



UNIVERSITY OF LEEDS

This is a repository copy of *Vehicle mass and injury risk in two-car crashes: a novel methodology*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/43781/>

---

**Article:**

Tolouei, R, Maher, MJ and Titheridge, H (2013) Vehicle mass and injury risk in two-car crashes: a novel methodology. *Accident Analysis and Prevention*, 50. 155 - 166. ISSN 0001-4575

<https://doi.org/10.1016/j.aap.2012.04.005>

---

© 2012, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

**Reuse**

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# Vehicle Mass and Injury Risk in Two-Car Crashes: A Novel Methodology

Reza Tolouei<sup>1 a</sup>

(Corresponding author)

<sup>a</sup> Centre for Transport Studies, University College London, Gower Street, London  
WC1E 6BT, UK

Email: tolouei\_reza@hotmail.com

Telephone: + (44) 7858 758 291

Mike Maher<sup>b</sup>

<sup>b</sup> Institute for Transport Studies, University of Leeds, Leeds  
LS2 9JT, UK

Email: M.J.Maher@its.leeds.ac.uk

Telephone: + (44) 16 2089 5607

Helena Titheridge<sup>c</sup>

<sup>c</sup> Centre for Transport Studies, University College London, Gower Street, London  
WC1E 6BT, UK

Email: h.titheridge@ucl.ac.uk

Telephone: + (44) 20 7679 7775

## Abstract

This paper introduces a novel methodology based on disaggregate analysis of two-car crash data to estimate the partial effects of mass, through the velocity change, on absolute driver injury risk in each of the vehicles involved in the crash when absolute injury risk is defined as the probability of injury when the vehicle is involved in a two-car crash. The novel aspect of the introduced methodology is in providing a solution to the issue of lack of data on the speed of vehicles prior to the crash, which is required to calculate the velocity change, as well as a solution to the issue of lack of information on non-injury two-car crashes in national accident data. These issues have often led to focusing on relative measures of injury risk that are not independent of risk in the colliding cars. Furthermore, the introduced methodology is used to investigate whether there is any effect of vehicle size above and beyond that of mass ratio,

---

<sup>1</sup> Present address: AECOM Consultancy, 63-77 Victoria Street, St Albans, Herts AL1 3ER

and whether there are any effects associated with the gender and age of the drivers. The methodology was used to analyse two-car crashes to investigate the partial effects of vehicle mass and size on absolute driver injury risk. The results confirmed that in a two-car collision, vehicle mass has a protective effect on its own driver injury risk and an aggressive effect on the driver injury risk of the colliding vehicle. The results also confirmed that there is a protective effect of vehicle size above and beyond that of vehicle mass for frontal and front to side collisions.

Keywords: Two-car collision, driver injury risk, secondary safety, vehicle mass, velocity change

## 1. Background

Amongst various vehicle design features, vehicle mass is a key variable from a policy perspective because of its effect on fuel consumption and emissions on the one hand, and its effect on safety performance of vehicles within the fleet on the other hand. A decrease in mass of the vehicles within the fleet is associated with a decrease in overall fuel consumption and emissions (Tolouei and Titheridge, 2009); however, the effect of vehicle mass on safety is more complicated. While it is generally accepted that decreasing the mass of a vehicle, all other factors being constant, imposes a greater risk of injury to its occupants when the vehicle is involved in a crash, it is not clear what effect a change in the distribution of vehicles' mass within a fleet has on the overall safety of the fleet. This is mainly due to the uncertainty on the combined effect of mass of a vehicle on the safety performance of that vehicle as well as that on the safety performance of the other vehicles with which the vehicle collides.

There are two distinct aspects of the safety performance of a vehicle in a fleet: primary safety performance, which is linked to the risk of crash involvement of the vehicle, and secondary safety performance, which is linked to the risk of occupant injury (to a specific level) when the vehicle is involved in a particular type of crash. While there is no evidence of any direct effect of vehicle mass on the primary safety performance of a vehicle, mass is a key variable that is directly related to the secondary safety performance of the vehicle. In a two-vehicle crash, the injury risk of occupants in the lighter vehicle tends to be higher than that in the heavier vehicle due to the greater velocity change during the collision. In the case of a frontal collision, for example, between two vehicles with masses  $m_1$  and  $m_2$  travelling with speeds  $v_1$  and  $v_2$ , it can be easily shown using Newtonian mechanics that the velocity change of the first vehicle during the collision ( $\Delta v_1$ ) depends on the proportion of the total mass contained by the other vehicle ( $\frac{m_2}{m_1+m_2}$ ) and the closing speed ( $v_1 + v_2$ ):

$$\Delta v_1 = \left( \frac{m_2}{m_1+m_2} \right) (v_1 + v_2) . \quad (1)$$

There are therefore two aspects of the effects of mass of a subject vehicle on its secondary safety performance in a crash with another vehicle: a protective effect related to the injury risk (injury probability) of the occupants in the subject vehicle, and an aggressive effect related to the injury risk that mass imposes on the occupants of the other vehicles in collision with the subject vehicle. In order to investigate the relationship between vehicle mass and secondary safety performance, two-car crashes have been studied intensively in vehicle safety

research. This is because they form a case for vehicle crashes where both protective and aggressive effects of mass are best represented since the closing speed is identical for both drivers in the crash. Two-car crashes can also provide insight into crashes between any pair of vehicles and also into single-vehicle crashes (Evans, 1994). However, there are certain disadvantages or shortcomings associated with the methodologies used so far.

In a two-car crash, Equation 1 implies that the relative mass of the two cars directly influences the velocity change ( $\Delta v$ ).  $\Delta v$  has been regarded and used in vehicle safety research as the best measure of crash severity contributing to the injury risk of vehicle occupants (Evans, 1994). The main difficulty in investigating the relationship between injury risk and  $\Delta v$  is lack of information on the speed of the vehicles prior to a crash, which is required together with mass of the vehicles to calculate  $\Delta v$  (see Equation 1).

Equation 1 implies that in a two-car crash, the velocity change ratio is inversely related to the mass ratio of the cars ( $\Delta v_1/\Delta v_2 = m_2/m_1$ ). As a result of this relationship and lack of data on vehicle speed, several studies have investigated the relative injury risk in two-car crashes as a function of mass ratio. Evans (2004) has intensively studied the effect of vehicle mass in two-car crashes using 1978-1998 US crash fatality data and he has shown empirically that in a crash between two cars of different masses, the fatality risk ratio ( $R$ ) of the lighter to the heavier car increases as a power function of mass ratio ( $\mu = m_2/m_1$ ) of the heavier to the lighter car (Evans and Frick, 1993):

$$R = \mu^u. \tag{2}$$

The value of parameter  $u$  for a given set of two-car crashes is estimated by aggregating the crash data into categories associated with values of  $\mu$  in given ranges and estimating a least square fit to  $\log(R_k) = u \times \log(\mu_k)$  where  $R_k$  and  $\mu_k$  for each crash category  $k$  are the ratio of driver fatalities of the lighter cars to the heavier cars and the average mass ratio, respectively. Equation 2, which is regarded by Evans (2004) as the “first law of two-car crashes”, has been commonly accepted and used by the researchers and practitioners in the area of vehicle safety. Different values of exponent  $u$  for various sets of US crash data are estimated ranging from 2.70 (crashes in all directions) to 3.80 (frontal crashes) (Evans and Frick, 1992, Evans and Frick, 1993; Evans, 1994; Evans and Frick, 1994; Evans, 2001, Evans, 2004). While Evans’ relationship provides a simple approach to estimate injury and fatality risk ratio as a function of mass ratio in two-car crashes, it is associated with some

disadvantages (see Tolouei, 2011 for details). For example, Evans (1994) shows that the underlying assumption behind Equation 2 on the relationship between absolute driver injury and fatality risk ( $P$ ) and vehicle velocity change ( $\Delta v$ ) has the following form:

$$P = (\Delta v / \alpha)^k \quad (3)$$

where  $\alpha$  and  $k$  are parameters that are estimated from the crash data. As Evans (1994) points out, this relationship suffers from a major structural problem that results in values of risk greater than 1 when  $\Delta v > \alpha$ . Besides, this does not provide the relationship between  $P$  and mass of vehicles in two-car crashes explicitly. A few other studies have investigated a similar relationship to that of Evans (Equation 2) between fatality and injury risk ratio and mass ratio in two-car crashes (e.g. Ernst et al., 1991; Ernvall et al., 1992; Joksch, 1998; Ross and Wenzel, 2001). These are all empirical studies based on aggregate analysis of crash data which have used a similar approach to that of Evans and Frick (1993) as explained earlier.

Other studies have investigated the relationship between vehicle mass and driver risk of injury directly. For example, Broughton (1996a) discussed the effect of vehicle mass on injury risk in two-car crashes based on British crash data where injury risk is defined as the probability of driver injury when the vehicle is involved in a two-car crash in which at least one of the drivers is injured. He found that driver risk falls steadily with increasing mass according to a linear relationship and that mass could explain a high proportion of variation in the casualty data. This generally reflected the greater protection of drivers in the heavier cars compared to that of drivers in the lighter cars in fleet; however, this relationship alone does not provide any information on the aggressive effect of vehicle mass in fleet as it does not include the mass of the colliding vehicle. Besides, the measure of injury risk that he has used is not the ideal measure as it is a relative measure and hence is not independent of risk of injury in the colliding car (Broughton, 1996b). Wenzel and Ross (2005) defined risk as the driver deaths per year per million registered vehicles for a given car model and all types of crashes and found that mass alone is only a modestly effective predictor of risk. The difference between their results and those from previous studies on the effect of mass could arise because they used a different measure of vehicle safety; one which is a measure of both primary safety (crash involvement) and secondary safety (injury risk).

