Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: MICHALAKI, P. ... et al., 2015. Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model. Journal of Safety Research, 55, pp.89-97.

Additional Information:

- This is an open access article published by Elsevier under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Metadata Record: https://dspace.lboro.ac.uk/2134/20122

Version: Published

Publisher: © The Authors. National Safety Council and Elsevier Ltd

Rights: This work is made available according to the conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. Full details of this licence are available at: http://creativecommons.org/licenses/by/4.0/

Please cite the published version.
Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model

Paraskevi Michalaki, a, *, Mohammed A. Quddus, a David Pitfield, a Andrew Huetson b

a School of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire LE11 3TU, United Kingdom
b Connect Roads — Balfour Beatty, Bramham Maintenance Compound, Spenn Common Lane, Tadcaster LS24 9NS, United Kingdom

Abstract

Problem: The severity of motorway accidents that occurred on the hard shoulder (HS) is higher than for the main carriageway (MC). This paper compares and contrasts the most important factors affecting the severity of HS and MC accidents on motorways in England. Method: Using police reported accident data, the accidents that occurred on motorways in England are grouped into two categories (i.e., HS and MC) according to the location. A generalised ordered logistic regression model is then applied to identify the factors affecting the severity of HS and MC accidents on motorways. The factors examined include accident and vehicle characteristics, traffic and environment conditions, as well as other behavioral factors. Results: Results suggest that the factors positively affecting the severity include: number of vehicles involved in the accident, peak-hour traffic time, and low visibility. Differences between HS and MC accidents are identified, with the most important being the involvement of heavy goods vehicles (HGVs) and driver fatigue, which are found to be more crucial in increasing the severity of HS accidents. Practical applications: Measures to increase awareness of HGV drivers regarding the risk of fatigue when driving on motorways, and especially the nearside lane, should be taken by the stakeholders.

© 2015 The Authors. National Safety Council and Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The hard-shoulder of a motorway – also called shoulder or emergency lane – can only be used either by the public for emergency reasons or by public agencies and private companies that are responsible for the maintenance and operation of the road. In many past campaigns aimed at both the public and the operators, it has been highlighted that the hard-shoulder is a hazardous place either to stop or drive (SURVIVE Group, 2006); however, this has not been investigated through rigorous research. Significant research has been conducted in the area of road safety, and especially motorways. Many studies have focused on the severity of motorway accidents since historical data suggest it is higher than occurs generally on the rest of the road network (SURVIVE Group, 2006).

In Great Britain, the severity of accidents has three categories: slight, serious and fatal injury. Motorway accidents can be divided into two groups according to the location where the initial impact happened: hard-shoulder (HS) and main carriageway (MC). When comparing the accident severity of these two groups in Great Britain, it can be concluded that the severity of HS accidents is significantly higher. More specifically, the percentage of accidents on the HS is almost five times than on the MC and the proportion of serious injury accidents is also noticeably higher on the HS. However in the past, no significant research has been conducted on the factors affecting the severity of hard-shoulder accidents.

The aim of this study is to investigate the main factors affecting the severity of accidents in these two distinct parts of the motorway and to identify any differences between them. In this context, focus is drawn on a number of factors that are commonly reported as important in road accidents. These are driver fatigue, accidents involving heavy goods vehicles (HGVs), and driver errors.

The rest of the paper is organized as follows: the next section provides a discussion on the factors that are commonly reported as important in road accidents, followed by the statistical methods used in this study. Then the estimation results along with a discussion on the findings are presented. The conclusions at the end are followed by some relevant practical applications of the work.

2. Factors affecting road accidents

Identifying causes of accidents and possible factors that increase the accident risk for the driver improves the development of countermeasures related to traffic safety (Häkkänen & Summala, 2000). The factors
affecting road accident severity are generally divided into two categories: engineering (road/vehicle/environment related) and human.

Engineering factors refer to road infrastructure characteristics, traffic conditions, and ambient conditions. The relationship between road traffic accidents and geometric design variables, such as curvature, vertical grade, lane width, and hard-shoulder width has been empirically investigated through statistical models in several studies (Haynes, Jones, Kennedy, Harvey & Jewell, 2007; Kononov, Bailey & Allery, 2008).

Regarding traffic conditions, speed appears to be important for both severity and frequency of accidents and is included in several alternative forms, for instance average speed, speed limit, and speed variation (Elvik, Christensen & Amundsen, 2004; Aarts & van Schagen, 2006).

Traffic is also expressed using a range of variables, such as traffic flow, traffic density, and congestion (Golob & Recker, 2003; Wang, Quddus & Ison, 2009). Inclement prevailing weather conditions appear to be quite significant for road accidents in various aspects, namely precipitation, wind, and fog. Road vehicles in strong winds can experience a variety of problems depending upon the vehicle type, shape and the wind dynamic (Edwards, 1994), while rainfall is suggested to affect drivers in different ways across different geographic areas and times of the day (Brijs, Karlis, & Wets, 2008).

Human factors also play a very important role in road accidents. In general, different groups of drivers have a different risk of being involved in an accident. HGV drivers have proven to be in the high risk group (Charbotel, Chiron, Martin, & Bergeret, 2002). In the United Kingdom (UK), fatal road accidents per 100 million vehicle kilometers involving HGVs are almost double of those involving cars (Department of Environment, Transport and the Regions, 1998).

