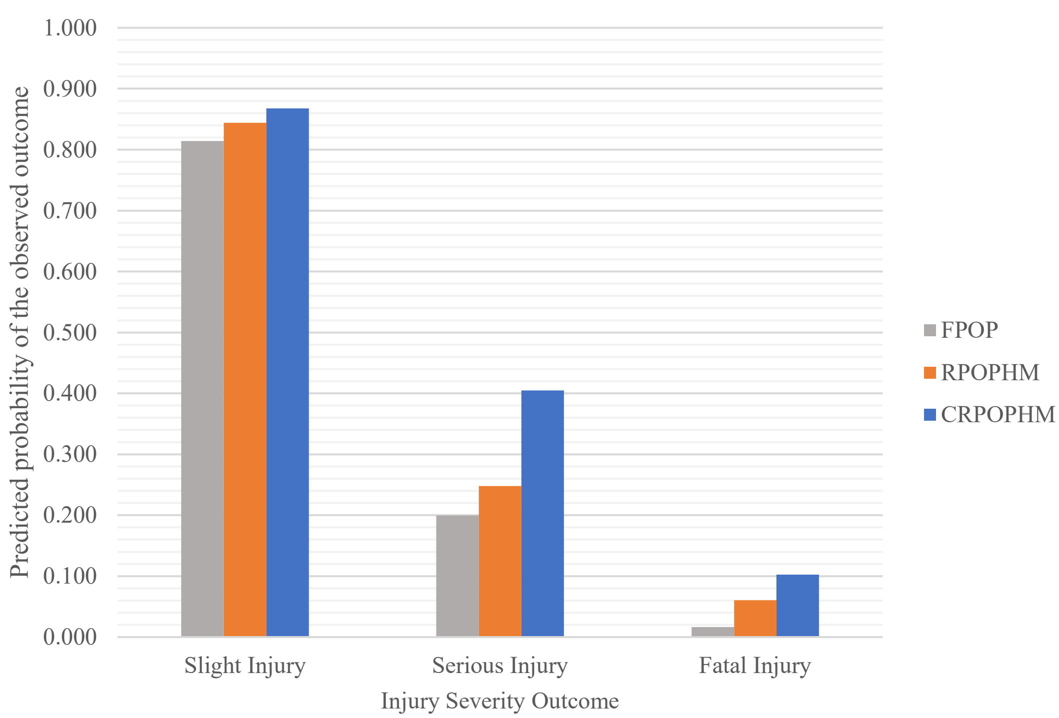

Figure 7. Predicted probability of the observed outcome by injury severity level for bicycle-motor vehicle crashes

SUMMARY AND CONCLUSIONS

This study provides new insights into the factors affecting the injury severities of crashes involving bicyclists by using a correlated random parameters ordered probit approach with heterogeneity in the means for the analysis of the crash data. The approach accounts for the impact of unobserved heterogeneity acknowledging that the sources of unobserved heterogeneity are not independent. The correlation among random parameters shed light on the interactive effects of the unobserved factors, whereas the heterogeneity-in-the-means structure captures the impact of unobserved factors on the distributional characteristics of the random parameters. The empirical analysis focused on two major groups of bicycle-involved crashes, the single-bicycle crashes and the bicycle-motor vehicle crashes.

Upon extensive testing of a broad range of explanatory variables, the results of the analysis showed that fine weather and dry road surface are significant determinants of injury severities for both single-bicycle and bicycle-motor vehicle crashes. However, their effects on injury severities exhibit heterogeneous patterns, as they resulted in correlated random parameters. Speed limit and presence of carriageway hazard (for bicycle-motor vehicle crashes) as well as daylight and urban locations (for single-bicycle crashes) were also found to have heterogeneous impacts on injury severities. Crashes occurred on Sunday and male drivers or cyclists were observed to affect the means of the random parameters' distributions for the single-bicycle and bicycle-motor vehicle crashes, respectively. The correlation coefficients between the random parameters also unmasked interactive effects of the unobserved characteristics on injury severities, which could not be identified through the traditional random parameters modeling. Overall, the identification of different determinants as well as the observed variations in the effect of the common determinants between single-bicycle and bicycle-motor vehicle crashes show that their injury-generation mechanisms differ between the two types of crashes.