The ideal measure of secondary safety in two-car crashes is the absolute injury risk defined as the probability of driver injury when the vehicle is involved in a crash, whether or not the driver in the colliding vehicle is injured. However, the relationship between vehicles' mass

and absolute driver injury risk in two-car crashes has remained unclear in the literature. The major issue is that absolute driver injury risk cannot be directly estimated from the crash data because data on non-injury crashes (crashes in which neither of the drivers are injured) is not normally available. Besides, data on the speed of the vehicles prior to the crash, which contribute to the injury risk through  $\Delta v$  (see Equation 1), is not available.

One other issue which has not been addressed properly in the analysis of injury risk in two-car crashes is the partial effects of vehicle mass and size. There is generally a high level of correlation between vehicle mass and size (vehicle length or wheelbase has been often used as a proxy for vehicle size in the literature). Many of the studies that have investigated the effect of mass on risk of injury and fatality have not controlled for the effect of vehicle size appropriately; therefore, their estimates could contain the effects of vehicle size as well. There is evidence in the literature suggesting different effects of mass and size on risk of injury and fatality given a crash; however, there are inconsistencies in the results of different studies (for example, see Grime and Hutchinson, 1982; Evans and Wasielewski, 1987; Broughton, 1999; Ross and Wenzel, 2001; Van Auken and Zellner, 2005). The main question, which has remained unclear in the literature, is whether there is any effect of vehicle size above and beyond that of mass ratio (Hutchinson and Anderson, 2009). This is of particular importance because there is the potential to reduce vehicle mass while maintaining its size through various mass-reduction technologies (Wenzel and Ross, 2001).

The study reported in this paper introduces a novel methodology based on a disaggregate analysis of two-car crash data to estimate partial effects of mass, through the velocity change, on absolute driver injury risk in each of the vehicles involved in the crash when absolute injury risk is defined as the probability of injury when the vehicle is involved in a two-car crash. The novel aspect of the introduced methodology is in providing a solution to the issue of the lack of data on the speed of vehicles prior to the crash as well as a solution to the issue of the lack of information on non-injury two-car crashes in national accident data, which has often led to focusing on relative measures of injury risk that are not independent of risk in the colliding cars. Furthermore, the introduced methodology is used to investigate whether there is any effect of vehicle size above and beyond that of mass ratio, and whether there are any effects associated with the gender and age of the drivers. The methodology is then used to analyse two-car collisions to investigate the partial effects of vehicle mass and size on absolute driver injury risk.

## 2. Methodology

As was mentioned in the previous section, in the case of a collision between two vehicles, the velocity change of the vehicles during the collision ( $\Delta v_1$  and  $\Delta v_2$ ) are given by:

$$\Delta v_1 = \left( \frac{m_2}{m_1 + m_2} \right) v \quad (4)$$

$$\Delta v_2 = \left( \frac{m_1}{m_1 + m_2} \right) v, \quad (5)$$

where  $v$  is the closing speed of the vehicles. In the case of a front to front and front to back collision,  $v = v_1 + v_2$  and  $v = v_1 - v_2$  ( $v_1 > v_2$ ), respectively. In the case of a front to side collision at right angle, it can be shown that  $v = \sqrt{v_1^2 + v_2^2}$  (see Appendix 1 for more details).  $\Delta v_1$  and  $\Delta v_2$  in Equations 4 and 5 can be rearranged as the following:

$$\Delta v_1 = v \left( \frac{\mu}{\mu + 1} \right) \quad (6)$$

$$\Delta v_2 = v \left( \frac{1}{\mu + 1} \right). \quad (7)$$

In a two-vehicle collision, the probability of injury of the driver of vehicle 1,  $p_1(v)$ , increases with closing speed  $v$  and with increasing the value of mass ratio  $\mu$  while the probability of injury of the driver of vehicle 2,  $p_2(v)$ , increases with closing speed  $v$  and with decreasing the value of mass ratio  $\mu$ . One of the functional forms having the appropriate properties to describe  $p_1(v)$  and  $p_2(v)$ , both of which range between zero and one, is the logistic function; this was therefore chosen in this study to describe  $p_1(v)$  and  $p_2(v)$  as

$$p_1(v) = \frac{\exp[C_1 + \beta \Delta v_1]}{1 + \exp[C_1 + \beta \Delta v_1]} = \frac{\exp\left[C_1 + \beta v \left(\frac{\mu}{\mu + 1}\right)\right]}{1 + \exp\left[C_1 + \beta v \left(\frac{\mu}{\mu + 1}\right)\right]} \quad (8)$$

and

$$p_2(v) = \frac{\exp[C_2 + \beta \Delta v_2]}{1 + \exp[C_2 + \beta \Delta v_2]} = \frac{\exp\left[C_2 + \beta v \left(\frac{1}{\mu + 1}\right)\right]}{1 + \exp\left[C_2 + \beta v \left(\frac{1}{\mu + 1}\right)\right]}. \quad (9)$$

In these equations,  $C_1$  and  $C_2$  represent the characteristics of the driver (age, gender, etc) and the vehicle (dimensions, make, model, etc) that could contribute to the driver injury risk.  $C_1$  and  $C_2$  can be expressed as the following:

$$C_j = \sum_i \alpha_{ji} x_{ji}; \quad j = 1, 2 \text{ \& } i = 0, 1, 2, \dots \text{ \& } x_{j0} = 1 \quad (10)$$

where  $j$  denotes the driver number (1 or 2),

$\{x_{ji}\}$  is a set of driver and vehicle characteristics for vehicle  $j$ , and

$\{a_{ji}\}$  is a set of parameters to be estimated in the model fitting process.

As can be seen from Equations 8 and 9, the probability of driver injury in each vehicle depends on the closing speed  $v$ . As mentioned earlier, the main difficulty associated with the analysis of injury risk in two-car crashes arises because the vehicles' speeds immediately prior to the crash ( $v_1$  and  $v_2$ ) are not usually observed; therefore, the closing speed  $v$  is rarely known.

Suppose  $f(v)$  represents the probability distribution of closing speed  $v$  that is generally characterised by a mean  $m$  and a vector  $\phi$  of constant parameters depending on the types of distribution (such as standard deviation for the case of normal and log-normal distribution, scale and shape for the case of Gamma distribution, etc). Then the overall probabilities of any collision resulting in injury of the drivers of vehicles 1 and 2 are:

$$P_1 = \int_v p_1(v)f(v)dv \quad (11)$$

and

$$P_2 = \int_v p_2(v)f(v)dv . \quad (12)$$

There are four possible driver injury outcomes of any two-vehicle collision depending on the driver injury outcome of each vehicle. Since the probability of injury depends on the closing speed  $v$ , and  $v$  is common to the two vehicles in the collision, the two events of driver 1 and driver 2 being injured are jointly dependent. Therefore the four possible injury outcomes are in fact joint injury probabilities as shown in Table 1. In this case, we expect that if the driver of vehicle 1 is injured, it is more likely that the driver of vehicle 2 is injured too.

Table 1: Possible joint injury outcomes of a two-vehicle collision

|                      | Driver 1 not injured                               | Driver 1 injured                             |
|----------------------|--|--|
| Driver 2 not injured | $\pi_{00} = \int_v (1 - p_1(v))(1 - p_2(v))f(v)dv$ | $\pi_{10} = \int_v p_1(v)(1 - p_2(v))f(v)dv$ |
| Driver 2 injured     | $\pi_{01} = \int_v (1 - p_1(v))p_2(v)f(v)dv$       | $\pi_{11} = \int_v p_1(v)p_2(v)f(v)dv$       |

According to Equations 8 to 10, the  $\pi_{ij}$  are functions of the parameter  $\beta$ , the parameters  $\alpha$  relating to the vehicle and driver characteristics, and the parameters  $m$  and  $\phi$  characterising the distribution of closing speed  $f(v)$ . Since the speeds of the vehicles prior to the collision

are not observed,  $m$  and  $\phi$  are nuisance parameters that will be estimated in the model fitting process.

It is noted that  $f(v)$  is the prior distribution of closing speed in two-car crashes. When  $f(v)$  is combined with the probability of injury outcome conditional on  $v$  ( $p_1(v)$ ), we get the unconditional injury outcomes  $\pi_{ij}$ . Bayes' Theorem can then be used to determine the posterior speed distribution; that is, the distribution of speed  $v$ , given any of the injury outcomes  $f(v|x)$ , where  $x = 0$  or  $1$  indicates non-injury and injury respectively. In a simplified example where there is only one vehicle in the crash, Bayes' Theorem implies that

$$f(v|x = 1) = \frac{\text{prob}(x=1|v)f(v)}{P(x=1)} = \frac{p(v)f(v)}{P(x=1)} \quad (13)$$

where  $\text{prob}(x = 1|v)$  is the conditional probability of injury given speed  $v$ ,  $P(x = 1)$  is the combined probability of injury irrespective of speed  $v$ , and  $f(v|x = 1)$  is the posterior distribution of  $v$  given injury ( $x = 1$ ). Similarly, the posterior distribution of  $v$  given no injury ( $x = 0$ ) is given by

$$f(v|x = 0) = \frac{\text{prob}(x=0|v)f(v)}{P(x=0)} = \frac{(1-p(v))f(v)}{P(x=0)} \quad (14)$$

The posterior distribution of  $v$  in two-car crashes for any injury outcome shown in Table 1 can be determined in the same way; these will be different for the different cells of the injury outcomes matrix. This is shown below using a simple numerical example.

Suppose that in a two-car collision between two cars of the same mass ( $\mu = 1$ ), the values of parameters  $C_1$ ,  $C_2$ , and  $\beta$  in Equations 8 and 9 are respectively  $-9$ ,  $-9$ , and  $0.16$ . Also, suppose that  $v$  has a Normal distribution with mean  $100$  and standard deviation  $20$ . It can be shown that the joint injury outcomes in Table 1 will be  $\pi_{00}=0.51$ ,  $\pi_{01}=0.16$ ,  $\pi_{10}=0.16$ ,  $\pi_{11}=0.17$ . Figure 1 and 2 show, respectively, the prior speed distribution and the posterior speed distributions for each injury outcome; these are different distributions with different values of the mean.