Identifying drivers at risk could facilitate more effective traffic safety work and allow measures to be tailored toward a specific driver group.

Driver fatigue appears to be one of the most often reported factors in road accidents. According to Karreer and Roetting (2007) falling asleep at the wheel is one of the leading causes of fatal accidents and injuries, accounting for up to 15–20% of all traffic accidents in developed countries. However, it is often overlooked in police reports, as some drivers are deceased in the accidents, while surviving ones may be unwilling to admit that they may have been asleep (Corfitzen 1999). HGV drivers are a sensitive group fatigue-wise. Several previous studies have reported that they face fatigue-related problems while driving (Häkkänen & Summala, 2000; Zhang & Chan, 2014) and the drivers themselves recognize it (Häkkänen & Summala, 2001).

Nordbakke and Sagberg’s study (2007) showed that most drivers experience various symptoms of sleepiness while driving, such as difficulty keeping their eyes open before they fall asleep. However, these symptoms are not taken seriously enough; this may be due to an underestimation of the relation between the various physiological and behavioral signals. Even though drivers often fight the symptoms of sleepiness, even before the trip, they sometimes overestimate their own capability. In spite of the drivers’ knowledge of the risk, most drivers continue to drive even when recognizing sleepiness in themselves. The justification for this is most commonly related to the distance to be driven — total or remaining (Nordbakke & Sagberg, 2007). Horne and Reyner have contributed to the establishment of permanent awareness signs on the British motorways stating that “Tiredness Can Kill” and encouraging drivers to take a break (Horne & Reyner, 2001).

Several studies have also been conducted to correlate the accidents (both frequency and severity) with their causes or contributory factors. However, there is no previous significant research referring to accidents that happen specifically on the HS. These incidents have not been thoroughly examined to determine their specific primary causes or their contributory factors and how these have an effect on the level of severity.

3. Statistical models

Accident severity is, usually, defined as a categorical variable; the values of which vary according to the method being officially suggested by the authority designing and collecting accident data. These values represent the ‘level of severity’ in an ordinal scale (e.g., no injury, slight injury, serious injury, fatality). In GB, accident severity is recorded as slight, serious, and fatal. For this study, discrete choice modeling, in which a decision maker chooses an alternative from a set of exhaustive and mutually exclusive alternatives (Train, 2009), can be chosen in order to explore the most important factors affecting the severity of accidents on British motorways. In the case of road accident severity models, researchers are trying to explore the factors that can be related to road characteristics, the users, the vehicles, or the conditions on the road at the time of the accident.

When the dependent variable (accident severity) is discrete and contains more than two categories, the multinomial logistic regression (logit) model is a well-established method to use (e.g., Khorashadi, Niemeier, Shankar, & Manning, 2005). The multinomial logit model combines the independent variables to estimate the probability that a particular event will occur; in this case the probability of an accident to be slight, serious, or fatal. When the order of the values is taken into consideration, thus hierarchy has a natural meaning such as in this study, the commonly used model is an ordered logistic regression (Quddus, Wang, & Ison, 2010).

If the data are nested, such as accidents nested within roads and roads nested within areas (e.g., a census tract), the use of multilevel ordered logit models would be more appropriate. Such models can allow a possible correlation structure among a set of observations from the same level (Lord & Manning, 2010; Dupont, Papadimitriou, Martensen, & Yannis, 2013). Based on the data hierarchy, the first decision concerns the choice of the levels of analysis. Formulated generally, a level is a set of units, or equivalently a system of categories, or a classification factor in a statistical design. In order to justify the use of the multilevel model, the significance of the Intra-class Correlation Coefficient (ICC) must be tested. In the case of a two-level model, one ICC can be estimated. It varies from 0 to 1 and indicates whether the multiple levels are appropriate for the data, as it shows the similarity of observations within a group.

Intercorrelation among accidents that share some common characteristics has not been widely examined in past studies. Savolainen, Manning, Lord, & Quddus’ study (2011) involved the intercorrelation among the accidents that occur in the same geographical area. The area variable chosen to group correlated motorway accidents is county (Fig. 1). Counties derive from the geographical and administrative area division system in GB. For consistency of the area sizes, only data from England are used.

The multilevel mixed-effects ordered logit model allows for many levels of nested clusters of random effects. A quantity being random means that it fluctuates over units in some population, and which particular unit is being observed depends on chance. In this case, the coefficients of the model are not standard (fixed) for the whole sample. They follow a distribution and conclusions are drawn about the population from which the observed units were taken, rather than about these particular units themselves. It allows the model to account for unobserved effects that are difficult to quantify and may affect the model estimation (Ye & Lord, 2014).

