SECONDARY SAFETY PERFORMANCE OF VEHICLES IN FLEET: THE ROLE OF VEHICLE MASS

Reza Tolouei, PhD Candidate, Centre for Transport Studies, University College London, UK, email: reza@transport.ucl.ac.uk

ABSTRACT

There are two distinct safety performance aspects for a vehicle that is involved in a two-car crash: “Protectivity Performance” which is linked to the injury risk to the occupants of that vehicle, and “Aggressivity Performance” which is linked to the injury risk that the vehicle imposes to the occupants of the other vehicle. Current methodologies that estimate protectivity and aggressivity performance of vehicles based on all-injury two-car crash data have the main disadvantage that the estimates for the vehicle involved in two-car crashes are influenced by the injury risk in the colliding vehicle. This is mainly due to the lack of data on non-injury crashes. This paper introduces an alternative methodology to estimate the relative safety performance of makes and models in crashes. The introduced methodology overcomes the disadvantage associated with the previous methodologies that use all-injury crash data by introducing two independent indices. Protectivity Performance Index (PPI) is defined as the proportional difference in the absolute driver injury risk between a given make and model and the mean make and model in fleet, and Aggressivity Performance Index (API) is defined as the proportional difference in the imposed absolute driver injury risk by a given make and model and that by the mean make and model in fleet when involved in similar crashes with similar vehicles. PPI and API were estimated for popular makes and models in 2000-2004 Great Britain vehicle fleet before and after controlling for the effect of mass proportion of the colliding vehicles, which contributes to crash severity. The results showed that when mass proportion is controlled, many makes and models in the dataset that previously showed a significantly different safety performance than the average do not have a significantly different performance. This confirms that vehicle mass is the main contributory factor to the risk of driver injury. Cross-comparison of protectivity and aggressivity performance before controlling for the effect of mass proportion showed a general trade-off between the two; however, controlling for the effect of mass proportion confirmed that this trade-off is imposed by vehicle mass. It was found that even when the effect of mass is controlled, there are a few makes and models that are designed successfully in favour of both protectivity and aggressivity performance or in favour of one aspect without a negative impact on the other aspect.

Keywords: Secondary Safety, Vehicle Protectivity, Vehicle Aggressivity, Vehicle Mass

BACKGROUND

There are two distinct safety performance aspects for a vehicle that is involved in a two-car crash: “Protectivity Performance” which is linked to the injury risk to the occupants of that vehicle, and “Aggressivity Performance” which is linked to the injury risk that the vehicle imposes to the occupants of the other vehicle. In a vehicle fleet, different makes and models have different levels of protectivity performance and aggressivity performance depending on their design.

Amongst various design features, vehicle mass is a key variable contributing to the crash injury risk of vehicle’s occupants by influencing crash severity. In a two-car crash, the injury risk of occupants in the lighter car is higher than that of the heavier car due to the greater velocity change during a collision. For example, in the case of an end-on collision 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 collision (Δv_1) depends on the proportion of the total mass contained by the other vehicle and the closing speed:

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

Other specific design features of makes and models could also influence occupant injury risk. Apart from vehicle design, some human factors affect risk of injury as well. For example, for a given vehicle with a given velocity change in a crash, an older occupant tends to be more severely injured than a younger occupant. Figure 1 outlines different factors contributing to the occupant injury risk of a vehicle involved in a two-car crash. Therefore, when estimating and comparing secondary safety performance of different makes and models, it is important to control for the other effects including the effects of mass of the colliding vehicles from the design-specific effects to avoid misleading results.

Several studies have compared safety performance of different makes and models in vehicle fleet using various methodologies. Wenzel and Ross (2005) estimated a combined risk for each make and model in 1997-2001 US fleet. The combined risk was the sum of the risk to drivers in all kinds of crashes and the risk to drivers of other vehicles in two-car crashes when risk is defined as driver deaths per year per million registered vehicles. While this measure gives an indication of risk per ownership, it does not take into account the effect of vehicle usage. It is highly possible that some makes and models have a significantly different usage than others and hence significantly different exposure to risk. Besides, it is not an appropriate measure to compare secondary safety performance of makes and models as it is influenced by the primary safety performance (risk of crash involvement) of vehicles as well.