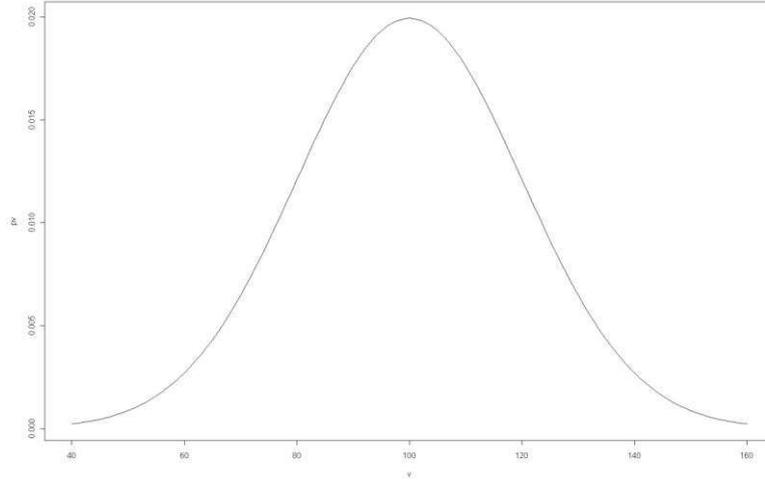


Figure 1: Prior speed distribution

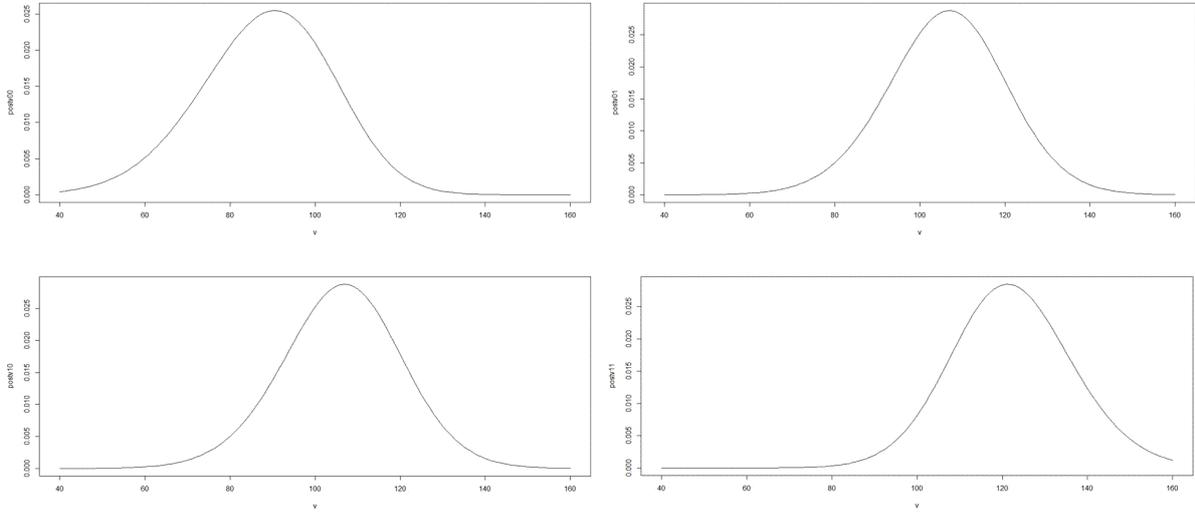


Figure 2: Posterior speed distributions for different injury outcomes shown in Table 1

The other difficulty in estimating the absolute driver injury risk ( $P_1$  and  $P_2$ ) is the fact that no observation is available if there are no injuries ( $\pi_{00}$  is unknown). However, the conditional driver injury risk, defined as the probability of driver injury in a collision in which there is at least one driver injury, can be calculated directly from the observed data. The conditional joint injury probabilities are defined as below:

$$q_{ij} = Prob(Z_1 = i, Z_2 = j \mid i = 0,1 \& j = 0,1 \& i + j \geq 1) = \frac{\pi_{ij}}{1 - \pi_{00}} \quad (15)$$

where  $Z$  for each driver represents the binary injury outcome (0 = no injury, 1 = injury). Closing speed  $v$  determines the “severity” of the collision (measured by  $i + j$ ) whilst mass ratio  $\mu$  determines the “imbalance” between the injuries of the two drivers (measured by  $|i - j|$ ). The three possible conditional joint injury outcomes as defined by Equation 15 are shown in Table 2. The observed values of these are available from the crash data.

Table 2: Conditional joint injury outcomes of a two-vehicle collision

|                      | Driver 1 not injured             | Driver 1 injured                 |
|----------------------|----------------------------------|----------------------------------|
| Driver 2 not injured | -                                | $q_{10} = \pi_{10}/1 - \pi_{00}$ |
| Driver 2 injured     | $q_{01} = \pi_{01}/1 - \pi_{00}$ | $q_{11} = \pi_{11}/1 - \pi_{00}$ |

## 2.1 Maximum likelihood estimation of parameters

The joint injury probabilities  $\pi_{ij}$  shown earlier in Table 1 can be formed using Equations 8 to 12 as a function of unknown parameters ( $\beta$ ,  $\alpha$ ,  $m$ ,  $\Phi$ ) where a probability density function is assumed for closing speed ( $f(v)$ ). Having formed  $\pi_{ij}$ , the three conditional joint injury probabilities ( $q_{ij}$ ) shown in Table 2 can be described as functions of these unknown parameters:

$$q_{10} = \frac{\int_v \frac{\exp[C_1 + \beta v(\frac{\mu}{\mu+1})] f(v) dv}{\left[1 + \exp[C_1 + \beta v(\frac{\mu}{\mu+1})]\right] \left[1 + \exp[C_2 + \beta v(\frac{1}{\mu+1})]\right]}{1 - \int_v \frac{f(v) dv}{\left[1 + \exp[C_1 + \beta v(\frac{\mu}{\mu+1})]\right] \left[1 + \exp[C_2 + \beta v(\frac{1}{\mu+1})]\right]}} \quad (16)$$

$$q_{01} = \frac{\int_v \frac{\exp[C_2 + \beta v(\frac{1}{\mu+1})] f(v) dv}{\left[1 + \exp[C_1 + \beta v(\frac{\mu}{\mu+1})]\right] \left[1 + \exp[C_2 + \beta v(\frac{1}{\mu+1})]\right]}{1 - \int_v \frac{f(v) dv}{\left[1 + \exp[C_1 + \beta v(\frac{\mu}{\mu+1})]\right] \left[1 + \exp[C_2 + \beta v(\frac{1}{\mu+1})]\right]}} \quad (17)$$

$$q_{11} = \frac{\int_v \frac{\exp[C_1 + \beta v(\frac{\mu}{\mu+1})] \exp[C_2 + \beta v(\frac{1}{\mu+1})] f(v) dv}{\left[1 + \exp[C_1 + \beta v(\frac{\mu}{\mu+1})]\right] \left[1 + \exp[C_2 + \beta v(\frac{1}{\mu+1})]\right]}{1 - \int_v \frac{f(v) dv}{\left[1 + \exp[C_1 + \beta v(\frac{\mu}{\mu+1})]\right] \left[1 + \exp[C_2 + \beta v(\frac{1}{\mu+1})]\right]}} \quad (18)$$

The unit of observation will be two-car collisions with three possible conditional joint injury outcomes. For any values of the parameters the probabilities of the observed conditional joint injury outcomes can be calculated for each collision. By combining these over the whole dataset, the likelihood function can be calculated as the following. For each observation, define,

$$y_{ij} = \begin{cases} 1, & \text{if } Z_1 = i \text{ and } Z_2 = j ; \\ 0, & \text{otherwise;} \end{cases} \quad (19)$$

where  $i$  and  $j$  show, respectively, the binary injury outcome for the drivers 1 and 2 (i.e.  $i = 0, 1, j = 0, 1, i + j \geq 1$ ). The likelihood function over the whole dataset can be calculated using the following:

$$L(\alpha, \beta, m, \phi) = \prod_{n=1}^N [(q_{10}^{y_{10}})_n (q_{01}^{y_{01}})_n (q_{11}^{y_{11}})_n] \quad (20)$$

where  $N$  denotes the total number of records in the dataset and  $n$  refers to the collision number in the data set ( $n = 1, 2, \dots, N$ ). An optimisation algorithm can then be applied to find the values of the parameters that maximise the logarithm of the likelihood function (log-likelihood function) shown below:

$$l(\alpha, \beta, m, \phi) = \sum_{n=1}^N [(y_{10} \log_e(q_{10}))_n + (y_{01} \log_e(q_{01}))_n + (y_{11} \log_e(q_{11}))_n]. \quad (21)$$

## 2.2 Probability distribution of closing speed

As mentioned earlier, a distributional form is required for the closing speed with a given probability density function  $f(v)$  the parameters of which ( $m, \phi$ ) will be estimated in the model estimation process. In this study, two continuous probability distributions were investigated separately to describe the distribution of closing speed. A Normal distribution was investigated first because it is a simple well-described distribution which is defined with only two parameters (mean and standard deviation). However, the disadvantage of the Normal distribution in this case is that it has an unbounded range; hence specific constraints are required on the distribution parameters to ensure the values of closing speed  $v$  remain positive during the model fitting process. Therefore, a log-normal distribution was also investigated for  $v$ . This has the advantageous property that it is bounded below by 0, and therefore it is free from any constraint required to ensure positive values for closing speed  $v$  during the model fitting process. Comparison of model estimation results for the different density functions formulated confirmed that a log-normal distribution of closing speed fits the two-car collisions data substantially better than a Normal distribution (Tolouei, 2011). Therefore in this paper, a log-normal distribution with parameters  $m$  and  $\sigma$  is assumed for the closing speed  $v$  where  $m$  and  $\sigma$  are, respectively, the mean and standard deviation of the associated Normal distribution (i.e.  $m$  and  $\sigma$  are the mean and standard deviation of  $\log_e v$ ).