In the two-level model, random effects of the accidents clustered in the English counties can be specified. The model estimates the possibility of an accident having a specific level of severity, given that the accident has occurred in a certain county — the second level of the model. For a series of independent clusters, and conditional on a set of fixed effects $x_p$, a set of cut-points $n$ that define the limit values between the different categories, and a set of random effects $u_p$, the probability of the response $y_p$ being in a category $m$ is:

$$p_{ij} = \Pr(y_{ij} = m | h, u_j) = \frac{1}{1 + \exp(-K_{m-1} + n_{ij})}$$

where:

$$n_{ij} = \frac{1}{K_m - 1}$$

$$K_m = \exp(-K_{m-1} + n_{ij})$$

The multinomial ordered logit model would be more appropriate. Such models can allow a possible correlation structure among a set of observations from the same level (Lord & Manning, 2010; Dupont, Papadimitriou, Martensen, & Yannis, 2013). Based on the data hierarchy, the first decision concerns the choice of the levels of analysis. Formulated generally, a level is a set of units, or equivalently a system of categories, or a classification factor in a statistical design. In order to justify the use of the multilevel model, the significance of the Intra-class Correlation Coefficient (ICC) must be tested. In the case of a two-level model, one ICC can be estimated. It varies from 0 to 1 and indicates whether the multiple levels are appropriate for the data, as it shows the similarity of observations within a group.

Intercorrelation among accidents that share some common characteristics has not been widely examined in past studies. Savolainen, Manning, Lord, & Quddus’ study (2011) involved the intercorrelation among the accidents that occur in the same geographical area. The area variable chosen to group correlated motorway accidents is county (Fig. 1). Counties derive from the geographical and administrative area division system in GB. For consistency of the area sizes, only data from England are used.

The multilevel mixed-effects ordered logit model allows for many levels of nested clusters of random effects. A quantity being random means that it fluctuates over units in some population, and which particular unit is being observed depends on chance. In this case, the coefficients of the model are not standard (fixed) for the whole sample. They follow a distribution and conclusions are drawn about the population from which the observed units were taken, rather than about these particular units themselves. It allows the model to account for unobserved effects that are difficult to quantify and may affect the model estimation (Ye & Lord, 2014).

In the two-level model, random effects of the accidents clustered in the English counties can be specified. The model estimates the possibility of an accident having a specific level of severity, given that the accident has occurred in a certain county — the second level of the model. For a series of independent clusters, and conditional on a set of fixed effects $x_p$, a set of cut-points $n$ that define the limit values between the different categories, and a set of random effects $u_p$, the probability of the response $y_p$ being in a category $m$ is:
where $j = 1, ..., j$ clusters (i.e. counties), with cluster $j$ consisting of $i = 1, ..., n_j$ observations (i.e. accidents).

The model, in terms of a latent linear response, where observed ordinal responses $y_{ij}$ are generated from the latent continuous responses, is written as:

$$y_{ij} = \beta_j + u_{ij} + \epsilon_{ij}$$

(2)

$$y_{ij} = \begin{cases} 1 & \text{if } -\infty \leq y_{ij} < \kappa_1 \\ 2 & \text{if } \kappa_1 \leq y_{ij} < \kappa_2 \\ ... & \text{if } \kappa_{m-1} \leq y_{ij} < \kappa_m = \infty \end{cases}$$

(3)

The errors $\epsilon_{ij}$ are distributed as logistic with mean 0 and variance $\pi^2/3$ and are independent of $u_{ij}$. The model is estimated using the adaptive Gauss–Hermit quadrature estimation.

When there are no theoretical or other prior guidelines about which variables should have a random effect, the researcher can be led by the substantive focus of the investigation, the empirical findings and the parsimonious modeling. This implies that those explanatory variables that are especially important – or have especially strong effects – could be modeled with random effects if the variances of these effects are important enough, as evidenced by their significance and size. Nonetheless one should take care that the number of variables with random effects should not be so large that the model becomes unwieldy (Snijders, 2005; Snijders & Bosker, 1999).

In cases where there is no identified inter-relationship among the observations, according to ICC, and/or a higher level random intercepts are statistically insignificant, a single level ordered logit should be used. However an important underlying assumption of the ordered logit model is the parallel regression assumption. Accordingly, the relationship between each pair of outcome groups (e.g., slight vs. serious and fatal; slight and serious vs. fatal) is the same. This assumption can be tested using the Brant test, which provides evidence if it is violated (Brant, 1990). In this case, the results of a simple ordered logit model can lead to incorrect or misleading results and a different model should be applied that allows the relaxation of the parallel regression assumption.

Such an alternative is the generalized ordered logit model. This model’s advantage is that it does not impose the parallel regression assumption, which is an important limitation of the ordered logit model. In this model, the coefficients of the variable vary among the categories of the variables. If this is not the case for all the variables, the model is called partially constrained. It has been recommended by Eluru (2013) for the case of ordinal data and has been adopted in the area of road safety (Abegaz, Berhane, Worku, Assrat, & Assefa, 2014; Wang, Qudus & Ison, 2009).

In the partially constrained generalized ordered logit model, only a subset of variables has a varying coefficient, the ones that violate the aforementioned assumption. The generalized ordered logit model can be written as follows (Williams, 2006):

$$P(Y_i > j) = \frac{\exp(a_j + X_i\beta_j)}{1 + \sum_{k=1}^{m-1} \exp(a_k + X_i\beta_k)}$$

(4)

where $m$ is the number of categories of the ordinal dependent variable, while the partially constrained model as:

$$P(Y_i > j) = \frac{\exp(a_j + X_i\beta_j + X_{ij}\kappa_j)}{1 + \sum_{k=1}^{m-1} \exp(a_k + X_i\beta_k + X_{ik}\kappa_k)}$$

(5)

In the last model, the first subset of variables has a non-constrained coefficient across the values, and the second subset has the same coefficient across the values of $j$. For the case of a dependent variable that has 3 values (e.g. 1, 2, 3), two panels of coefficients are provided as if the variable is recoded as 1 vs. 2 & 3 and 1 & 2 vs. 3. Positive coefficients suggest that higher values on the independent variables make higher values of the dependent variable more likely. The parameters are estimated using the maximum likelihood estimation technique.