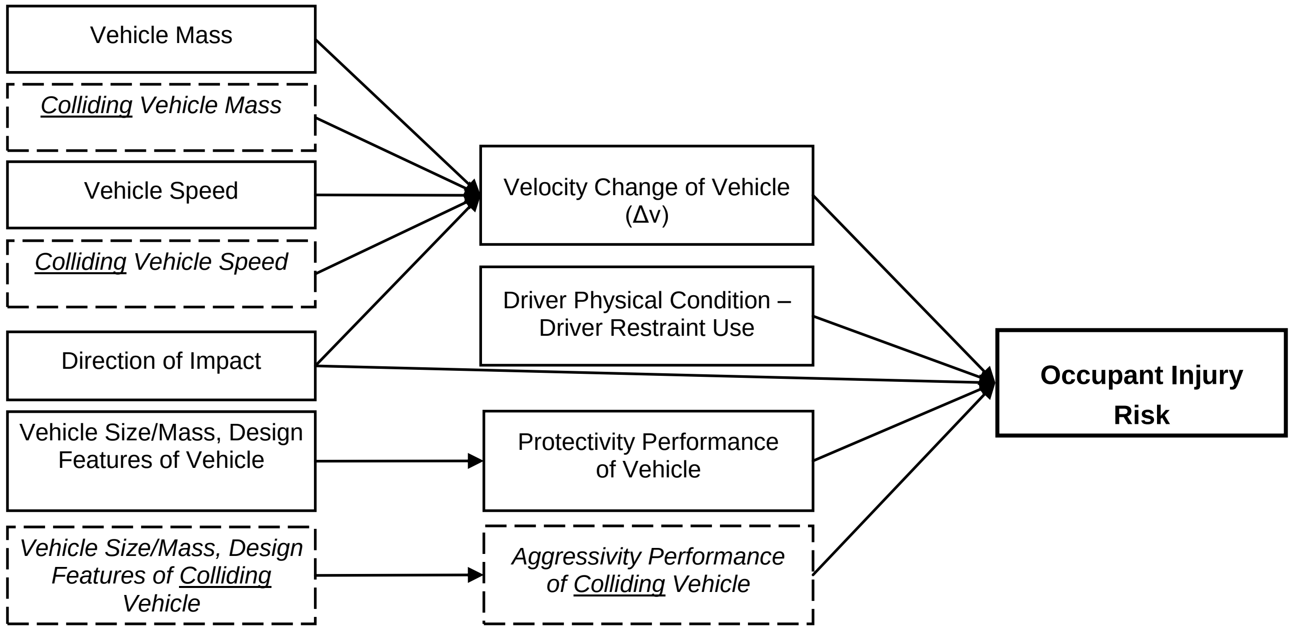

Figure 1 – Occupant injury risk of vehicle in a two-car crash

UK Department for Transport (DfT) estimates protectivity performance of popular makes and models in Great Britain as the driver injury risk and report the results for specific time periods. The latest DfT report on secondary safety of vehicles is available for 186 models of cars involved in traffic crashes during 2000 to 2004 in Great Britain (DfT, 2006). The DfT estimates of protectivity performance are calculated for cars using data from two-car crashes where at least one driver was injured. The DfT safety index for car model m is defined as

D_m = Proportion of drivers of a car model m who are injured when involved in two-car crashes where at least one driver is injured

Adjusted D for all makes and models are calculated using logistic regression models to allow for speed limit (proxy for accident severity), first point of impact, driver sex and driver age.

Broughton (1996a, 1996b) in two papers discussed the DfT method for estimating safety indices as a measure of secondary safety performance of vehicles[1]. In the first paper, he concluded from theoretical considerations that DfT indices provide the most satisfactory means of comparing the secondary safety performance of different models of cars compared to alternative available indices (Broughton, 1996a). Two different safety indices are calculated for each make and model: D_{all} which is based on all kinds of driver injury (killed, seriously injured, and slightly injured), and D_{ksi} which is based on Killed or Seriously Injuried (KSI) drivers. He suggested that it is sensible to concentrate on the “all casualties” index (D_{all}) as it is shown to be highly correlated with “ksi” index (D_{ksi}) and also it is more discriminating because of the much larger number of accidents used in its calculation. Broughton (1996b) discussed practical aspects of the indices in the second paper. He showed that the DfT indices are not biased by ignoring the differences in the distribution of

[1] The term “secondary safety” used by Broughton refers to the protectivity performance of vehicles as used throughout this paper

other cars for different car models involved in accidents. He found that indices calculated from individual years of data are consistent with the indices calculated for the grouped data from 1989-92 and argued that it is justified to accumulate data over several years to provide more reliable results. On the other hand, the fact that the index is a relative measure which compares the safety of different models at the same time limits the number of years over which the index should be calculated. This is because the design of vehicles in fleet changes over time. He also discussed that the indices are closely clustered when calculated for different model variants within makes and models; therefore, it is justifiable to calculate aggregate indices for each make and model to provide more reliable results.

To complement the DfT protectivity index, Broughton (1996a) defined an aggressivity index for car model m as

A_m = Proportion of drivers of cars who are injured when involved in collision with car model m where at least one driver is injured.

As Broughton (1996a) correctly states, the defined D and A indices for makes and models are not independent of injury risk in the other cars in collision with them. This is due to the correlation between the defined risk measures (D and A). For example, when estimating protectivity performance of a given make and model, if the other cars in collisions with this make and model are hypothetically replaced with less physically vulnerable drivers leading to a reduction in driver injury risk in the other car, this would result in an increase in the estimated protectivity performance of this make and model. In an ideal situation, the estimated protectivity and aggressivity performance of makes and models should be independent of risk of injury in the other cars in collision with them. This is the main disadvantage associated with DfT's defined protectivity index (D) as well as Broughton's defined aggressivity index (A) which is because non-injury crash data is not available in Great Britain. The other disadvantage of the DfT methodology is the fact that the estimated effects do not reflect partial effects of vehicle mass and other specific design features of makes and models. A few other studies have separated the effects of mass and other design features on secondary safety performance (Broughton, 1996c; Tolouei and Titheridge, 2009); however, these studies do not investigate the role of relative mass of vehicles on the relationship between protectivity performance and aggressivity performance.

Newstead et al. (2000) have used a preferred method of estimating protectivity and aggressivity performance to compare the secondary safety performance of makes and models in Australian fleet. The indices they have used are the products of two probabilities. The first is the probability of being injured when involved in a crash where a vehicle is towed-away, and the second is the probability of an injured driver being killed or hospitalised. While this results in two independent indices for protectivity and aggressivity, their estimation is solely dependent on the availability of non-injury crash data. Besides, the effects of other vehicles mass proportion (which contributes to crash severity) is not controlled.