Evidence suggests that vehicles have different average speeds on different types of roads as classified according to their speed limit. Speeds generally tend to be higher on roads that have a higher speed limit than on those that have a lower speed limit (DfT, 2006; DfT, 2007). Therefore, it can be argued that the distribution of closing speed is different for different

types of roads where the speed limit varies. This can be investigated by including the variable speed limit, which is observed for each collision, in the probability density function of the distributions being investigated. This allows us to specify a different prior distribution for different speed limits. Two cases are therefore investigated as described below.

### 2.2.1 Case A: same distribution for all speed limits

In this case, the same log-normal distribution for  $v$  is assumed for all the collisions irrespective of the speed limit of the road.  $v$  has a log-normal distribution with parameters  $m$  and  $\sigma$ , therefore  $u = \beta v$  has a log-normal distribution with parameters  $m' = m + \log_e \beta$  and  $\sigma$  (i.e.  $m'$  and  $\sigma$  are mean and standard deviation of  $\log_e u$ ). Transforming  $\log_e u$ , which has a normal distribution, to  $\log_e z$ , which has a standard normal distribution with mean 0 and standard deviation 1, results in the following relationship:

$$u = \exp[m' + \sigma \log_e z] = \beta \exp(m) z^\sigma . \quad (22)$$

In this equation, both  $\beta$  and  $m$ , in addition to the standard deviation  $\sigma$ , are unknown parameters. A new variable  $\theta$  is therefore defined as  $\theta = \beta \exp(m)$ . Therefore,

$$u = \theta z^\sigma . \quad (23)$$

The probability density function  $f(u)$  can be expressed based on the unit Normal density function for  $\log_e z$  according to the following:

$$f(u) = g(z) = \frac{1}{z} \phi(\log_e z) = \frac{1}{z\sqrt{2\pi}} \exp[-0.5(\log_e z)^2] \quad (24)$$

where  $\log_e z = \frac{\log_e u - m'}{\sigma}$ . The conditional joint injury probabilities that are required to form the likelihood function in Equation 21 are then calculated according to the following:

$$q_{10} = \frac{\int_z \frac{\exp[C_1 + u(z)\left(\frac{\mu}{\mu+1}\right)] g(z) dz}{\left[1 + \exp\left[C_1 + u(z)\left(\frac{\mu}{\mu+1}\right)\right]\right] \left[1 + \exp\left[C_2 + u(z)\left(\frac{1}{\mu+1}\right)\right]\right]}}{1 - \int_z \frac{g(z) dz}{\left[1 + \exp\left[C_1 + u\left(\frac{\mu}{\mu+1}\right)\right]\right] \left[1 + \exp\left[C_2 + u\left(\frac{1}{\mu+1}\right)\right]\right]}} \quad (25)$$

$$q_{01} = \frac{\int_z \frac{\exp\left[C_2 + u(z)\left(\frac{1}{\mu+1}\right)\right] g(z) dz}{\left[1 + \exp\left[C_1 + u(z)\left(\frac{\mu}{\mu+1}\right)\right]\right] \left[1 + \exp\left[C_2 + u(z)\left(\frac{1}{\mu+1}\right)\right]\right]}}{1 - \int_z \frac{g(z) dz}{\left[1 + \exp\left[C_1 + u(z)\left(\frac{\mu}{\mu+1}\right)\right]\right] \left[1 + \exp\left[C_2 + u(z)\left(\frac{1}{\mu+1}\right)\right]\right]}} \quad (26)$$

$$q_{11} = \frac{\int_z \frac{\exp\left[C_1+u(z)\left(\frac{\mu}{\mu+1}\right)\right] \exp\left[C_2+u(z)\left(\frac{1}{\mu+1}\right)\right] g(z) dz}{\left[1+\exp\left[C_1+u(z)\left(\frac{\mu}{\mu+1}\right)\right]\right] \left[1+\exp\left[C_2+u(z)\left(\frac{1}{\mu+1}\right)\right]\right]}{1-\int_z \frac{g(z) dz}{\left[1+\exp\left[C_1+u(z)\left(\frac{\mu}{\mu+1}\right)\right]\right] \left[1+\exp\left[C_2+u(z)\left(\frac{1}{\mu+1}\right)\right]\right]}} \quad (27)$$

where  $u(z)$  and  $g(z)$  are given by Equations 23 and 24, respectively. The integration is calculated numerically over the standardised normal variable  $z$ .

The parameters  $\alpha$  (representing driver and vehicle factors as shown in Equation 10), and  $\theta$  and  $\sigma$  that maximise Equation 21 are estimated subject only to the following constraint:

$$\sigma \geq 0. \quad (28)$$

The estimated values of these parameters are then used to predict the values of absolute driver injury risk for vehicle 1 and 2 for any given value of mass ratio  $\mu$  using the following equations:

$$P_1 = \int_v p_1(v) f(v) dv = \int_z \frac{\exp\left[C_1+u(z)\left(\frac{\mu}{\mu+1}\right)\right] g(z) dz}{1+\exp\left[C_1+u(z)\left(\frac{\mu}{\mu+1}\right)\right]} \quad (29)$$

$$P_2 = \int_v p_2(v) f(v) dv = \int_z \frac{\exp\left[C_2+u(z)\left(\frac{1}{\mu+1}\right)\right] g(z) dz}{1+\exp\left[C_2+u(z)\left(\frac{1}{\mu+1}\right)\right]}. \quad (30)$$

### 2.2.2 Case B: different distributions for different speed limits

We now consider the possibility that the distribution mean for  $\log_e v$  is related to the speed limit  $s^l$  according to the following:

$$m = \log_e(\omega s^l). \quad (31)$$

Replacing this in Equation 22 gives

$$u = \beta \omega z^\sigma s^l. \quad (32)$$

In this equations, both  $\beta$  and  $\omega$ , as well as the standard deviation  $\sigma$ , are unknown parameters. A new variable  $t$  is therefore defined as  $t = \beta \omega$ . Equation 32 can be rewritten as

$$u = tz^\sigma s^l. \quad (33)$$

Therefore, the only difference between this case and the formulation shown in Case A is that instead of parameter  $\theta$ , the parameter  $t$  is estimated which is used together with the estimated value of  $\sigma$  as well as the observed value of speed limit  $s^l$  to characterise the distribution of closing speed.

### 3. Two-car crash dataset

The methodology just introduced is now used to analyse driver injury risk in frontal two-car crashes. The data used is based on the UK STATS19 Police reported data which includes all road accidents that involve personal injury. Basic STATS19 data is publicly available for use as in annual basis<sup>1</sup>. Data from 2000 to 2006 was used to extract frontal two-car crashes in which at least one of the drivers was injured. Frontal crashes alone constitute about 43% of all crashes in the dataset. This is probably because these are high severity crashes resulting in a greater number of KSI drivers.

STATS19 data contains information on the age and gender of each driver, and the severity of their injury. Unfortunately, it does not include data on vehicle mass. However, the Vehicle Registration Mark (VRM) for each vehicle involved in a crash is recorded. This provides the opportunity to link the STATS19 vehicle data to vehicle makes and models. A separate dataset, developed by the UK Department for Transport (DfT) includes make and model data on each of the vehicles involved in the accidents where the data is available. This dataset was linked to the basic two-car crash dataset to include make and model information for each car involved in a two-car crash. Due to a considerable proportion of missing data on makes and models, however, only about 66% of two-car crashes included make and model information for both of the cars.

Investigation to find an available vehicle mass dataset in Great Britain was not successful. Therefore, detailed data on kerb mass for all the variants of popular makes and models were extracted from the UK Car Magazine<sup>2</sup>, which is a web-based data source, using a computer programme. It is notable that mass data is different for different variants of makes and models stratified by engine size, year of manufacture, and engine specifications. Having downloaded available mass data for all popular makes and models, another computer programme was used to assign mass to each make and model in the two-car crash dataset. This could be done with a high level of accuracy as the VRM data includes information on engine size, year of manufacture, and engine specification for each make and model. Due to a high proportion of missing mass data for makes and models involved in two-car crashes, especially older cars, only about 21% of all two-car crashes in the STATS19 data (about 85000 crashes) included mass data for both of the cars involved in the crash.

---

<sup>1</sup> Department for Transport, Local Government and the Regions, Road Accident Data, 2000-2006 [computer file]. Colchester, Essex: UK Data Archive [distributor], <http://www.data-archive.ac.uk>.

<sup>2</sup> <http://data.carmagazine.co.uk/cars/specs/>

There are three levels of casualty severity in STATS19 data: killed (within 30 days as a result of sustained injury), seriously injured, and slightly injured. The main analysis of two-car crashes was performed for serious or fatal injuries only due to the greater importance of these injuries. This is consistent with the similar studies based on STATS19 data (for example see Broughton, 1996c; Broughton, 2007). Therefore, the final sample dataset included two-car crashes where at least one of the drivers was either Killed or Seriously Injured (KSI); this included a total of 2485 two-car crashes.

Descriptive statistics of vehicle design variables in the dataset (mass, length, width, height, and wheelbase) are shown in Table 3 (the maximum number of observations in this table is twice the total number of collisions because there are two vehicles per collision). The average vehicle mass in the dataset is 1150 kg; it ranges from 690 kg (for a variant of Citroen AX) to about 2600 kg (for a variant of Land Rover Range Rover). Vehicle length varies from 270 cm (for a variant of Volkswagen Polo) to 516 cm (for a variant of Mercedes S class) with an average of 413 cm. As the statistics for “Length” and “Wheelbase” suggest, there is more variation in “Length” compared to “Wheelbase” in the dataset. This suggests that “Length” is preferred to “Wheelbase” as a variable that represents the vehicle size. The average width of vehicles in the dataset is 178 cm where the minimum and maximum widths belong to variants of Renault Clio and Ford Mondeo, respectively.