The methodological merits of the correlated random parameters ordered probit approach with heterogeneity in the means were also assessed. The comparison with less sophisticated approaches (i.e., fixed parameters ordered probit and random parameters ordered probit with heterogeneity in the means) proved that the use of a more generalized formulation for the random parameters, despite computationally demanding, can shed more light on the unobserved interdependencies underpinning the analysis of bicycle injury-severity data, thus resulting in more accurate and robust statistical inferences.

Even though the applied methodological framework addresses various layers of unobserved heterogeneity, the potential presence of temporal heterogeneity in the data (as they span over a 9-year period) – which may cause temporal instabilities in the identified effects – is not considered. This constitutes a limitation of the study. In addition, the presented modeling framework does not explicitly account for threshold heterogeneity, as the thresholds of the ordered probit models are specified as fixed parameters. Future extensions of the modeling framework may endeavor to provide a more generalized formulation of the model allowing the thresholds to vary as functions of exogenous variables, in a similar manner with the mixed generalized ordered or the random threshold random parameter hierarchical ordered models (Eluru et al., 2008; Eluru and Yasmin, 2015; Fountas and Anastasopoulos, 2017; Yu et al., 2021). Despite the latter, the findings of this study could establish the basis for more disaggregate analyses of bicycle-involved crashes and their underlying variations in the future using more granular data including richer information about cycling exposure, and human factors. This is particularly important for single-bicycle crashes, whose generation mechanism has not been thoroughly explored to date.

ACKNOWLEDGMENTS

This research was sponsored by the Exchequer Research Fund of the School of Engineering and the Built Environment at Edinburgh Napier University.

REFERENCES

Ahmed, S., Pantangi, S., Eker, U., Fountas, G., Still, S., Anastasopoulos, P., 2020. Analysis of safety benefits and security concerns from the use of autonomous vehicles: A grouped random parameters bivariate probit approach with heterogeneity in means. *Analytic Methods in Accident Research*, 28, 100134.

Ahmed, S.S., Cohen, J., Anastasopoulos, P.C., 2021. A correlated random parameters with heterogeneity in means approach of deer-vehicle collisions and resulting injury-severities. *Analytic Methods in Accident Research*, 100160.

Al-Bdairi, N., Behnood, A., Hernandez, S., 2020. Temporal stability of driver injury severities in animal-vehicle collisions: A random parameters with heterogeneity in means (and variances) approach. *Analytic Methods in Accident Research*, 26, 100120.

Anastasopoulos, P., Mannering, F., 2016. The effect of speed limits on drivers' choice of speed: a random parameters seemingly unrelated equations approach. *Analytic Methods in Accident Research*, 10:1-11.

Balusu, S., Pinjari, A., Mannering, F., Eluru, N., 2018. Non-decreasing threshold variances in mixed generalized ordered response models: A negative correlations approach to variance reduction. *Analytic Methods in Accident Research*, 20, 46-67.

Behnood, A., Mannering, F., 2017. Determinants of bicyclist injury severities in bicycle-vehicle crashes: a random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research*, 16, 35-47.

Billot-Grasset, A., Amoros, E., Hours, M., 2016. How cyclist behavior affects bicycle accident configurations?. *Transportation research part F*, 41, 261-276.

Bogue, S., Paleti, R., Balan, L., 2017. A Modified Rank Ordered Logit model to analyze injury severity of occupants in multivehicle crashes. *Analytic Methods in Accident Research*, 14, 22-40.

Boufous, S., de Rome, L., Senserrick, T., Ivers, R. Q., 2013. Single-versus multi-vehicle bicycle road crashes in Victoria, Australia. *Injury prevention*, 19(5), 358-362.

Byrnes, J.P., Miller, D.C., Schafer, W.D., 1999. Gender differences in risk taking: a meta-analysis. *Psychological bulletin*, 125(3), 367.

Cai, Q., Lee, J., Eluru, N., Abdel-Aty, M., 2016. Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis and Prevention*, 93, 14-22.

Chen, P., Shen, Q., 2016. Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accident Analysis and Prevention*, *86*, 239-246.