4. Accident data description

In GB all road accidents involving fatalities or personal injury, in which one or more vehicles are involved, are notified within 30 days of occurrence to the police, who attend the scene and collect the data required. Details recorded are organized in three datasets which include complete information on the accident itself, the casualties (any persons injured or killed) and every vehicle involved or contributing to the accident. The subsequent collated output reports, entitled STATS19, have been available since 1985.

For this study, the three sub-sets are merged, using the accident and vehicle reference numbers provided. The unit of analysis is the accident. After merging, if there is more than one record per accident (for instance, where there is more than one vehicle involved), the duplicates are removed. Motorway accidents are then extracted from the whole database and divided into two groups: (1) accidents that occurred on the MC and (2) accidents that occurred on the HS (i.e. entering, leaving or parked on the HS). From 1985 to 2011, there were a total of 199,388 accidents (in which 2.3% were fatal, 13.0% serious injury and 84.7% slight injury accidents) on GB motorways.

The distinction between the HS and MC accidents is based on the location where the accident happened. As described in STATS20 – the document for the specification of the variables included in STATS19 – an accident should be located where the first impact occurred. For example, an accident where two vehicles on the MC collide and one ends up on the HS is not labelled as a HS accident.

The dependent variable of the models is ‘severity of accident,’ a discrete variable that can obtain three values: slight injury, serious injury and fatality. Fatal injury includes the cases where death occurs in less than 30 days as a result of the accident. Serious injuries are those
where either immediate or later detention in hospital as an in-patient, was required.

For this study, new variables were created to keep more detailed information regarding the vehicles involved in the accident. These variables refer to whether the accident involved one vehicle or more, at least one HGV, at least one left-hand-side drive vehicle or whether roadworks were present at the time of the accident. For instance, if at least one HGV was involved in an accident, the variable has the value 1 (thus 0 for the opposite). Also, in order for the models to be estimated, all categorical variables were transformed into the required number of dummy variables (day of the week, month, speed limit, traffic, weather, surface and light conditions). Trend is also included in the model as a yearly dummy variable.

In 2005, the police were asked to record additional data, which are more subjective and related to drivers, vehicle or road characteristics; these are known as contributory factors. These are factors in a road accident – in the opinion of the attendant police officer – that are the key actions and failures that led directly to the actual impact. They postulate why the accident occurred and give readers clues as to how it may have been prevented. In comparison to the accident, vehicle and casualty record sets, which mainly record objective details, the contributory factors depend on the skill and experience of the investigating officer to reconstruct the events that led directly to the accident. While instructions are also provided, factors should be identified on the basis of evidence rather than guess-work about what may have happened. Up to six factors are collected for each accident. They are clustered in certain categories according to STATS20; however, a different grouping was followed in this study, in order to minimize the number of categories and to create more cohesive groups. In addition, as special focus is given to fatigue, a single category only for this is created.

The groups of contributory factors which are studied are related to:

- **driver/rider fatigue**, which in the initial police records comes under the impairment category, but is individually examined in this study;
- **driver/rider error** (e.g., failed to look properly, failed to judge other person’s path or speed, sudden braking, swerved, loss of control, nervous or uncertain behavior, learner or inexperienced driver);
- **driver/rider behavior (aggressive or illegal)** (e.g. exceeding speed limit, travelling too fast for conditions, aggressive/careless/reckless driving or in a hurry);
- **driver/rider impairment** (e.g., impaired by alcohol/drugs, defective eyesight, illness);
- **road conditions** (e.g., poor road surface, foreign deposit on the road, slippery road due to weather, road layout, vision affected by dazzling headlight/sun/rain);
- **vehicle** (e.g., tyres illegal or defective, defective lights/indicators/brakes/steering/mirrors, overloaded or poorly loaded vehicle or trailer);
- **distraction** (e.g., driver using mobile phone, eating/drinking, distraction outside the vehicle); and
- **pedestrians**.

Multicollinearity among the contributory factors was checked via the calculation of pair-wise correlations. Only error and behavior factors were found to have high correlation in both datasets (0.87 for HS and 0.88 for MC accidents). Thus, these two variables were not included in the models simultaneously.

The data used for both models are from 2005 to 2011. Table 1 provides the frequencies and severity of the accidents for the period examined. It is noted that the severity of HS accidents is significantly higher than the MC, as the relative frequency of HS accidents is almost 5 times higher. The absolute frequency of HS accidents might be quite low; however, when these accidents occur, it is quite probable they tend to a serious or fatal injury. The relative frequencies of some characteristics between HS and MC accidents are significantly different. For instance, out of all the fatal accidents, the percentage of those that involved driver’s fatigue is double on the HS than that of the corresponding percentage for MC accidents (23.08% vs. 11.63%). Other such examples for fatal accidents are:

- **HGVs**: 80% HS while 44.36% MC
- **Single-vehicle accidents**: 7.69% HS while 39.09% MC
- **Morning peak traffic**: 13.55% HS while 7.69% MC
- **Pedestrian(s) involved**: 15.38% HS while 9.83% MC
- **Contributory factor ‘road’**: 18.46% HS while 8.39% MC.