Table 3: Descriptive statistics of vehicle design variables in the dataset (injury level: KSI)

| Vehicle variable | Descriptive statistics |      |      |                |      |
|------------------|------------------------|------|------|----------------|------|
|                  | Min                    | Mean | Max  | Std. Deviation | Obs. |
| Mass (kg)        | 690                    | 1150 | 2599 | 250            | 4970 |
| Length (cm)      | 270                    | 413  | 516  | 36             | 4970 |
| Width (cm)       | 142                    | 178  | 223  | 15             | 4970 |
| Height (cm)      | 122                    | 143  | 194  | 9              | 4302 |
| Wheelbase (cm)   | 142                    | 255  | 448  | 13             | 4970 |

Driver factors in a vehicle that is involved in a two-vehicle crash can potentially contribute to the risk of injury to the driver of that vehicle (through correlation with the physical strength of the driver) as well as to that to the driver of the colliding vehicle (through influencing driving style and aggressivity). The distribution of drivers involved in frontal injury crashes by age and gender are reflected in Table 4. The category with the largest number of records in the dataset is that of male drivers aged 35-54.

Table 4: Distribution of drivers by age and gender in the dataset (injury level: KSI)

| Driver age | Driver gender |        |         | Total |
|------------|---------------|--------|---------|-------|
|            | Male          | Female | Unknown |       |
| 17-24      | 712           | 315    | 1       | 1028  |
| 25-34      | 666           | 390    | 1       | 1057  |
| 35-54      | 1099          | 642    | 0       | 1741  |
| +55        | 680           | 330    | 0       | 1010  |
| Unknown    | 60            | 28     | 46      | 134   |
| Total      | 3217          | 1705   | 48      | 4970  |

#### 4. Injury risk modelling

The effects of different factors on driver injury probability (Equations 29 and 30) in frontal two-car collisions were estimated by forming and maximising the log-likelihood function, described by Equation 21, over the two-car crash dataset. Two different distributional assumptions, defined and formulated in detail in Section 2 (Case A and Case B), are investigated for the closing speed  $v$ ; these are summarised in Table 5.

Table 5: Summary of the defined distributions for closing speed  $v$

| Distribution of $v$ | Case A (distribution is independent of speed limit $v^l$ )  | Case B (distribution is dependent on speed limit $v^l$ )   |
|---------------------|---|--|
| Log-normal          | $v \sim \text{LogN}(m, \sigma)$<br>$u = \beta v \sim \text{LogN}(m + \log_e \beta, \sigma)$<br>$u = \beta \exp(m) z^\sigma$<br>$u = \theta z^\sigma$<br>$f(u) = 1/z \phi(\log_e z)$<br><br>Unknown parameters: $\theta, \sigma$ | $v \sim \text{LogN}(m, \sigma)$<br>$u = \beta v \sim \text{LogN}(m + \log_e \beta, \sigma)$<br>$m = \log_e(\omega s^l)$<br>$u = \beta \omega z^\sigma s^l$<br>$u = t z^\sigma s^l$<br>$f(u) = 1/z \phi(\log_e z)$<br><br>Unknown parameters: $t, \sigma$ |

$v^l$ : speed limit

The maximum likelihood estimation was performed using the ‘‘R’’ package<sup>1</sup>. In calculating the log-likelihood function value, the integration in Equations 25 to 27 was performed numerically using Simpson’s rule (Moin, 2001).

##### 4.1 Maximum likelihood estimation results

In the model estimation process for frontal collisions, vehicles 1 and 2 had the same labels (vehicle 1, vehicle 2) as those in the original STATS19 data. These are believed to be arbitrary). In the first step, the simplest model form that includes no driver or vehicle effects except mass ratio  $\mu$  (i.e.  $C_1 = C_2 = \alpha_0$ ) was estimated for the two distributional assumptions (Case A and Case B in Table 5) to find the distributional form that led to the best description

<sup>1</sup> R Development Core Team, 2005. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>.

of the injury severity distribution. Therefore there were three parameters to estimate for each distribution form: the constant  $\alpha_0$  that represents  $C_1$  and  $C_2$  in Equations 29 and 30, and two parameters that describe the closing speed distribution as summarised in Table 5. The maximum likelihood estimation results are shown in Table 6. In this table, the models have the same labels as their assumed distribution equivalents. Figure 1 shows how the maximised log-likelihood value varies by the standard deviation of logarithm of closing speed ( $\sigma$ ).

Table 6: Maximum likelihood estimation results: Log-normal distribution of  $v$

| Parameters | Model A0 |            |       |       | Model B0 |            |       |       |
|------------|----------|------------|-------|-------|----------|------------|-------|-------|
|            | Est.     | Std. Error | CI-   | CI+   | Est.     | Std. Error | CI-   | CI+   |
| $\alpha_0$ | -4.39    | 0.60       | -5.58 | -3.21 | -5.18    | 0.28       | -5.73 | -4.63 |
| $\theta$   | 5.00     | 1.46       | 2.13  | 7.87  | -        | -          | -     | -     |
| $t$        | -        | -          | -     | -     | 0.07     | 0.01       | 0.05  | 0.09  |
| $\sigma$   | 0.47     | 0.22       | 0.05  | 0.90  | 0.73     | 0.06       | 0.62  | 0.85  |
| LL value   | -2507.84 |            |       |       | -2443.42 |            |       |       |
| AIC        | 2514     |            |       |       | 2449     |            |       |       |
| Obs        | 2485     |            |       |       | 2485     |            |       |       |

CI- and CI+ show, respectively, lower and upper bound of 95% confidence interval of the estimates

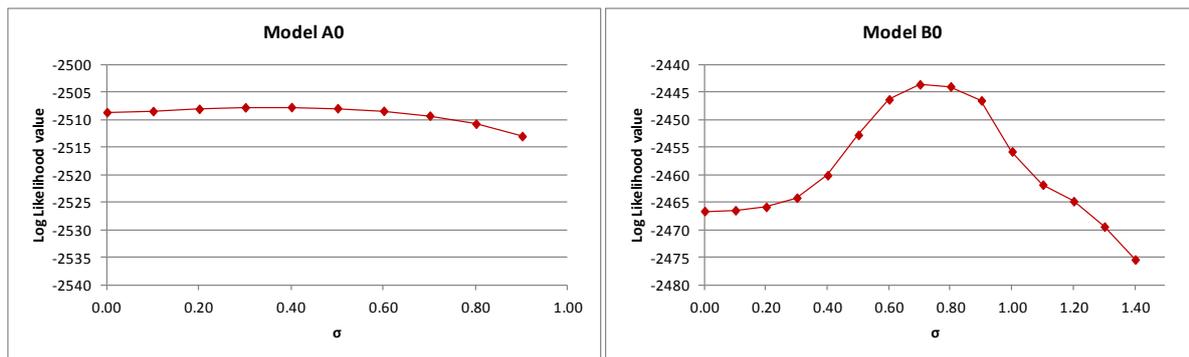


Figure 3: Maximised log-likelihood versus standard deviation of logarithm of closing speed ( $\sigma$ )

The results show that the best model with highest log-likelihood value is Model B0, in which all the estimated parameters are statistically significant and the log-likelihood is substantially higher than that for Model A0. According to this model, the closing speed has a log-normal distribution with a mean value that depends on the speed limit ( $m = \log_e(\omega s^l)$ ). The spread of the distribution of closing speed is described by the single parameter  $\sigma$ , which in the log-normal distribution gives the relationship to the mean speed:  $Var(v) = [E(v)]^2[\exp(\sigma^2) - 1]$ . This model, not surprisingly, fitted the data substantially better than the corresponding one that did not use speed limit in the model for the distribution of closing speed. Therefore, Model B0 was chosen as the one to be expanded to include further explanatory variables.

The variables related to driver age and driver gender were added to Model B0 to investigate the effects of these factors. These contribute to  $C_1$  and  $C_2$  in Equations 29 and 30 as shown by Equation 10. The first hypothesis was that injury risk to the driver in each vehicle is influenced by the physical condition of the driver as represented by driver age and gender. These variables were added to Model B0 to test this hypothesis.  $C_1$  and  $C_2$  for this model, labelled as Model B1, are described as the following:

$$C_1 = \sum_i \alpha_{1i}x_{1i}; i = 0,1,2, \dots \& x_{10} = 1 \quad (34)$$

$$C_2 = \sum_i \alpha_{2i}x_{2i}; i = 0,1,2, \dots \& x_{20} = 1. \quad (35)$$

The maximum likelihood estimation results for Model B1 are shown in Table 7. The widely-used dummy coding method has been used to code the categorical variables relating to driver age and gender that were shown in Table 4 where male driver and the age category 35-54 are taken as the reference categories.

Table 7: Maximum likelihood estimation results: adding the effects of driver characteristics

| Parameters                              | Model B1 |            |        |        |
|---|----------|------------|--------|--------|
|   | Est.     | Std. Error | CI-    | CI+    |
| $\alpha_{10} = \alpha_{20}$             | -5.178   | 0.268      | -5.703 | -4.653 |
| $\alpha_{11} = \alpha_{21}$ (Female)    | -0.417   | 0.080      | -0.574 | -0.260 |
| $\alpha_{12} = \alpha_{22}$ (Age 17-24) | -0.241   | 0.102      | -0.441 | -0.040 |
| $\alpha_{13} = \alpha_{23}$ (Age 25-34) | -0.190   | 0.100      | -0.385 | 0.006  |
| $\alpha_{14} = \alpha_{24}$ (Age +55)   | 0.646    | 0.100      | 0.449  | 0.843  |
| $t$                                     | 0.060    | 0.010      | 0.041  | 0.079  |
| $\sigma$                                | 0.797    | 0.061      | 0.677  | 0.917  |
| LL value                                | -2393.72 |            |        |        |
| AIC                                     | 2408     |            |        |        |
| Obs                                     | 2485     |            |        |        |

CI- and CI+ show, respectively, lower and upper bound of 95% confidence interval of the estimates

It was then hypothesised that injury risk to the driver in each vehicle is influenced not only by the physical vulnerability of its driver, but also by the driving style of the driver of the colliding vehicle as represented by variables age and gender (e.g. the effect of more aggressive driving). Therefore in the new model, labelled as Model B2,  $C_1$  and  $C_2$  are described as the following:

$$C_1 = \sum_i \alpha_{1i}x_{1i} + \alpha_{2i}x_{2i}; i = 0,1,2, \dots \& x_{10} = x_{20} = 1 \quad (36)$$

$$C_2 = \sum_i \alpha_{1i}x_{2i} + \alpha_{2i}x_{1i}; i = 0,1,2, \dots \& x_{10} = x_{20} = 1. \quad (37)$$

The estimation results for Model B2 are shown in Table 8.