The study reported in this paper introduces an alternative methodology and new indices to estimate and compare secondary safety performance (protectivity and aggressivity) of makes and models based on generalized linear modelling techniques and disaggregate analysis of crash data. While this method uses all-injury crash data, the new indices are independent of the injury risk in the other car. Therefore, the introduced methodology overcomes the main disadvantage associated with methods currently used to estimate secondary safety performance of vehicles based on all-injury crash data. The risk of a vehicle being involved in a crash is beyond the scope of this paper. Disaggregate cross-sectional data of a sample of two-car injury crashes in Great Britain were used to investigate the relationship between protectivity and aggressivity performance of popular makes and models before and after controlling for the effect of relative mass of vehicles in crash.

METHODOLOGY

New Secondary Safety Indices: PPI and API

Protectivity Performance Index (PPI) and Aggressivity Performance Index (API) for make and model m in vehicle fleet is defined as

PPI_m = the ratio of driver injury risk of make and model m to that of the average make and model in fleet.

API_m = the ratio of driver injury risk of cars in collision with make and model m to that of cars in collision with the average make and model in fleet.

Assume a hypothetical set of two-car crashes between own cars (an average make and model) and the other cars. Table 1 shows the injury distribution between drivers where $X_1 + X_2 + X_3 + X_4 = N$ (total number of crashes) and $\pi_i = x_i / N$ (probability of driver injury). This scenario of two-car crashes is referred to as the Base Scenario (BS).

Table I – Driver injury distribution in the Base scenario (BS)

		Drivers in own car (average make and model)	
		Not injured	Injured
Driver in other car	Not injured	$x_1(\pi_1)$	$x_2(\pi_2)$
	Injured	$x_3(\pi_3)$	$x_4(\pi_4)$

Absolute crash injury risk (P) to the driver of own cars and the other cars are,

$$P_{own}(BS) = \frac{X_2 + X_4}{N} = \pi_2 + \pi_4 \qquad (2)$$

$$P_{other}(BS) = \frac{X_3 + X_4}{N} = \pi_3 + \pi_4 \qquad (3)$$

Two measures of conditional risks are defined: crash injury risk to the driver of own cars when involved in two-car crashes with the other cars where at least one of the drivers is injured (Q_{own}), and crash injury risk to the driver of other cars when involved in two-car crashes with the own cars where at least one of the drivers in injured (Q_{other}). By definition,

$$Q_{own}\,(BS) = \frac{X_2 + X_4}{X_2 + X_3 + X_4} = \frac{\pi_2 + \pi_4}{\pi_2 + \pi_3 + \pi_4} = D \qquad (4)$$

$$Q_{other}\,(BS) = \frac{X_3 + X_4}{X_2 + X_3 + X_4} = \frac{\pi_3 + \pi_4}{\pi_2 + \pi_3 + \pi_4} = A \qquad (5)$$

Q_{own} is equivalent to DfT's secondary safety index D, and Q_O is equivalent to Broughton's aggressivity index A. As these relationships show, the conditional injury risk in own cars (Q_{own}=D) is not independent of the conditional injury risk in the other cars (Q_{other}=A). This is due to the fact that data on x_1 is not available in the crash data, therefore for each vehicle only conditional risk can be calculated and not the absolute risk which is the preferred measure of risk.

An Alternative Scenario (AS) is introduced where for the same two-car crashes in the base scenario, the make and model of the own cars is replaced with a different make and model (model m) while the other cars are kept unchanged. Therefore, the only difference between the base scenario and the alternative scenario is the make and model of the own cars. Suppose that model m cars have a design feature and hence a different protectivity performance that changes crash injury risk to their own driver α times and a different aggressivity performance that changes crash injury risk to the driver of the other car β times.

$$\frac{P_{own}\,(AS)}{P_{own}\,(BS)} = \hat{PPI}_m \qquad (6)$$

$$\frac{P_{other}\,(AS)}{P_{other}\,(BS)} = \hat{API}_m \qquad (7)$$

Expected Values of PPI and API

Based on the two assumptions that PPI and API are independent from each other and crashes in the base and alternative scenarios have the same severity, the resulting change in the distribution of driver injury based on the definitions and simple algebra is as shown in Table 2.

Table 2 – Driver injury distribution in the alternative scenario (AS)

		Drivers in own car (models m)	
		Not injured	Injured
Driver in other car	Not injured	π_1+(1- PPI) π_2+(1-API) π_3+(1- PPI - API + PPI×API) π_4	PPI [π_2+(1- API) π_4]
	Injured	API [π_3+(1- PPI) π_4]	PPI ×API π_4

By definition,

$$Q_{own}\,(BS) = \frac{\hat{PPI}\left(\pi_2 + \pi_4\right)}{\hat{PPI}\,\pi_2 + \hat{API}\,\pi_3 + (\,\hat{PPI} + \hat{API} - \hat{PPI} \times \hat{API}\,)\pi_4} = D_m \tag{8}$$

$$Q_{other}\,(BS) = \frac{\hat{API}\left(\pi_3 + \pi_4\right)}{\hat{PPI}\,\pi_2 + \hat{API}\,\pi_3 + (\,\hat{PPI} + \hat{API} - \hat{PPI} \times \hat{API}\,)\pi_4} = A_m \tag{9}$$

As it is clear from (8) and (9), a change in the driver injury risk of car model m (PPI_m) not only changes the conditional driver injury risk in own car ($Q_{own} = D_m$), but also changes the relative driver injury risk in other car ($Q_{other} = A_m$). As discussed earlier in this paper, this is the main disadvantage of studies that estimate protectivity and aggressivity performance based on the defined conditional risks Q_{own} and Q_{other}.