For other variables, such as the surface condition, the above ratios are similar for HS and MC accidents; i.e. 70% of the fatal accidents occur when the surface is dry and the remaining 30% when the opposite is true (e.g., wet, flooded). If all accidents, and not only fatal, are taken into consideration, the ratios are sometimes substantially different; for example, the percentage of all accidents that are single-vehicle is 19.07% for HS and 22.49% for MC vs. 7.69% and 39.09% for fatal.

### 5. Estimation results and discussion

The statistical models are applied to the HS and MC motorway accidents separately and the relationships between the levels of severity of these accidents are explored with a series of explanatory variables. As mentioned earlier, the unit of analysis is the accident. The models that are initially tested are the simple ordered logit model and the multilevel ordered logit model. Firstly, all the independent variables considered to be potentially significant and to have an effect on accident severity are included in the model. A random intercept multilevel ordered logit model was initially estimated and then, as a second step, the random effects of the independent variables are included in the model one at a time. After all the variables have been tested independently, models that include two or more independent variables with possible random effects are estimated. However, the ICC was found to be low (i.e., 0.016 for MC accidents and 0.1 for HS accidents) suggesting that the correlation among the accidents that occur in the same county is not strong enough to support the use of the multilevel model. Therefore, a single-level model is then considered.

When the Brant test is used, the ordered logit model is also considered inappropriate for this study, as the parallel regression assumption is violated. Thus, it is necessary to apply the generalized ordered logit model, and more specifically, the partially constrained model. In these models, as there are three categories in the dependent variable (M = 3), two coefficients (M-1) are estimated for each of the explanatory variables that violated the parallel regression assumption: one coefficient represents the effect of an explanatory variable on the outcome of slight relative to serious and fatal; and the other indicates the effect of the same explanatory variable on the outcome of slight and serious relative
to fatal. The models for the MC and HS accidents are separately estimated while following the same process as above with the explanatory variables. Only the statistically significant variables are kept in the final models.

Table 2 presents separately the results for the two partially constrained generalized ordered logit models for the two groups of accidents. The variables kept in the model are those whose parameters are statistically significant at the 90% level. An alternative parameterization, called Gamma ($\gamma$), provides a more parsimonious layout and another way of understanding the parallel regression assumption (Peterson & Harrell, 1990). In this alternative parameterization, Beta coefficients refer to the effect of variables that have the same coefficients for all possible pairs of outcome categories, while Gamma coefficients refer to the differential effect of the variables on each pair of outcomes. The Gammas indicate the extent to which the parallel regression assumption is violated by the variable. Thus, if Gammas are statistically different from 0 for an explanatory variable, then the parallel regression assumption is considered to be violated for this variable. If all Gammas are equal to 0, the model would reduce to the ordered logit model. Table 3 presents the marginal effects, which show the probabilities of having a specific outcome with respect to changes in the dependent variables. The variables included in the table are the ones with non-zero Gammas plus fatigue and two variables with high marginal effects in comparison to the others: number of casualties and single-vehicle accidents.

### 5.1. Accident characteristics

Regarding the accident characteristics in the MC model, the variables referring to the number of vehicles, number of casualties as well as single-vehicle accidents are significant. In the case of HS accidents, only the number of casualties appears to be significant, while having a higher

---

**Table 2**

Gologit2 model results for main carriageway and hard-shoulder accident severity.