Cycling Scotland, 2018. Annual Cycling Monitoring Report. Glasgow, UK.

Cycling Scotland, 2019. Annual Cycling Monitoring Report. Glasgow, UK.

Department for Transport, 2019. Road Safety Data. Available from: https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

Eker, U., Ahmed, S. S., Fountas, G., Anastasopoulos, P. C., 2019. An exploratory investigation of public perceptions towards safety and security from the future use of flying cars in the United States. *Analytic Methods in Accident Research*, 23, 100103.

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 and Prevention*, 40(3), 1033-1054.

Eluru, N., Yasmin, S., 2015. A note on generalized ordered outcome models. *Analytic Methods in Accident Research*, 8, 1-6.

Fanyu, M., Sze, N.N., Cancan, S., Tiantian, C., Yiping, Z., 2021. Temporal instability of truck volume composition on non-truck-involved crash severity using uncorrelated and correlated grouped random parameters binary logit models with space-time variations. *Analytic Methods in Accident Research*, 100168.

Fountas, G., Anastasopoulos, P., 2017. A random thresholds random parameters hierarchical ordered probit analysis of highway accident injury-severities. *Analytic Methods in Accident Research*, 15, 1-16.

Fountas, G., Anastasopoulos, P., 2018.Analysis of accident injury-severity outcomes: The zero-inflated hierarchical ordered probit model with correlated disturbances. *Analytic Methods in Accident Research*, 20, 30-45.

Fountas, G., Sarwar, M. T., Anastasopoulos, P. C., Blatt, A., Majka, K., 2018a. Analysis of stationary and dynamic factors affecting highway accident occurrence: a dynamic correlated grouped random parameters binary logit approach. *Accident Analysis and Prevention*, 113, 330-340.

Fountas, G., Anastasopoulos, P. C., Abdel-Aty, M., 2018b. Analysis of accident injury-severities using a correlated random parameters ordered probit approach with time variant covariates. *Analytic Methods in Accident Research*, 18, 57-68.

Fountas, G., Fonzone, A., Gharavi, N., Rye, T., 2020a. The joint effect of weather and lighting conditions on injury severities of single-vehicle crashes. *Analytic Methods in Accident Research*, 100124.

Fountas, G., Pantangi, S. S., Hulme, K. F., Anastasopoulos, P. C., 2019. The effects of driver fatigue, gender, and distracted driving on perceived and observed aggressive driving behavior: A correlated grouped random parameters bivariate probit approach. *Analytic Methods in Accident Research*, *22*, 100091.

Fountas, G., Rye, T., 2019. A note on accounting for underlying injury-severity states in statistical modeling of injury accident data. *Procedia Computer Science*, 151, 202-209.

Fountas, G., Sun, Y.Y., Akizu-Gardoki, O., Pomponi, F., 2020b. How do people move around? National data on transport modal shares for 131 countries. *World*, *1*(1), 34-43.

Greene W. LIMDEP Version 11.0. NY: Econometric Software, Inc; 2016.

Halton, J. H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, *2*(1), 84-90.

Hamed, M. M., Al-Eideh, B. M., 2020. An exploratory analysis of traffic accidents and vehicle ownership decisions using a random parameters logit model with heterogeneity in means. *Analytic Methods in Accident Research*, *25*, 100116.

Hollingworth, M.A., Harper, A.J., Hamer, M., 2015. Risk factors for cycling accident related injury: The UK Cycling for Health Survey. *Journal of Transport and Health*, *2*(2), 189-194.

Intini, P., Berloco, N., Fonzone, A., Fountas, G., Ranieri, V., 2020. The influence of traffic, geometric and context variables on urban crash types: A grouped random parameter multinomial logit approach. *Analytic Methods in Accident Research*, *28*, 100141.

Islam, M., Alnawmasi, N., Mannering, F., 2020. Unobserved Heterogeneity and Temporal Instability in the Analysis of Work-Zone Crash-Injury Severities. *Analytic Methods in Accident Research*, 100130.

Jordan, G. A., Anastasopoulos, P. C., Peeta, S., Somenahalli, S., Rogerson, P. A., 2019. Identifying elderly travel time disparities using a correlated grouped random parameters hazard-based duration approach. *Research in Transportation Business and Management*, 30, 100369.