The odds ratio (OR) of the conditional risks Q_{own} and Q_{other} in two scenarios, which are the ratios of odds of injury risk if the own car is model m to the odds of injury risk if the own car is an average car model, is defined as,

$$OR_{(Q)}\left(AS,BS\right) = \frac{Q(AS)/\left(1 - Q(AS)\right)}{Q(BS)/\left(1 - Q(BS)\right)} \tag{10}$$

Using equations (4), (5), (8), and (9), the odds ratios are

$$\hat{OR}\left(Q_{own}\right) = \frac{\hat{PPI}\,\pi_3}{\hat{API}\,\pi_3 + \hat{API}(1 - \hat{PPI}\,)\pi_4} \tag{11}$$

$$\hat{OR}\left(Q_{other}\right) = \frac{\hat{API}\,\pi_2}{\hat{PPI}\,\pi_2 + \hat{PPI}(1 - \hat{API}\,)\pi_4} \tag{12}$$

Solving for PPI and API in (11) and (12) results in the following formulae:

$$\hat{PPI} = \frac{1 + \left(\frac{\pi_4}{\pi_3}\right) + \left(\frac{\pi_4}{\pi_2}\right) + \left(\frac{\pi_4}{\pi_3}\right)\left(\frac{\pi_4}{\pi_2}\right) - \left(\frac{1}{\hat{OR}\left(Q_{own}\right) \times \hat{OR}\left(Q_{other}\right)}\right)}{\left(\frac{\pi_4}{\pi_3}\right)\left(\frac{\pi_4}{\pi_2}\right) + \left(\frac{1}{\hat{OR}\left(Q_{own}\right)}\right)\left(\frac{\pi_4}{\pi_2}\right) + \left(\frac{\pi_4}{\pi_3}\right)} \tag{13}$$

$$\hat{API} = \frac{1 + \left(\frac{\pi_4}{\pi_3}\right) + \left(\frac{\pi_4}{\pi_2}\right) + \left(\frac{\pi_4}{\pi_3}\right)\left(\frac{\pi_4}{\pi_2}\right) - \left(\frac{1}{\hat{OR}\left(Q_{own}\right) \times \hat{OR}\left(Q_{other}\right)}\right)}{\left(\frac{\pi_4}{\pi_3}\right)\left(\frac{\pi_4}{\pi_2}\right) + \left(\frac{1}{\hat{OR}\left(Q_{other}\right)}\right)\left(\frac{\pi_4}{\pi_3}\right) + \left(\frac{\pi_4}{\pi_2}\right)} \tag{14}$$

For a sample of two-car crashes with the same severity, if the estimated values of odds ratios of conditional risks (OR(Q_{own}) and OR(Q_{other})) as well as the ratios of $\frac{\pi_4}{\pi_2} = \frac{x_4}{x_2}$ and $\frac{\pi_4}{\pi_3} = \frac{x_4}{x_3}$ are known, it would be possible to estimate PPI and API using equations (13), and (14).

Since Q_{own} and Q_{other} are both measures of risk changing between zero and one, the appropriate functional form reflecting variation in Q is a cumulative logistic function:

$$Q = \frac{exp\left(\beta_0 + \beta_i x_i + \lambda_{j(own)} z_{j(own)} + \lambda_{j(other)} z_{j(other)}\right)}{1 + exp\left(\beta_0 + \beta_i x_i + \lambda_{j(own)} z_{j(own)} + \lambda_{j(other)} z_{j(other)}\right)} \quad (15)$$

where x_i is a set of controlling variables and z_j is a binary variable for make and model j. The appropriate statistical model to estimate Q_{own} and Q_{other} is a logistic regression which is a generalized linear model with binomial distribution for the dependent variable and logit link function.

The appropriate dataset to estimate Q_{own} and Q_{other} should include makes and models involved in two-car crashes in which at least one of the drivers is injured. Each observation includes data on the two makes and models involved in the crash. One of them is referred to as the own car and the other is referred to as the other car. Therefore, the size of the dataset would be twice the number of two-car crashes and symmetric on the makes and models involved in these crashes resulting in each make and model to appear twice in the dataset, once as the “own car” and once as the “other car”.

A logistic regression model is estimated where the dependent variable is conditional crash injury risk to the driver (Q). Other variables are included in the model to control for the effect of other factors that contribute to driver injury outcome (including crash severity). Make and model of both own car and other car are included in the estimated model. Since the dataset is symmetric, the estimated coefficient of the own make and model reflects the effect of that make and model on conditional risk in the own car, Q_{own} and the estimated coefficient of the other make and model reflects the effect of the own make and model on conditional risk in the other car, Q_{other}.

It can be easily shown that for a logistic regression model with binary explanatory variables, the estimates of odds ratio is the exponential of the estimated coefficient of that variable in the model (Hosmer and Lemeshow, 2000). Therefore, for make and model m, if the estimated coefficient of the own make and model and other make and model in the estimated model are respectively $\lambda_{m(own)}$ and $\lambda_{m(other)}$, then odds ratios are

$$\hat{OR}(Q_{own}) = exp\left(\hat{\lambda}_{m(own)}\right) \quad (16)$$

$$\hat{OR}(Q_{other}) = exp\left(\hat{\lambda}_{m(other)}\right) \quad (17)$$

The estimated odds ratio is in fact the ratio of odds of conditional injury risk if the own car is make and model m to the odds of conditional injury risk if the own car is a make and model with an average protectivity and aggressivity performance in the fleet. Using the estimated odds ratios, PPI and API can be estimated using equations (13) and (14). These estimates reflect the protectivity performance and aggressivity performance of a given make and model relative to the mean performance of makes and models in the dataset.