<table>
<thead>
<tr>
<th>Explanatory variables</th>
<th>Main carriageway</th>
<th>Hard-shoulder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>z</td>
</tr>
<tr>
<td><strong>Beta</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accident characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>0.0549</td>
<td>3.55</td>
</tr>
<tr>
<td>Number of casualties</td>
<td>0.2802</td>
<td>23.83</td>
</tr>
<tr>
<td>Single vehicle accident</td>
<td>0.9494</td>
<td>22.53</td>
</tr>
<tr>
<td>Vehicle characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heavy goods vehicle</td>
<td>0.3353</td>
<td>9.54</td>
</tr>
<tr>
<td>Left-hand-side drive</td>
<td>−0.2858</td>
<td>−3.82</td>
</tr>
<tr>
<td>Seasonality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturday</td>
<td>0.1275</td>
<td>2.93</td>
</tr>
<tr>
<td>Sunday</td>
<td>0.1649</td>
<td>3.88</td>
</tr>
<tr>
<td>Traffic characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traffic peak morning</td>
<td>−0.2837</td>
<td>−4.76</td>
</tr>
<tr>
<td>Traffic peak afternoon</td>
<td>−0.4111</td>
<td>−7.34</td>
</tr>
<tr>
<td>Traffic peak normal</td>
<td>−0.3627</td>
<td>−7.23</td>
</tr>
<tr>
<td>Reference: traffic non peak</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speed limit 70 mph</td>
<td>0.6365</td>
<td>3.85</td>
</tr>
<tr>
<td>Speed limit 60 mph</td>
<td>0.6962</td>
<td>3.51</td>
</tr>
<tr>
<td>Speed limit 50 mph</td>
<td>0.5601</td>
<td>3.12</td>
</tr>
<tr>
<td>Conditions on the motorway</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daylight</td>
<td>−0.2825</td>
<td>−5.89</td>
</tr>
<tr>
<td>Dark and lights on</td>
<td>−0.2909</td>
<td>−5.80</td>
</tr>
<tr>
<td>Reference: dark and no lights</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface dry</td>
<td>0.1324</td>
<td>3.15</td>
</tr>
<tr>
<td>Reference: surface non-dry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weather fine</td>
<td>0.1509</td>
<td>3.01</td>
</tr>
<tr>
<td>Weather fog</td>
<td>0.3199</td>
<td>2.10</td>
</tr>
<tr>
<td>Contributory factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CF fatigue</td>
<td>0.3891</td>
<td>6.91</td>
</tr>
<tr>
<td>CF error</td>
<td>0.0736</td>
<td>2.36</td>
</tr>
<tr>
<td>CF impairment</td>
<td>0.6468</td>
<td>11.58</td>
</tr>
<tr>
<td>CF distraction</td>
<td>0.1756</td>
<td>2.49</td>
</tr>
<tr>
<td>CF pedestrian</td>
<td>1.4370</td>
<td>13.09</td>
</tr>
<tr>
<td>CF vehicle</td>
<td>0.1372</td>
<td>1.91</td>
</tr>
<tr>
<td>CF other</td>
<td>0.2538</td>
<td>3.81</td>
</tr>
<tr>
<td>Roadworks</td>
<td>−0.2017</td>
<td>−3.56</td>
</tr>
<tr>
<td>CF fatigue</td>
<td>0.3891</td>
<td>6.91</td>
</tr>
<tr>
<td>CF error</td>
<td>0.0736</td>
<td>2.36</td>
</tr>
<tr>
<td>CF impairment</td>
<td>0.6468</td>
<td>11.58</td>
</tr>
<tr>
<td>CF distraction</td>
<td>0.1756</td>
<td>2.49</td>
</tr>
<tr>
<td>CF pedestrian</td>
<td>1.4370</td>
<td>13.09</td>
</tr>
<tr>
<td>CF vehicle</td>
<td>0.1372</td>
<td>1.91</td>
</tr>
<tr>
<td>CF other</td>
<td>0.2538</td>
<td>3.81</td>
</tr>
<tr>
<td>Reference: CF road</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gamma_2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of vehicles</td>
<td>0.0545</td>
<td>3.07</td>
</tr>
<tr>
<td>Heavy goods vehicle</td>
<td>0.4001</td>
<td>5.80</td>
</tr>
<tr>
<td>Saturday</td>
<td>0.1986</td>
<td>2.30</td>
</tr>
<tr>
<td>Traffic peak afternoon</td>
<td>−0.2098</td>
<td>−2.06</td>
</tr>
<tr>
<td>Daylight</td>
<td>−0.3800</td>
<td>−5.57</td>
</tr>
<tr>
<td>CF error</td>
<td>−0.1901</td>
<td>−2.83</td>
</tr>
<tr>
<td>CF pedestrian</td>
<td>1.0940</td>
<td>9.80</td>
</tr>
<tr>
<td>CF distraction</td>
<td>0.2790</td>
<td>2.09</td>
</tr>
<tr>
<td>Alpha</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant 1</td>
<td>−3.5187</td>
<td>−19.80</td>
</tr>
<tr>
<td>Constant 2</td>
<td>−5.6330</td>
<td>−29.22</td>
</tr>
<tr>
<td>Summary statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>−17,984.448</td>
<td>522.214</td>
</tr>
<tr>
<td>Pseudo R squared</td>
<td>0.0708</td>
<td>0.1241</td>
</tr>
<tr>
<td>Number of observations</td>
<td>47,094</td>
<td>776</td>
</tr>
</tbody>
</table>

*na = non-applicable.*
coefficient than MC. These three variables all have a positive coefficient, affecting in an increasing way the severity of accidents if all else are held constant. In terms of the number of vehicles involved in an accident, for the case of MC model, the variable does not have the same effect across both pairs of outcome categories, which would not have been found if the simple ordered logit model was employed. The Beta coefficient for number of vehicles is 0.0549 denoting the effect of this variable on the slight versus serious and fatal outcome. The Gamma value for the variable is 0.0545 suggesting that the effect of this variable on the slight and serious injury versus fatal outcome is 0.1094 (i.e. 0.0549 + 0.0545). The severity of accidents is decided according to the most severe casualty. Thus, accidents that have only slight injuries, regardless of the number of casualties, are recorded as slight injury accidents. The results of the MC model suggest that the higher the number of casualties, the higher the possibility of the accident to be serious or fatal, which is also shown from the marginal effects being positive for these two outcomes.

5.2. Vehicle characteristics

The next set of variables refers to the characteristics of the vehicles embroiled in the accident; both variables refer to whether at least one of the following types of vehicles was involved: HGVs or left-hand-side drive vehicles. HGVs appear to have a positive coefficient, suggesting that the severity of accident is increased when there is at least one HGV involved. HGVs do have the ability to ‘protect’ its own occupants due to the size and structure of the vehicle; however, for the same reasons it causes a more serious impact to the other vehicles involved (Özkan & Lajunen, 2010). It is worthwhile to note that in the case of HS accidents, this variable has a much greater effect and also has a varying coefficient for the second threshold. In opposition to HGVs, left-hand-side driving vehicles tend to reduce the level of severity of accidents. The opposite result could be expected due to added difficulty of driving or inexperience of foreign drivers in the British driving system. However, a possible explanation for this result can be that these drivers are more cautious while driving in a non-familiar environment and from their driving position have a better visibility of the HS.