Katanalp, B.Y., Eren, E., 2020. The novel approaches to classify cyclist accident injury-severity: hybrid fuzzy decision mechanisms. *Accident Analysis and Prevention*, *144*, 105590.

Kim, J. K., Kim, S., Ulfarsson, G. F., Porrello, L. A., 2007. Bicyclist injury severities in bicycle–motor vehicle accidents. *Accident Analysis and Prevention*, 39(2), 238-251.

Langley, J. D., Dow, N., Stephenson, S., Kypri, K., 2003. Missing cyclists. *Injury prevention*, 9(4), 376-379.

Mannering, F., 2018. Temporal instability and the analysis of highway accident data. *Analytic Methods in Accident Research*, 17, 1-13.

Mannering, F., Bhat, C., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. *Analytic Methods in Accident Research*, 25, 100113.

Mannering, F., Shankar, V., Bhat, C., 2016. Unobserved heterogeneity and the statistical analysis of highway accident data. *Analytic Methods in Accident Research*, 11, 1-16.

Marcoux, R., Yasmin, S., Eluru, N., Rahman, M., 2018. Evaluating temporal variability of exogenous variable impacts over 25 years: an application of scaled generalized ordered logit model for driver injury severity. *Analytic Methods in Accident Research*, 20, 15-29.

Marshall, W., Ferenchak, N., 2019. Why cities with high bicycling rates are safer for all road users. *Journal of Transport and Health*, 13, 100539.

Myhrmann, M.S., Janstrup, K.H., Møller, M., Mabit, S.E., 2020. Factors influencing the injury severity of single-bicycle crashes. *Accident Analysis and Prevention*, 149, 105875.

Nashad, T., Yasmin, S., Eluru, N., Lee, J., Abdel-Aty, M. A., 2016. Joint modeling of pedestrian and bicycle crashes: copula-based approach. *Transportation Research Record*, 2601, 119-127.

Pantangi, S. S., Fountas, G., Anastasopoulos, P., Pierowicz, J., Majka, K., Blatt, A., 2020. Do high visibility enforcement programs affect aggressive driving behavior? An empirical analysis using naturalistic driving study data. *Accident Analysis and Prevention*, 138, 105361.

Pantangi, S.S., Ahmed, S.S., Fountas, G., Majka, K., Anastasopoulos, P.C., 2021. Do high visibility crosswalks improve pedestrian safety? A correlated grouped random parameters approach using naturalistic driving study data. *Analytic Methods in Accident Research*, *30*, 100155.

Poulos, R. G., Hatfield, J., Rissel, C., Flack, L. K., Murphy, S., Grzebieta, R., McIntosh, A. S., 2015. An exposure based study of crash and injury rates in a cohort of transport and recreational cyclists in New South Wales, Australia. *Accident Analysis and Prevention*, 78, 29-38.

Prati, G., De Angelis, M., Marín Puchades, V., Fraboni, F., Pietrantoni, L., 2017. Characteristics of cyclist crashes in Italy using latent class analysis and association rule mining. *PLoS one*, 12(2), 0171484.

Reynolds, C.C., Harris, M.A., Teschke, K., Cripton, P.A., Winters, M., 2009. The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. *Environmental health*, 8(1), 47.

Robinson, D.L., 2006. No clear evidence from countries that have enforced the wearing of helmets. *Bmj*, *332*(7543), 722-725.

Schepers, P., Agerholm, N., Amoros, E., Benington, R., Bjørnskau, T., Dhondt, S., ..., Niska, A., 2015. An international review of the frequency of single-bicycle crashes (SBCs) and their relation to bicycle modal share. *Injury prevention*, 21, e138-e143.

Schepers, P., Wolt, K., 2012. Single-bicycle crash types and characteristics. *Cycling Research International*, *2*(1), 119-135.

Shackel, S. C., Parkin, J., 2014. Influence of road markings, lane widths and driver behaviour on proximity and speed of vehicles overtaking cyclists. *Accident Analysis and Prevention*, 73, 100-108.

Shannon, D., Fountas, G., 2021. Extending the Heston Model to Forecast Motor Vehicle Collision Rates. *Accident Analysis and Prevention,* 159, 106250.