Confidence Intervals of PPI and API

According to the delta method, the variance of f(x) and f(x,y), where x and y are random variables, is estimated using the following formulae:

$$Var\left[f(x)\right]=(f_x)^2 Var(x) \tag{18}$$

$$Var\left[f(x,y)\right]=(f_x)^2 Var(x)+2\left(f_x f_y\right)Cov(x,y)+\left(f_y\right)^2 Var(y). \tag{19}$$

Relationships (11) and (12) can be rewritten as,

$$\hat{OR}(Q_{own})=\frac{\hat{PPI}}{\hat{API}}\left[1+(1-\hat{PPI})\left(\frac{\pi_4}{\pi_3}\right)\right]^{-1} \tag{20}$$

$$\hat{OR}(Q_{other})=\frac{\hat{API}}{\hat{PPI}}\left[1+(1-\hat{API})\left(\frac{\pi_4}{\pi_2}\right)\right]^{-1}. \tag{21}$$

It can be shown that for each binary variable related to makes and models in the logistic regression,

$$Var\left[Log\left(\hat{OR}\right)\right]=Var\left(\hat{\lambda}_m\right)=\left(SE(\hat{\lambda}_m)\right)^2 \tag{22}$$

where $SE(\hat{\lambda}_m)$ is the relevant estimated standard error. Using (18), the variance of odds ratio can be estimated as

$$Var\left(\hat{OR}\right)=(OR)^2\left(SE(\hat{\lambda})\right)^2. \tag{23}$$

Using (19), the variance of odds ratios in (20) and (21) can be calculated.

$$Var\left[\hat{OR}(Q_{own})\right]=\left[\frac{1}{API}\left[1+(1-PPI)(\frac{\pi_4}{\pi_3})\right]^{-1}+\left(\frac{PPI}{API}\right)\left(\frac{\pi_4}{\pi_3}\right)\left[1+(1-PPI)(\frac{\pi_4}{\pi_3})\right]^{-2}\right]^2$$
$$Var(PPI)+\left[\left(\frac{-PPI}{API^2}\right)\left[1+(1-PPI)(\frac{\pi_4}{\pi_3})\right]^{-1}\right]^2 Var(API) \tag{24}$$

and

$$Var\left[\hat{OR}(Q_{other})\right] = \left[\left(\frac{-API}{PPI^2}\right)\left[1+(1-API)(\frac{\pi_4}{\pi_2})\right]^{-1}\right]^2 Var(PPI) + \left[\frac{1}{PPI}\left[1+(1-API)(\frac{\pi_4}{\pi_2})\right]^{-1} + \left(\frac{API}{PPI}\right)\left(\frac{\pi_4}{\pi_2}\right)\left[1+(1-API)(\frac{\pi_4}{\pi_2})\right]^{-2}\right]^2 Var(API) \quad (25)$$

Solving for $Var(PPI)$ and $Var(API)$ in equations (24) and (25), $Var(PPI)$ and $Var(API)$ could be written as a functions of $Var\left[\hat{OR}(Q_{own})\right]$ and $Var\left[\hat{OR}(Q_{other})\right]$ which are known from the model estimation results. The confidence intervals for PPI and API would be

$$CI_{PPI} = \hat{PPI} \pm z_{1-\alpha/2}\sqrt{Var(PPI)} \quad (26)$$

$$CI_{API} = \hat{API} \pm z_{1-\alpha/2}\sqrt{Var(API)} \quad (27)$$

SECONDARY SAFETY ANALYSIS

Study Dataset

The data used in this study is based on the British database of police reports (known as STATS19) which includes road accidents that involve personal injury or death. STATS19 data from 2000 to 2004 was used to extract two-car crashes in which at least one of the drivers was injured. About 300,000 two-car crashes were extracted.

Unfortunately, STATS19 data 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 STATS19 vehicle data to vehicle makes and models. A separate dataset, developed by Department for Transport (DfT) includes make and model data on 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, only about 71% 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 "car magazine"[2], 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.

[2] http://data.carmagazine.co.uk/cars/specs/

This was 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 18% of two-car crashes in the STATS19 data included mass data for both of the cars involved in the crash.

To assure design consistency within makes and models, as suggested by DfT (DfT, 2006), crashes involving makes and models registered before January 1995 were excluded from the dataset. The make and model definitions in the DfT report of secondary safety (DfT, 2006) was used as the basis to define make and model categories in this study. To obtain statistically reliable estimates, the threshold of 150 crash involvement (used by DfT) was used to define make and model categories. As a result of this process, 68 make and model categories were defined followed by one additional category named "other" for the rest of makes and models in the dataset. The final dataset included 37,081 two-car crashes with the first car referred to as the Reference car and the second car referred to as the Other car. A new dataset was developed swapping the first and the second car in each record. This was added to the previous dataset to make the final symmetric dataset of vehicles with 74,162 records (2 x 37,081) with each car repeated twice in the dataset, once as the Reference car and once as the other car.