5.3. Seasonality

Another set of variables that appears to have a significant effect on the severity of MC accidents is the day of the week when the accident happened. From the results of the models, it is shown that during working days, the accidents are less severe than during the weekend.

5.4. Traffic characteristics

Since no detailed traffic data were available, the way of incorporating this information in the model was by creating four variables according to the hour that the accident happened. These were based on historical data of peak in the morning, the late afternoon peak, normal and non-peak (quiet) traffic hours. The model illustrates that if the hour is non-peak, the severity tends to be higher. This might be related to speed, which is generally higher during non-peak hours and is frequently related to more severe and fatal accidents (Elvik, Christensen, & Amundsen, 2004). The reference variable for this set was non-peak and the rest of the variables appeared to have a negative coefficient. In addition, ‘traffic peak afternoon’ has a different second coefficient; the Gamma value is negative suggesting that the probability of having a fatal accident is even lower. However, the difference between serious/fatal is lower than slight/serious, as it is suggested by the marginal effects.

Another traffic related variable is speed limit. When comparing the higher speed limits with the variable left out of the model, which represents a speed limit of 40 mph or lower, it is suggested that higher speed limits cause an increase of the level of severity of the road accidents. Past studies have shown contradictory results regarding the relationship between speed and accident severity, thus more research is required in this field (Wang, Quddus, & Ison, 2013).

5.5. Conditions on the motorway

Regarding lighting conditions, the same conclusions as previous studies are drawn: it is supported that the presence of light (daylight or street lighting) has an effect on the accident severity by decreasing it; thus it is expected to have lower severity accidents when visibility on the motorway is better.

Two types of road surface conditions were considered: dry and non-dry. The latter includes snow, frost, and flood. It is estimated that when the condition of the surface is dry, it is more probable to have a more severe accident, which is consistent with other studies (Quddus, Wang & Ison, 2010). This result is plausible, as under adverse surface conditions, drivers tend to drive at a lower speed and awareness in general is increased. Similarly, when weather is fine, accidents tend to be more severe for the case of MC accidents, as the variable is significant with a positive coefficient. Wind was also tested in the initial models, but did not appear to be significant in any.

‘Roadworks’ is a variable that represents whether they were present at the time of the accident. This variable is significant in both models, while having a much greater effect in the HS model. It is estimated that if roadworks are present, the severity of an accident tends to drop. Again, this can be attributed to the speed restrictions that are
always imposed in these instances, as well as the drivers being more cautious. It has to be noted that if an accident happens in a closed lane that is normally running, the accident is still considered a HS accident.

5.6. Contributory factors

Even though these data are partially based on the police officer’s judgment, they can still provide useful information. Since the contributory factors are represented by a group of dummy variables, as described earlier, one of the dummy variables must be left out of the model. The variable representing the factors related to the road is chosen not to be included, as it is the factor that is least related to the driver. All contributory factors’ variables have a statistically significant coefficient and present the same sign. As it can be noticed, the highest coefficient is the one related to pedestrians involved in the accident. As this group of people is the most vulnerable, it is expected when a pedestrian is involved, for the accident to be more severe. After pedestrians the next highest coefficients belong to impairment and fatigue, which play an important role in road accidents by increasing significantly their severity (Evans, 1991).

Driver fatigue is a significant factor for both the MC and HS models. While it is the second most important contributory factor for HS accidents, after ‘pedestrians’, in the HS model, it is a much more important factor, as the value of the coefficient is substantially higher at 0.79, while it is 0.39 for MC accidents. Past studies (e.g., Häkkänen H., and Summala, H., 2001) have shown that very often HGV drivers feel tired while driving. This can be the result of long and monotonous journeys or the inability to stop due to timetable restrictions. It is crucial for HGV drivers to be aware of risk arising from fatigue, especially for HS accidents. The marginal effects are negative for slight injury, both for MC and HS accidents, while they are positive for serious and fatal injuries. They are also much higher for the case of HS. For instance, the marginal effects of fatigue for fatal injury are 0.0054 and 0.0522 for MC and HS respectively; thus, if the value of dummy variable ‘fatigue’ changes from 0 to 1, the possibility of having an accident to be fatal is getting 0.54% higher for the MC and 5.2% higher for the HS.

Error and behavior related factors appear to have a very high correlation (0.88 for MC and 0.87 for HS), thus it is important to control whether one for these variables should be removed from the model. After including each of them individually and both together, it was concluded that only one should be kept; error is the selected one, as it is of a high interest and also has a varying coefficient. The high correlation between the factors above shows that most of the time they occur in the same accident. Even though contributory factors data are, up to a point, inevitably subjective; we can recognize that a driver’s irresponsible behavior may lead to driving errors both to the driver or even other users of the motorway. From the marginal effects, which are positive only for serious injury accidents, it can be concluded that when driving errors are involved, the possibility of slight and fatal injury accidents drop.

Fig. 2 illustrates the observed and predicted values of the probabilities of an accident that occurred to be slight, serious, or fatal, under...
the conditions that fatigue was involved or not (for HS accidents) and an HGV was involved or not (for MC accidents). The left set of bars shows the probabilities if the condition is true and the right if it is not.