Shinar, D., Valero-Mora, P., van Strijp-Houtenbos, M., Haworth, N., Schramm, A., De Bruyne, G., ..., Fyhri, A., 2018. Under-reporting bicycle accidents to police in the COST TU1101 international survey: Cross-country comparisons and associated factors. *Accident Analysis and Prevention*, 110, 177-186.

Shriner, D., Yi, N., 2009. Deviance information criterion (DIC) in Bayesian multiple QTL mapping. *Computational statistics and data analysis*, 53(5), 1850-1860.

Smith L, Chowdhury S, Hammond J, Kaminski A, Wallbank C., 2019. Healthy mobility and road safety Report. Project Report, PPR865, TRL, UK.

Transport Scotland, 2018a. Scottish transport statistics. No. 36, 2017 Edition. Edinburgh, UK.

Transport Scotland, 2018b. Key Reported Road Casualties Scotland 2017. Edinburgh, UK.

Utriainen R., 2020. Characteristics of commuters' single-bicycle crashes in insurance data. *Safety*, 6(1).

Wang, C., Lu, L., Lu, J., 2015. Statistical analysis of bicyclists' injury severity at unsignalized intersections. *Traffic injury prevention*, 16(5), 507-512.

Waseem, M., Ahmed, A., Saeed, T.U., 2019. Factors affecting motorcyclists' injury severities: An empirical assessment using random parameters logit model with heterogeneity in means and variances. *Accident Analysis and Prevention*, 123, 12-19.

Washington, S., Karlaftis, M. G., Mannering, F., Anastasopoulos, P., 2020. *Statistical and econometric methods for transportation data analysis*. CRC press.

Whyte B, Waugh C., 2015. Trends in pedestrian and cyclist road casualties in Scotland. Glasgow, UK.

Yan, X., He, J., Zhang, C., Liu, Z., Wang, C., Qiao, B., 2021. Temporal analysis of crash severities involving male and female drivers: A random parameters approach with heterogeneity in means and variances. *Analytic Methods in Accident Research*, 30, 100161.

Yasmin S, Eluru N., 2013. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accident Analysis and Prevention*, 59, 506–21.

Yasmin, S., Eluru, N., Bhat, C.R., Tay, R., 2014. 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., Pinjari, A., 2015. Pooling data from fatality analysis reporting system (FARS) and generalized estimates system (GES) to explore the continuum of injury severity spectrum. *Accident Analysis and Prevention*, 84, 112-127.

Young M, Whyte B, 2020. Cycling in Scotland: a review of cycling casualties, near misses and under-reporting. Available from: www.gcph.co.uk

Yu, M., Ma, C.. Shen, J., 2021. Temporal stability of driver injury severity in single-vehicle roadway departure crashes: a random thresholds random parameters hierarchical ordered probit approach. *Analytic Methods in Accident Research*, 29, 100144.

Zheng, L., Sayed, T., Mannering, F., 2020. Modeling traffic conflicts for use in road safety analysis: A review of analytic methods and future directions. *Analytic Methods in Accident Research*, 100142.

APPENDIX

TABLE A1. Comprehensive overview of the variables used for the statistical analysis.