Statistical Modelling

The developed symmetric dataset provided the opportunity for cross-sectional analysis of makes and models to estimate their design-specific effects on crash injury risk to their own driver and to the driver of the other vehicle. A logistic regression model was estimated where the dependent variable was conditional driver injury risk of the own car (equation (4)). A set of variables defined and included to the model to control for the effects of other factors contributing to driver injury risk. They are summarised in Table 3. Driver age and gender of own car were used as a proxy for driver physical vulnerability while driver age and gender of the other car were used as a proxy for driving style which could influence the speed of impact. Speed limit was included as a proxy for crash severity (DfT, 2006). This variable together with type of crash and driver type variables account for closing speed of vehicles.

Table 3 – Controlling variables

Factor		**Variable**	**Description**
Driver physical condition		Age_{own}	Own driver age *18-24, 25-34, 35-54, +55*
		$Gender_{own}$	Own driver gender *male, female*
Velocity change (Δv = mass proportion x closing speed)	closing speed	Impact	Type of impact *frontal, front to back, back to front, front to side, side to front*
		SPL	Speed limit *20 or 30, 40 or 50, 60, 70*
		Age_{other}	Other driver age *18-24, 25-34, 35-54, +55*
		$Gender_{other}$	Other driver gender *male, female*
	mass proportion	MP	$m_{other}/(m_{own}+m_{other})$

Makes and models are coded using "deviation from the mean" coding method. According to this method, the number of variables would be one less than the number of categories (makes and models). Each category assigns the value of 1 to the variable associated with that category and 0 to the rest except the last category which assigns the value of -1 to all variables. In this method, the reference category represents the mean of the means of the defined categories. Therefore, the estimated coefficient of each make and model represents the difference between the effect of that make and model and the mean of the mean effects of all defined makes and models on dependent variable.

A total of 68 variables were introduced to the logistic model for the 69 defined make and model categories (where the last category was named "other" to include the other makes and models in the dataset) for each of the own car and the other car. This resulted in overall 136 (68 x 2) variables to account for the make and model of the own and the other cars. The estimated coefficient for each defined make and model reflects the difference between the effect of that make and model on the dependent variable (Q_{own}) and the mean of mean effect of all defined makes and models.

As discussed earlier, vehicle mass is a key design variable influencing crash injury risk to the driver of both vehicles in a two-car crash. According to the equation (1), vehicle mass contributes to velocity change which is the best available measure of crash severity. According to this equation, the proportion of vehicle mass contained by the other vehicle in collision (m_{other}/(m_{own}+ m_{other}) is calculated and used in the model as the explanatory variable for relative mass. To investigate the effect of mass of vehicles in collision on vehicle's secondary safety performance, two statistical models were estimated. Model 1 excludes mass proportion variable (MP) and model 2 includes MP as an explanatory variable. The estimation results of logistic regression models are shown in Table 4. The estimated coefficients of makes and models are not shown due to shortage of space in the table.

As the results show, including mass proportion to the logistic model significantly improves performance of the estimated model (measured by log likelihood values through the AIC). This confirms the significant effect of vehicle mass on driver injury risk in own and other car. Comparison of the estimated coefficients of controlling variables between model 1 and 2 shows that including mass proportion variable does not significantly influence the estimated effects of these variables. However, it does influence the estimated coefficients of makes and models. Before including mass proportion (model 1), from 68 estimated make and model effects, 37 (for the own car) and 39 (for the other car) makes and models had significantly different effect from the mean. When mass proportion is included to the model (model 2), only 14 (for the own car) and 8 (for the other car) makes and models have significantly different effect from the mean. This shows that vehicle mass has an important role in driver injury risk in both vehicles. It should be noted that estimated coefficients from the model can not be used directly to investigate the effects on injury risk because the dependent variable is the conditional injury risk in the own car (equation 4) which is influenced by changes in absolute risk in the other vehicle as well.

Table 4 – Model estimation results (dependent variable: conditional driver injury risk in own vehicle, Q_{own})

Variable	**Model 1**			**Model 2**		
	Coefficient	**Std. Error**	**z-value**	**Coefficient**	**Std. Error**	**z-value**
	0.679	0.033	20.759	-2.377	0.136	-17.434
SPL (mile/hr) – reference: 20 or 30 mile/hr						
40 or 50	0.178	0.027	6.570	0.179	0.027	6.573
60	0.381	0.023	16.329	0.385	0.023	16.432
70	0.198	0.032	6.125	0.199	0.032	6.131
Impact – reference: Own: Front - Other: Front						
Own: Front - Other: Back	-1.690	0.028	-60.369	-1.705	0.028	-60.581
Own: Back - Other: Front	0.743	0.031	24.157	0.749	0.031	24.266
Own: Front - Other: Side	-0.410	0.028	-14.376	-0.414	0.029	-14.456
Own: Side - Other: Front	-0.009	0.029	-0.292	-0.006	0.030	-0.220
Other	-0.471	0.033	-14.082	-0.475	0.0336	-14.145
Age_{own} – reference: 35-54						
17-24	0.044	0.026	1.655	-0.012	0.027	-0.454
25-34	0.033	0.023	1.430	0.014	0.023	0.595
55+	-0.008	0.027	-0.281	-0.028	0.027	-1.032
Unknown	-2.363	0.061	-38.559	-2.381	0.061	-38.696
Age_{other} – reference: 35-54						
17-24	0.023	0.026	0.891	0.075	0.026	2.855
25-34	-0.045	0.023	-1.929	-0.028	0.023	-1.224
55+	0.045	0.027	1.647	0.064	0.027	2.322
Unknown	2.054	0.075	27.479	2.064	0.075	27.487
$Gender_{own}$ – reference: Male						
Female	0.809	0.019	42.013	0.782	0.019	40.396
Unknown	-3.907	0.407	-9.604	-3.923	0.407	-9.6380
$Gender_{other}$ – reference:						
Female	-0.643	0.019	-34.458	-0.615	0.019	-32.777
Unknown	4.146	0.562	7.376	4.165	0.564	7.386
MP						
Mass proportion	-	-	-	6.120	0.265	23.067
Make and Model						
Own: 68 variables (not shown)						
Other: 68 variables (not shown)						
Models Statistics						
Observations	74162			74162		
Model D.O.F	157			158		
Log L value	-37,507			-37,234		
AIC	75,328			74,784		