Table 8: Maximum likelihood estimation results: adding the effects of driver characteristics and driving style

| Parameters                  | Model B2  |            |        |        |
|-----------------------------|-----------|------------|--------|--------|
|                             | Est.      | Std. Error | CI-    | CI+    |
| $\alpha_{10} = \alpha_{20}$ | -5.185    | 0.581      | -6.324 | -4.046 |
| $\alpha_{11}$ (Female)      | 0.320     | 0.320      | -0.307 | 0.946  |
| $\alpha_{12}$ (Age 17-24)   | -0.001    | 0.378      | -0.742 | 0.740  |
| $\alpha_{13}$ (Age 25-34)   | -0.150    | 0.391      | -0.916 | 0.616  |
| $\alpha_{14}$ (Age +55)     | -0.410    | 0.418      | -1.230 | 0.410  |
| $\alpha_{21}$ (Female)      | 0.718     | 0.316      | 0.099  | 1.337  |
| $\alpha_{22}$ (Age 17-24)   | 0.189     | 0.366      | -0.528 | 0.906  |
| $\alpha_{23}$ (Age 25-34)   | 0.023     | 0.394      | -0.751 | 0.796  |
| $\alpha_{24}$ (Age +55)     | -1.063    | 0.425      | -1.895 | -0.230 |
| $t$                         | 0.086     | 0.009      | 0.069  | 0.103  |
| $\sigma$                    | 0.629     | 0.051      | 0.528  | 0.729  |
| LL value                    | -2393.724 |            |        |        |
| AIC                         | 2416      |            |        |        |
| Obs                         | 2485      |            |        |        |

CI- and CI+ show, respectively, lower and upper bound of 95% confidence interval of the estimates

Comparison of the maximum likelihood estimation results for Models B0, B1, and B2 reflected in Tables 6 to 8 shows that Model B1 has the best goodness of fit (measured by log-likelihood through the AIC) as well as the estimated parameters that are statistically significant (except for  $\alpha_{13} = \alpha_{23}$  which shows that there is no difference between age range 25-34 and the reference category of 35-54). Therefore it is the model that best represents the effects of drivers' age and gender on driver injury probability in two-car crashes. Thus the set of data used in this study does not show any age and gender-specific effect on driving style that influences injury risk in the colliding vehicles.

One of the fundamental questions in the analysis of injury risk in two-car crashes which has remained unclear is whether there is any effect of vehicles' size beyond the effect of mass ratio. In order to examine this, the variables related to vehicle size were added to Model B1 as explanatory variables (see Equations 34 and 35). Two models were estimated: Model B3 in which vehicle size is represented by "vehicle length" (m), and Model B4 in which vehicle size is represented by "vehicle length  $\times$  vehicle width" (m<sup>2</sup>)<sup>1</sup>. The results, reflected in Table 9, shows that both models have a better goodness of fit than that of Model B1 (measured by log-likelihood through the AIC) as well as statistically significant estimated coefficients for the variable "Size". This confirms that there is an effect of vehicle size above and beyond that of mass ratio in frontal collisions. The negative coefficient of size in these models, which are statistically significant at the 5% level, shows that vehicle size is protective. The goodness of

<sup>1</sup> Due to a substantial number of missing data on vehicle height, the effect of volume is not examined in these models.

fit of Model B4 is significantly better than that of Model B3. Therefore Model B4, in which vehicle size is represented by “vehicle length  $\times$  vehicle width”, is the model that best reflects the partial effects of different contributing factors on driver injury probability in frontal crashes where injury is defined by either fatality or serious injury. It should be noted that vehicle age was also included in the model as a variable to investigate its effect on injury risk. The results showed that the estimated coefficient of vehicle age was not statistically significant and the model fit was not improved, therefore it was excluded from the final model.

Table 9: Maximum likelihood estimation results: adding the effects of vehicle size

| Parameters                              | Model B3 |            |        |        | Model B4 |            |        |        |
|---|----------|------------|--------|--------|----------|------------|--------|--------|
|   | Est.     | Std. Error | CI-    | CI+    | Est.     | Std. Error | CI-    | CI+    |
| $\alpha_{10} = \alpha_{20}$             | -3.002   | 1.006      | -4.973 | -1.031 | -2.982   | 0.691      | -4.336 | -1.627 |
| $\alpha_{11} = \alpha_{21}$ (Female)    | -0.388   | 0.081      | -0.547 | -0.230 | -0.401   | 0.081      | -0.559 | -0.243 |
| $\alpha_{12} = \alpha_{22}$ (Age 17-24) | -0.283   | 0.104      | -0.486 | -0.080 | -0.274   | 0.103      | -0.477 | -0.072 |
| $\alpha_{13} = \alpha_{23}$ (Age 25-34) | -0.206   | 0.100      | -0.403 | -0.010 | -0.194   | 0.101      | -0.391 | 0.003  |
| $\alpha_{14} = \alpha_{24}$ (Age +55)   | 0.625    | 0.101      | 0.427  | 0.822  | 0.635    | 0.101      | 0.438  | 0.833  |
| $\alpha_{15} = \alpha_{25}$ (Size)      | -0.401   | 0.174      | -0.741 | -0.060 | -0.204   | 0.056      | -0.314 | -0.094 |
| $t$                                     | 0.047    | 0.011      | 0.025  | 0.068  | 0.038    | 0.011      | 0.016  | 0.059  |
| $\sigma$                                | 0.899    | 0.092      | 0.719  | 1.078  | 0.983    | 0.115      | 0.758  | 1.209  |
| LL value                                | -2389.86 |            |        |        | -2383.36 |            |        |        |
| AIC                                     | 2406     |            |        |        | 2399     |            |        |        |
| Obs                                     | 2485     |            |        |        | 2485     |            |        |        |

CI- and CI+ show, respectively, lower and upper bound of 95% confidence interval of the estimates

As mentioned above, the assumption of log-normal distribution for closing speed was made on the basis of its simplicity and advantageous properties over Normal distribution. It was also mentioned that comparison of model estimation results for the different density functions formulated confirmed that a log-normal distribution of closing speed fits the two-car collisions data substantially better than a Normal distribution (Tolouei, 2011). In order to investigate the sensitivity of the estimated parameters on the distributional assumption, we also assumed a Gamma distribution for the closing speed and fitted the final model (Model B4) using this assumption to compare the estimated effects of different variables. The results showed that the estimated effects of different variables on the injury risk are reasonably similar (in terms of values of estimates and significance of effects) for the two distributional assumptions considered. This suggests that the conclusions from the modelling are not unduly sensitive to the assumption of shape of  $f(v)$ .

The estimated values of parameters in Model B4 were used to predict driver injury probabilities for different values of the explanatory variables using Equations 29 and 30 when in these equations,  $C_1$  and  $C_2$  are given by Equations 34 and 35, and  $u(z)$  and  $g(z)$  are given by Equations 23 and 24, respectively. As was mentioned earlier, the integration is calculated numerically using Simpson's rule<sup>1</sup> over the values of  $z$  in the interval  $[0,40]$  with increments of 0.01.

## 4.2 Effects of vehicle mass

The estimated injury probabilities for a few examples of two-car crashes are shown in Table 10 where examples are defined depending on the values of mass ratio  $\mu$  and speed limit; these are for crashes with drivers in the reference category (male drivers aged 35-54). As the model estimation results showed, the driver injury probabilities ( $P_1$  and  $P_2$ ) are influenced not only by mass ratio, but also by "Size" of the vehicles in a collisions. Therefore, the estimated values of  $P_1$  and  $P_2$  can be different for a given value of  $\mu$  depending on the dimensions of the vehicles. The relationship between vehicle mass and "Size" in the dataset is shown in Figure 2 when "Size" is defined by "vehicle length  $\times$  vehicle width". The trend in the data is closer to an exponential function than a linear one. In the two-car crash examples in Table 10, an average value of "Size" is calculated for the given values of mass using the relationship shown in Figure 1; these are used in estimating  $P_1$  and  $P_2$ .

Table 10: The effect of mass ratio ( $\mu$ ) on injury probabilities ( $P_1$  and  $P_2$ ) in frontal collisions

| Crash                         | $\mu$ ( $m_2/m_1$ ) | Speed limit | $P_1$ | $P_2$ | $R=P_1/P_2$ |
|-------------------------------|---------------------|-------------|-------|-------|-------------|
| 1<br>$m_1=1000$<br>$m_2=1000$ | 1.0                 | 40          | 0.079 | 0.079 | 1.00        |
| 2<br>$m_1=1000$<br>$m_2=1500$ | 1.5                 | 40          | 0.101 | 0.043 | 2.38        |
| 3<br>$m_1=1000$<br>$m_2=2000$ | 2.0                 | 40          | 0.116 | 0.025 | 4.56        |
| 4<br>$m_1=1000$<br>$m_2=1000$ | 1.0                 | 60          | 0.135 | 0.135 | 1.00        |
| 5<br>$m_1=1000$<br>$m_2=1500$ | 1.5                 | 60          | 0.170 | 0.078 | 2.19        |
| 6<br>$m_1=1000$<br>$m_2=2000$ | 2.0                 | 60          | 0.194 | 0.048 | 4.00        |

The results suggest that, for example, if two cars with a similar mass (1000 kg) crash into each other on a road where the speed limit is 60 mile/hr, the probability of each driver being killed or seriously injured is about 13.5%. However, if car 2 had a mass twice that of car 1 (1000 kg compared to 2000 kg), the probability of driver of car 1 (the lighter car) being killed

<sup>1</sup> See, for example, (Moin, 2001) for a description of Simpson's rule

or seriously injured would increase to about 19.4% while the probability of driver of car 2 (the heavier car) being killed or seriously injured would decrease to about 4.8%. These results are consistent with vehicle mass having both protective and aggressive effects in two-car crashes. The results also show that, in general, the probability of injury increases with speed limit; this represents the effect of the closing speed of the vehicles involved in the collision on driver injury probability.