6. Conclusions

The motivation of this study was to extend research in the field of hard-shoulder accidents as the severity of these accidents appears to be significantly higher. The aim was to identify any differences between the factors that contribute to the severity of HS and MC motorway accidents using disaggregated accident data from motorways in England. Driver fatigue appears to be a much more common contributory factor for HS accidents, thus it was considered important to highlight its impact.

Severity of accidents is a discrete ordered variable and several models were examined in order to identify the most appropriate. The multilevel ordered logit model could have incorporated a second level that would represent any correlation among the accidents. In this study this level was counties, English geographical, and administrative areas. However, it was estimated that the correlation among the motorway accidents in the same county was not strong enough to support the use of this model. In addition, the ordered logit model was also not suitable due to the violation of the parallel regression assumption; thus the generalized ordered logit model, which allows the relaxation of the assumption, was preferred.

Two models, one for MC and one for HS accidents, were estimated and the results suggested substantial differences in the statistically significant variables. The variable with the highest impact for MC accidents was speed limit, which increases the severity. In addition, when an accident is single-vehicle, it is more likely for it to be fatal. Traffic volume was incorporated indirectly, which was one of the limitations of the study and it was estimated that during non-peak hours, the severity of MC accidents tends to increase. Dry surface conditions also have the same effect. On the contrary, left-hand-side drive vehicles and the presence of roadworks at the time of the accident, as well as good visibility, have a positive effect, giving more probabilities for a slight injury. According to the generalized ordered logit model, some of the variables may not have the same effect on all the categories of severity, which would not have been detected with the simple ordered logit model.

For the MC accident model, several variables, such as HGV and the number of vehicles involved have two different coefficients, one for slight injury versus serious and fatal, and one for slight and serious versus fatal injury. In terms of HS accidents, the number of significant variables is lower. Variables such as HGVs and fatigue are common for both models, but appear to have a much higher impact in this case. Fatigue also has a varying coefficient increasing even more the probability of a fatal accident when involved in a HS accident.

The limitations of this study are mainly related to data integrity. As mentioned, STATIS91 accident data are collected by the police attending the accident scene and despite thorough training and special instructions provided, the data might be recorded subjectively; thus the contributory factors may not always be accurate. Furthermore, some proxies for risk factors (peak time and speed limits for traffic flow and speed) were used in this study. It would be of interest for future studies to further examine the effect of the actual traffic characteristics (e.g., flow and speed) on the HS and MC accident severity.

7. Practical application

Considering that the level of severity of hard-shoulder accidents is five times higher than for the MC, the need for extra safety measures through informative campaigns and training should be considered. This study concludes that HGV drivers are a high risk group, especially for hard-shoulder accidents, and also fatigue appears to be a crucial factor. Therefore, it is important for the public sector or other related organizations to focus their safety campaigns to this specific group and especially raising the drivers’ awareness of the hazards arising when using the nearside lane of the motorway. Professional drivers can also be targeted via their employers in order to be provided with additional training. Lastly, private vehicle drivers can also be informed by campaigns that are focusing on the risks arising from tiredness.

Acknowledgments

This research was undertaken as part of an Engineering Doctorate project jointly funded by the Centre of Innovative and Collaborative Construction Engineering (CICE) at Loughborough University and Balfour Beatty. The support of the Engineering and Physical Sciences Research Council is gratefully acknowledged (EPSRC Grant EP/F037272/1).

References


Paraskevi Michalaki, who specialised in transportation engineering graduated from the Civil Engineering School of the National Technical University of Athens in 2012 before joining the Loughborough University Engineering Doctorate research programme related to risk estimation of hard-shoulder incursions. Further to her studies, Paraskevi has been involved in cross-division health and safety and risk forum initiatives in Balfour Beatty and has participated in road-worker safety programmes with Highways England. She also works as an STEM ambassador for the Institute of Civil Engineers in the United Kingdom.

**Prof. Mohammed Quddus** is a professor of Intelligent Transport Systems at the School of Civil and Building Engineering at Loughborough University. His primary research areas include: accident modelling, Intelligent Transport Systems (ITS) and transport modelling. Prof Quddus has led and been actively involved in numerous research projects funded by research councils, government agencies and industry. He has published over 70 peer-reviewed journal articles. He currently serves on the editorial board of *Transportation Research Part C: Emerging Technologies* as an associate editor.

**Dr. David Pittfield** has been a senior member of the Transport Studies Group at Loughborough for many years. He has taught, supervised, and published in many fields of transport but has, in the last 10 years or so, focused on issues in air transport research. This has notably covered the modelling of aircraft accidents the assessment of the start-up impact of low cost carriers on traffic and the assessment of regulatory changes such as the EU-US Open Skies Agreement. The impact of this body of work has resulted in his appointment as an Academician of the Academy of Social Sciences.

**Andrew Huetson** has twenty-five year experience in civil engineering contracting at both site and commercial level, leading on innovation and product deployment. In six years as business manager for Connect Roads, Andrew has worked closely with their clients to introduce and adopt new ideas and processes. He has overseen various research partnerships with Universities and acts as Industrial Supervisor to two Engineering Doctorate programs. Previously, Andrew spent six years in the Irish Construction market, responsible for bidding major highway, heavy civil and term maintenance contract, including quality bids and post contract claims.