Estimating PPI and API

The driver injury distribution between the own and the other car in the dataset is shown in Table 5. This was used to calculate the ratios of $\frac{\pi_4}{\pi_2} = \frac{x_4}{x_2}$ and $\frac{\pi_4}{\pi_3} = \frac{x_4}{x_3}$ in equations (13) and (14).

Table 5 – Driver injury distribution (2000-2004)

		Drivers in own cars	
		Not injured	Injured
Driver in other cars	Not injured	-	28,814
	Injured	28,814	16,534

Estimated coefficients of makes and models in statistical models 1 and 2 were used to estimate odds ratios of conditional risks ($OR(Q_{own})$ and $OR(Q_{other})$). The estimated odds ratios were then used in equations (13) and (14) to estimate PPI and API for each make and model. Table 6 shows estimated values and confidence intervals of PPI and API for defined makes and models in the dataset before and after controlling for the effect of vehicles mass. Only values significantly different from the mean are shown.

For each make and model, the values of PPI and API represent the proportional change in absolute injury risk to own driver and other driver, respectively, from the mean risk in fleet. Values greater than one show a significantly worse performance than the average make and model (increased absolute risk) and values less than one show a significantly better performance than the average make and model (decreased absolute risk). The "-" symbol shows that the performance of that make and model is not significantly different from the mean.

As noted earlier, comparison of PPI and API before and after allowing for mass proportion shows that in several cases, the estimated change in risk is in fact the effect of vehicles mass, however, there are still a few makes and models having significantly different performance even after the effect of mass proportion is controlled. For example, the results of model 1 for Land Rover Discovery which is one of the heaviest models in fleet shows that it decreases risk of injury to its own driver 48% and increases risk of injury to the driver of other vehicle 18%. However, when the effect of mass is controlled (model 2), the results suggest that Land Rover Discovery protects its driver only 17% better than the average car in fleet and it is not significantly more aggressive than the average car in fleet. On the other hand, results of model 1 for Daewoo Matiz, which is one of the lightest and smallest cars in fleet shows that it increases the risk to its driver by 25% and decreases the risk to the driver of the other car by 23%. However, when the effect of mass is controlled, the results suggest that Daewoo Matiz does not have a protectivity or aggressivity performance significantly different from the mean.

Figures 2 and 3 show cross-comparison of protectivity and aggressivity performance for the 68 makes and models before and after controlling for mass proportion of vehicles in crash, respectively. The percent change in risk is calculated and used in the graphs based on the estimates of PPI and API. Where the estimated PPI or API is not statistically significant for a make and model, it is taken as one for that make and model. Negative values show a more desirable performance than the average (decrease in risk) and positive values show a less desirable performance than the average (increase in risk). Therefore, for example, points in the top left quadrant of the charts represent makes and models with a better protectivity performance and a worse aggressivity performance than the average and those in the bottom right quadrant represents makes and models with a worse protectivity performance and a better aggressivity performance than the average.

Table 6 – Protectivity Performance Index (PPI) and Aggressivity Performance Index (API) of makes and models