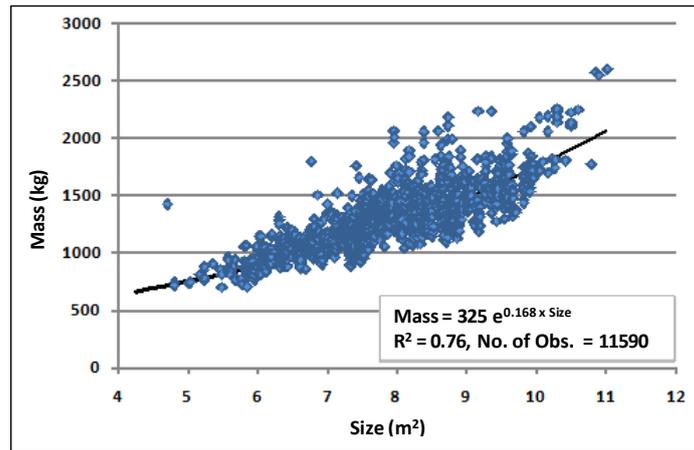


Figure 4: Relationship between vehicle mass and size (Length  $\times$  Width)

### 4.3 Effects of vehicle size

It was shown that vehicle size has a protective effect over and above the effect of mass ratio in frontal two-car crashes. It was also shown that the best variable representing the effect of vehicle size (of those tested) is the product of vehicle length and vehicle width. The estimated effects of vehicle size, based on the estimated Model B4, are shown for a few examples of frontal two-car crashes in Table 11; these are for crashes with drivers in the reference category (male drivers aged 35-54).

Table 11: The effects of vehicle mass (kg) and vehicle size (Length  $\times$  Width (m<sup>2</sup>)) on injury probabilities (P<sub>1</sub> and P<sub>2</sub>) in frontal collisions

| Crash   | $\mu$ (m <sub>2</sub> /m <sub>1</sub> ) | Speed limit | P <sub>1</sub> | P <sub>2</sub> | R=P <sub>1</sub> /P <sub>2</sub> |
|---|---|-------------|----------------|----------------|----------------------------------|
| 1<br>m <sub>1</sub> =1000 Size <sub>1</sub> =6<br>m <sub>2</sub> =1000 Size <sub>2</sub> =6 | 1.0                                     | 60          | 0.145          | 0.145          | 1.00                             |
| 2<br>m <sub>1</sub> =1000 Size <sub>1</sub> =6<br>m <sub>2</sub> =1000 Size <sub>2</sub> =7 | 1.0                                     | 60          | 0.145          | 0.131          | 1.10                             |
| 3<br>m <sub>1</sub> =1000 Size <sub>1</sub> =6<br>m <sub>2</sub> =1000 Size <sub>2</sub> =8 | 1.0                                     | 60          | 0.145          | 0.119          | 1.21                             |
| 4<br>m <sub>1</sub> =1000 Size <sub>1</sub> =6<br>m <sub>2</sub> =1500 Size <sub>2</sub> =9 | 1.5                                     | 60          | 0.181          | 0.079          | 2.30                             |
| 5<br>m <sub>1</sub> =1000 Size <sub>1</sub> =7<br>m <sub>2</sub> =1500 Size <sub>2</sub> =9 | 1.5                                     | 60          | 0.166          | 0.079          | 2.11                             |
| 6<br>m <sub>1</sub> =1000 Size <sub>1</sub> =8<br>m <sub>2</sub> =1500 Size <sub>2</sub> =9 | 1.5                                     | 60          | 0.152          | 0.079          | 1.94                             |

The results are shown for two sets of mass ratios (1.0 and 1.5) where the size of one of the cars varies while all other factors including mass of the two cars are kept constant. Comparison of the estimated values of  $P_1$  and  $P_2$  for crashes 1 to 3 in Table 11 shows that increasing “Size” for car 2 from 6 m<sup>2</sup> to 8 m<sup>2</sup>, when its mass is constant (1000 kg), decreases the probability of its driver being killed or seriously injured from about 14.5% to about 11.9%. On the other hand, in a frontal crash where mass ratio is 1.5 (1000 kg compared to 1500 kg), increasing “Size” for the lighter car (car 1) from 6 m<sup>2</sup> to 8 m<sup>2</sup> decreases its driver injury probability from about 18.1% to about 15.2% without affecting the driver injury probability of car 2.

The findings on the effects of vehicle size is important from the policy point of view because the relationship between mass and size reflected in Figure 2 suggests that there is the potential to make changes to vehicle design to increase the size of vehicles while vehicle mass is maintained. This could increase the safety performance of a vehicle without any adverse impact on the safety performance of the other vehicles in the fleet.

#### 4.4 Effects of driver factors

The estimated coefficients of driver age and driver gender variables were used to estimate their partial effects on driver injury risks in both vehicles using the same methodology that was used to estimate the partial effects of mass ratio and vehicle size. The results for a few examples of frontal collisions where they are different only in a driver factor are shown in Table 12.

Table 12: The effects of driver age and gender on injury probabilities ( $P_1$  and  $P_2$ ) in frontal collisions

| Crash   | $\mu$ ( $m_2/m_1$ ) | Speed limit | $P_1$ | $P_2$ | $R=P_1/P_2$ |
|---|---------------------|-------------|-------|-------|-------------|
| 1<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =male aged 35-54<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =male aged 17-24     | 1.0                 | 60          | 0.135 | 0.119 | 1.13        |
| 2<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =male aged 35-54<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =male aged +55       | 1.0                 | 60          | 0.135 | 0.183 | 0.74        |
| 3<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =female aged 35-54<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =female aged 17-24 | 1.0                 | 60          | 0.112 | 0.099 | 1.13        |
| 4<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =female aged 35-54<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =female aged +55   | 1.0                 | 60          | 0.112 | 0.151 | 0.74        |
| 5<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =male aged 17-24<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =female aged 17-24   | 1.0                 | 60          | 0.119 | 0.099 | 1.20        |
| 6<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =male aged 35-54<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =female aged 35-54   | 1.0                 | 60          | 0.135 | 0.112 | 1.21        |
| 7<br>m <sub>1</sub> =1000 Driver <sub>1</sub> =male aged +55<br>m <sub>2</sub> =1000 Driver <sub>2</sub> =female aged +55       | 1.0                 | 60          | 0.183 | 0.151 | 1.21        |

A negative coefficient for female driver in Model B4 shows a lower injury probability for female drivers than male drivers. The results show that, for example, in a frontal collision between two cars with the same mass (1000 kg) but different driver genders, the probability of injury for the male driver is about 13.5% while the probability of injury for the female driver is about 11.2%. This effect is not in accordance with the general expectation that female drivers are generally more vulnerable than male drivers when involved in similar crashes due to a relatively less physical strength. One possible explanation might be given by the type of cars female drivers tend to drive compared to male drivers. For example, they might tend to drive model variants that are newer or have better secondary safety features. Examination of the available two-car crash data (crashes between 2000-2006) shows that, for example, about 60% of the vehicles that were driven by male drivers were registered for the first time before 2000 while, for female drivers, this figure is about 50%. This suggests there is a general tendency for female drivers to drive vehicles that are newer compared to male drivers. It should be noted that in the analysis of two-car crashes in Great Britain during 2000-2004, DfT (2006) found the consistent results that female drivers are less likely to be killed than men drivers when involved in the crashes.

On the other hand, the estimated effects for driver age show that a younger driver has a lower risk of injury than an older driver of a similar vehicle when involved in crashes ( $P_2=0.119$  in crash 2 compared to  $P_2=0.183$  in crash 3 in Table 12); this is in accordance with the prevailing wisdom.

#### **4.5 Other types of collisions**

The same methodology has been used to analyse driver injury risk in front to side and front to back collisions (Tolouei, 2011). Similar to frontal collisions, it was found that the distributional form that best describes the closing speed in both front to side and front to back crashes is a log-normal distribution in which mean is related to the speed limit of the road. The results also confirmed that, identical to frontal collisions, the probability of injury of the driver of a vehicle 1 increases with speed limit and with increasing mass ratio ( $\mu = m_2/m_1$ ) while the probability of injury of the driver of vehicle 2 increases with speed limit and with decreasing mass ratio.

The results for front to side collisions confirmed that similar to the case of frontal collisions, there is a protective effect of vehicle size above and beyond that of vehicle mass; the data did not show any effect of vehicle size in front to back crashes. The findings on the effects of

driver age and gender in both front to side and front to back collisions were in general similar to those in the case of frontal collisions (see Section 4.4 for details).

## 5. Summary and discussion

This paper has introduced a novel methodology to estimate the partial effects of mass and size on absolute driver injury risk in each of the vehicles in a two-car crash. In the introduced methodology, the driver injury probability is described by a logistic function that includes, for each vehicle involved in the crash, the velocity change (defined as a function of mass ratio and closing speed) as well as various driver and vehicle characteristics. Because data on the speed of the vehicles prior to the crash is not available, a distribution for closing speed is assumed, the parameters of which are estimated in model estimation process. The methodology uses the conditional joint injury probabilities in two-car crashes as the basis of analysis to solve the issue related to lack of data on crashes where no driver is injured; these conditional joint probabilities are used to form the likelihood function. The parameters describing the driver injury probability in each vehicle are estimated by maximising the likelihood function over the two-car crash dataset.