Variable	Description	Outcomes	Single-bicycle crashes (%)	Bicycle - motor vehicle crashes (%)
Injury severity	Injury outcome of the most severely injured person in the crash	Slight injury Serious injury Fatal injury	54.86 41.14 4.00	81.03 18.22 0.75
Demographic generation of Cyclist or Motor-vehicle occupant	Demographic generation of cyclist and/or motor-vehicle occupant involved in the crash	Generation Z Millennials Generation X Baby Boomers The Silent Generation The Greatest Generation	16.88 27.27 39.61 12.31 0.65 -	20.15 30.60 26.90 16.99 5.15 0.03
Gender of Cyclist or Motor-vehicle occupant	Gender of cyclist and/or motor vehicle occupants involved in the crash	Male Female	81.17 18.83	70.97 29.03
Day of the Week	Day that the crash occurred	Sunday Monday Tuesday Wednesday Thursday Friday Saturday	11.71 16.29 13.43 15.71 14.86 11.14 14.86	8.91 14.88 16.90 17.14 16.55 15.16 10.52
Weekday *vs* Weekend	Type of day when the crash occurred	Weekday Weekend	72.57 26.86	80.56 19.44
Speed Limit	The speed limit on the road where the crash occurred	20 mph 30 mph 40 mph 50 mph 60 mph 70 mph	9.14 66.00 3.43 0.00 19.14 00.57	5.92 77.85 4.68 1.11 9.78 0.64
Time	Peak or off-peak period of traffic when the crash occurred	Morning Peak Morning Off-peak Evening Peak Evening Off-peak Night Off-peak	19.71 35.71 24.86 13.43 4.86	21.21 32.14 29.44 13.01 4.20
Weather Condition	The weather condition at the time when the crash occurred	Fine no high winds Raining no high winds Snowing no high winds Fine + high winds Raining + high winds Snowing + high winds Fog or mist Other Unknown	80.75 10.06 0.00 0.72 0.86 0.00 0.57 1.72 1.72	82.27 10.58 0.14 1.25 1.47 0.03 0.24 1.31 2.70

Road Surface Condition	The surface condition of the road at the time when the crashes occurred	Dry Wet or damp Snow Frost or ice Flood over 3cm. deep Oil or diesel Mud	69.77 27.91 0 2.32 0 0 0	74.34 24.55 0.14 0.87 0.10 0 0
Lighting Condition	Whether it is bright or dark at the time when the crash occurred	Daylight Darkness - lights lit Darkness - lights unlit Darkness - no lighting Darkness - lighting unknown	79.71 16.29 00.86 01.43 00.29	80.85 16.14 0.95 1.45 0.65
Urban/Rural	Whether the crash occurred in urban or rural area	Urban Rural	68.87 31.13	77.79 22.21
Special condition at site	Whether road furniture and conditions are in order when the crash occurred	None Auto traffic signal - out Auto signal partially defective Road sign defective Roadworks Road surface defective Oil or diesel Mud	91.71 0 0 0 0.29 4.57 1.14 0.29	98.52 0.24 0 0.01 0.50 0.21 0.06 0.03
Skidding and Overturning	How the vehicle acted when the crash occurred	None Skidded Skidded and overturned Jackknifed Jackknifed and overturned Overturned	84.10 10.40 1.45 0 0 4.05	96.90 1.88 0.29 0.01 0 0.93
Hit Object in Carriageway	Whether the vehicles hit any object in carriageway when the crash occurred	None Previous crash Road works Parked vehicle Bridge (roof) Bridge (side) Bollard or refuge Open door of vehicle Central island of roundabout Curb Other object Any animal (except ridden horse)	90.67 0 0.29 0 0 0 0.29 0 0 1.46 5.54 1.75	97.68 0 0 1.54 0 0 0.01 0.20 0 28.49 28.49
Vehicle Leaving Carriageway	Whether the vehicle leaves the carriageway when the crash occurred	Did not leave carriageway Nearside Nearside and rebounded Straight ahead at junction	89.70 6.00 0.60 0	98.02 1.27 0.06 0.07

		Offside on to central reservation Offside on to central reservation + rebounded Offside - crossed central reservation Offside Offside and rebounded	0 0 0.30 3.40 0	0 0 0 0.50 0.09
Hit Object Off Carriageway	Whether the vehicles hit any object off carriageway when the crash occurred	None Road sign or traffic signal Lamp post Telegraph or electricity pole Tree Bus stop or bus shelter Central crash barrier Near/Offside crash barrier Submerged in water Entered ditch Other permanent object Wall or fence	93.7 0.6 0.6 0.3 1.1 0 0 0.3 0 0.9 2.2 0.3	99.45 0.10 0.01 0 0.06 0 0 0 0 0.09 0.20 0.09
Carriageway Hazards	Whether there is any object in carriageway, which can be hazardous, when the crash occurred	None Vehicle load on road Other object on road Previous crash Dog on road Other animal on road Pedestrian in carriageway - not injured Any animal in carriageway (except ridden horse)	90.3 0 7.2 0 0 0 1.1 1.4	99.06 0.13 0.67 0.03 0 0 0.07 0.04