This methodology was used to analyse frontal two-car crashes in Great Britain. The results confirmed that in a two-car collision, the probability of injury of the driver of vehicle 1 increases with speed limit and with increasing mass ratio ( $\mu = m_2/m_1$ ) while the probability of injury of the driver of vehicle 2 increases with speed limit and with decreasing mass ratio; that is, in a two-car collision vehicle mass has a protective effect on its own driver injury risk and an aggressive effect on the driver injury risk of the colliding vehicle. Another novel aspect of the analysis based on the introduced methodology was separating the effect of vehicle mass from that of vehicle size on absolute driver injury risks of the vehicles involved in a two-car crash, where vehicle size is represented by “vehicle length  $\times$  vehicle width”. The results confirmed that there is a protective effect of vehicle size above and beyond that of vehicle mass for frontal collisions. It was mentioned briefly that the findings on the effects of vehicle mass, vehicle size, and driver factors were similar for all types of two-car collisions; the only exception was the effects of vehicle size in front to back collisions where the data did not show any effect of vehicle size above and beyond that of mass ratio.

Investigation of the relationship between vehicle mass and injury risk in detail is in particular important as it is the key to understand the relationship between environmental and safety outcomes within the vehicle fleet. Although it has been discussed that vehicle mass imposes a

trade-off between individual vehicles' secondary safety performance and fuel consumption (Tolouei and Titheridge, 2007), the results presented in this paper suggest that mass might not necessarily impose a trade-off between safety and environmental goals in the vehicle fleet as a whole. This is because the secondary safety performance of a vehicle depends on both its own mass and the mass of the other vehicles in the fleet. It was shown that absolute driver injury risk in a frontal collision changes by mass ratio and not the absolute mass of the vehicles. Therefore, in a uniform fleet downsizing scenario, which is a favourable scenario for the fleet fuel economy due to a reduction in vehicles' mass within the fleet, no subsequent increase in the number of casualties is expected provided that the downsizing only focuses on vehicle mass whilst maintaining vehicle size.

On the other hand, a reduction in fleet diversity by decreasing the variance of mass within the fleet could be a favourable scenario regarding safety. It can be shown using the findings presented in this paper that increasing the size of lighter cars within the fleet while maintaining their mass on one hand, and decreasing the mass of heavier cars within the fleet while maintaining their size on the other hand is a desirable scenario in favour of both safety and environmental goals (Tolouei, 2011). The relationship between vehicle size and mass reflected in Figure 2 shows a considerable variation in mass for a given size. This suggests that there is the potential to decrease the mass of many vehicles in the fleet whilst maintaining their dimensions. As Ross and Wenzel (2001) discussed, there are a number of mass-reduction techniques (e.g. use of lightweight materials in design, use of lighter high-efficiency propulsion systems) which could be used by manufacturers to reduce the kerb mass of their new car models. A combination of policies targeting both new vehicle design through various mass-based regulations and the consumer car purchase process through various economic incentives could be an effective way in achieving an informed change in the mass distribution of vehicle fleet. These together with various policy options, and their advantage and disadvantages should be discussed in detail in future studies.

### **Acknowledgement**

The authors would like to thank Rees Jeffreys Road Fund for providing financial support and UK Department for Transport for providing relevant data for this study. The authors also wish to thank Professor Benjamin Heydecker for his valuable suggestions and comments on this study.

## References

- Broughton, J., 1996a. The British index for comparing the accident records of car models. *Accident Analysis and Prevention*, 28 (1), 101-109.
- Broughton, J., 1996b. The theoretical basis for comparing the accident record of car models. *Accident Analysis and Prevention*, 28 (1), 89-99.
- Broughton, J., 1996c. Further aspects of the British index of secondary car safety. *Accident Analysis and Prevention*, 28 (6), 791-798.
- Broughton, J., 2007. Casualty rates by type of car. Transport Research Laboratory, Report No. PPR 203, TRL Limited, Crowthorne, UK.
- Department for Transport (DfT), 2007. Transport statistics for Great Britain: 2007 edition. *Transport Statistics Bulletin*, HMSO, London.
- Department for Transport (DfT), 2006. Cars: make and model: The risk of driver injury in Great Britain: 2000-2004. *Transport Statistics Bulletin*, HMSO, London.
- Ernst, E., Bruhning, E., Glaeser, K.P., Schmid, M., 1991. Safety in small and large passenger cars: the compatibility problem in head-on collisions. In: *Proceedings of the International IRCOBI Conference on the Biomechanics of Impacts*, 1991, Berlin.
- Ernvall, T., Pirtala, P., Hantula, L., 1992. The effect of car make and model on accidents. University of Oulu, Road and Transport Laboratory, Report No. SF-90 100 OULU, Finland.
- Evans, L., Wasielewski, P., 1987. Serious or fatal driver injury rate versus car mass in head-on crashes between cars of similar mass, *Accident Analysis and Prevention*, **19** (2), 119-131.
- Evans, L., Frick, M.C., 1992. Car size or car mass: which has greater influence on fatality risk? *American Journal of Public Health*, 82 (8), 1005-1112.
- Evans, L., Frick, M.C., 1993. Mass ratio and relative driver fatality risk in two-vehicle crashes. *Accident Analysis and Prevention*, 25 (2), 213-224.
- Evans, L., 1994. Driver injury and fatality risk in two-car crashes versus mass ratio inferred using Newtonian mechanics. *Accident Analysis and Prevention*, 26 (5), 609-616.

Evans, L., Frick, M.C., 1994. Car mass and fatality risk: has the relationship changed? *American Journal of Public Health*, 84 (1), 33-36.

Evans, L., 2001. Causal influence of car mass and size on driver fatality risk. *American Journal of Public Health*, 91 (7), 1076-1081.

Evans, L., 2004. *Traffic safety*. Michigan: Bloomfield Hills. Chapter 4.

Grime, G., Hutchinson, T.P., 1982. The influence of vehicle weight on the risk of injury to drivers. In: *Proceedings of Ninth International Technical Conference on Experimental Safety Vehicles*, National Highway Traffic Safety Administration, Washington, DC., 726–741.

Hutchinson, T.P., Anderson, R.G., 2009. Vehicle mass as a determinant of fuel consumption and secondary safety performance: A comment. *Transportation Research Part D*, 15(2), 123-125.

Joksch, H.C., 1993. Velocity change and fatality risk in a crash—a rule of thumb. *Accident Analysis and Prevention*, 2 (5), 103-104.

Moin, P. 2001. *Fundamentals of engineering numerical analysis*. Cambridge University Press, Cambridge.

Ross, M., Wenzel, T., 2001. Losing weight to save lives: A review of the role of automobile weight and size in traffic fatalities. Report No. ACEEE-T013, American Council for an Energy-Efficient Economy.

Tolouei, R., Titheridge, H., 2009. Vehicle mass as a determinant of fuel consumption and secondary safety performance. *Transportation Research Part D*, 14 (6), 385-399.

Tolouei, R., 2011. Interactions between environmental and safety performance in vehicle design. PhD Thesis. University College London.

Van Auken, R.M., Zellner, J.W., 2005. Supplemental results on the independent effects of curb weight, wheelbase, and track of fatality risk in 1985-1998 model year passenger cars and 1985-1997 model year LTVs. Report No. DRI-TR-05-01, Dynamic Research, Inc.

Wenzel, T.P., Ross, M., 2005. The effects of vehicle model and driver behaviour on risk. *Accident Analysis and Prevention*, 37 (3), 479-494.

## Appendix 1

In this appendix, we provide a proof of the results for the speed change in a two-car collision given in section 2.

Suppose that car 1, of mass  $m_1$ , is travelling in the direction of the x axis at a speed  $v_1$ , and car 2, of mass  $m_2$ , is travelling at speed  $v_2$  in a direction that is at an angle  $\theta$  to the x axis (so that if  $\theta = 0$ , the collision is head-on; if it is 180 degs, it is a rear-end shunt; and if it is 90 degs, it is a right-angle collision). After impact, we assume the two cars are effectively one object, of mass  $m_1+m_2$ , travelling at a velocity that has a component  $u_x$  in the x direction and a component  $u_y$  in the y direction.

By conservation of momentum in each of the two directions we have:

$$(m_1+m_2)u_x = m_1v_1 - m_2v_2\cos\theta \quad \text{and} \quad (m_1+m_2)u_y = m_2v_2\sin\theta.$$

The change in velocity for car 1 has components:  $v_1 - \frac{m_1v_1 - m_2v_2\cos\theta}{m_1+m_2}$  in the x direction, and  $\frac{m_2v_2\sin\theta}{m_1+m_2}$  in the y direction. Therefore the square of the magnitude of this change in velocity is:

$$\frac{(m_2v_1 + m_2v_2\cos\theta)^2 + (m_2v_2\sin\theta)^2}{(m_1+m_2)^2} = \left(\frac{m_2}{m_1+m_2}\right)^2 (v_1^2 + v_2^2 + 2v_1v_2\cos\theta) = \left(\frac{m_2}{m_1+m_2}\right)^2 v^2. \quad \text{The}$$

symmetrical result:  $\left(\frac{m_1}{m_1+m_2}\right)^2 v^2$  for car 2 can also be found.

Therefore, when  $\theta = 0$  (head-on),  $v$  reduces to  $v_1 + v_2$ ; when  $\theta = 180$  (shunt),  $v$  reduces to  $v_1 - v_2$ ; and when  $\theta = 90$  (right-angle),  $v$  reduces to  $\sqrt{v_1^2 + v_2^2}$ .