Make and Model	**Before allowing for mass proportion (model 1)**				**After allowing for mass proportion (model 2)**			
	PPI		**API**		**PPI**		**API**	
	Mean	**Confidence Interval**	**Mean**	**Confidence Interval**	**Mean**	**Confidence Interval**	**Mean**	**Confidence Interval**
Alfa Romeo 156	1.23	1.46 – 1.00	1.36	1.6 - 1.12	1.30	1.53 - 1.08	1.29	1.54 - 1.05
Audi A3	-	-	-	-	-	-	-	-
Audi A4	0.72	0.79 - 0.64	-	-	0.80	0.87 - 0.72	-	-
BMW 300B	-	-	-	-	-	-	0.83	0.94 - 0.73
BMW 300C	0.91	0.99 - 0.83	1.19	1.29 - 1.10	-	-	-	-
BMW 500B	0.74	0.86 - 0.62	1.21	1.38 - 1.05	-	-	-	-
Citroen Xantia	0.67	0.75 - 0.59	-	-	0.73	0.82 - 0.65	0.86	0.97 - 0.74
Citroen Xsara (00-03)	-	-	-	-	-	-	-	-
Daewoo Matiz	1.25	1.44 - 1.05	0.77	0.90 - 0.64	-	-	-	-
Fiat Brava	-	-	0.86	0.97 - 0.76	0.88	0.98 - 0.77	-	-
Fiat Cinquecento	1.28	1.46 - 1.09	0.80	0.94 - 0.66	-	-	-	-
Fiat Punto (94-98)	1.11	1.19 - 1.02	0.82	0.89 - 0.75	-	-	-	-
Fiat Punto (99-03)	1.20	1.28 - 1.13	0.89	0.96 - 0.83	-	-	-	-
Fiat Seicento	1.39	1.59 - 1.18	-	-	-	-	1.16	1.31 - 1.01
Ford Escort (90-99)	1.07	1.11 - 1.02	-	-	-	-	-	-
Ford Fiesta	1.16	1.2 - 1.12	0.95	0.99 - 0.91	1.06	1.10 - 1.01	1.06	1.10 - 1.02
Ford Focus	0.95	1.00 - 0.90	-	-	-	-	-	-
Ford Ka	1.23	1.31 - 1.16	0.90	0.96 - 0.84	1.08	1.16 – 1.00	-	-
Ford Mondeo	-	-	1.09	1.14 - 1.04	1.06	1.11 - 1.02	-	-
Ford Puma	-	-	-	-	-	-	1.19	1.36 - 1.02
Honda Accord (93-98)	-	-	0.82	0.98 - 0.67	-	-	0.75	0.90- 0.59
Honda Accord (98-03)	0.83	0.95 - 0.71	1.16	1.32 – 1.00	-	-	-	-
Honda CIVIC (93-03)	-	-	-	-	-	-	-	-
Land Rover Discovery	0.52	0.59 - 0.45	1.18	1.33 - 1.03	0.83	0.92 - 0.74	-	-
Land Rover Freelander	0.62	0.72 - 0.53	-	-	0.74	0.84 - 0.64	0.84	0.97 - 0.7
Mazda MX-5	1.23	1.41 - 1.04	-	-	-	-	1.19	1.35 - 1.02
Mercedes A class	-	-	-	-	-	-	-	-
Mercedes C Class	0.85	0.94 - 0.75	-	-	-	-	-	-
Mercedes E Class	0.48	0.55 - 0.40	0.82	0.96 - 0.69	0.62	0.70 - 0.53	0.69	0.82 - 0.56
MG MGF	-	-	0.84	0.99 - 0.69	-	-	-	-
Mini Mini	-	-	-	-	-	-	-	-
Nissan Almera	-	-	0.88	0.97 - 0.80	-	-	0.90	0.99 - 0.81
Nissan Micra (93-03)	1.33	1.4 - 1.26	0.88	0.94 - 0.82	1.17	1.25 - 1.09	1.07	1.13 - 1.01
Nissan Primera	0.83	0.92 - 0.74	0.84	0.94 - 0.74	0.89	0.98 - 0.80	0.78	0.87 - 0.68
Peugeot 106/Saxo	1.25	1.30 - 1.20	0.79	0.83 - 0.75	1.09	1.15 - 1.03	-	-
Peugeot 206	1.15	1.21 - 1.08	-	-	-	-	-	-
Peugeot 306	1.12	1.18 - 1.07	-	-	1.11	1.17 - 1.05	-	-
Peugeot 307	-	-	1.15	1.3 - 1.01	-	-	-	-
Peugeot 406	0.91	0.97 - 0.85	1.14	1.21 - 1.06	-	-	-	-
Renault Clio A (91-98)	1.40	1.55 - 1.25	-	-	1.27	1.42 - 1.11	-	-
Renault Clio B (98-04)	-	-	0.88	0.94 - 0.83	-	-	-	-
Renault Laguna	0.88	0.94 - 0.82	-	-	0.93	1.00 - 0.87	-	-
Renault Megane	-	-	-	-	-	-	-	-
Rover 200/400	-	-	0.92	0.99 - 0.86	-	-	-	-
Rover 25/45	-	-	-	-	-	-	-	-
Rover 75	0.66	0.78 - 0.54	1.24	1.43 - 1.04	0.81	0.94 - 0.68	-	-
Seat Ibiz/Co	-	-	-	-	-	-	1.22	1.37 - 1.06
Skoda Fabia	-	-	-	-	-	-	-	-
Skoda Octavia	1.17	1.32 - 1.03	-	-	1.28	1.43 - 1.14	-	-
Subaru Impreza	-	-	1.23	1.45 - 1.01	-	-	-	-
Suzuki Swift	-	-	0.74	0.87 - 0.60	-	-	-	-
Toyota Avensis	-	-	-	-	1.12	1.22 - 1.03	-	-
Toyota Celica	-	-	-	-	-	-	-	-
Toyota Corolla	-	-	-	-	-	-	-	-
Toyota Starlet	-	-	0.74	0.89 - 0.58	0.79	0.99 - 0.58	-	-
Vauxhall Astra (91-98)	-	-	-	-	-	-	-	-
Vauxhall Astra (98-04)	1.12	1.17 - 1.07	1.09	1.14 - 1.04	1.15	1.19 - 1.10	1.06	1.11 - 1.01
Vauxhall Cavalier	-	-	-	-	-	-	-	-
Vauxhall Corsa (00-03)	-	-	0.83	0.88 - 0.77	0.91	0.97 - 0.85	0.92	0.98 - 0.87
Vauxhall Corsa (93-00)	1.21	1.28 - 1.13	-	-	1.09	1.17 - 1.02	1.08	1.14 - 1.01
Vauxhall Omega	0.78	0.89 - 0.67	-	-	-	-	0.84	0.97 - 0.70
Vauxhall Vectra	0.94	0.99 - 0.90	-	-	-	-	0.94	0.99 - 0.89
Volkswagen Golf	-	-	1.11	1.17 - 1.04	-	-	1.08	1.14 - 1.01
Volkswagen Passat	-	-	1.13	1.23 - 1.03	-	-	-	-
Volkswagen Polo (93-98)	-	-	-	-	-	-	-	-
Volkswagen Polo (98-04)	-	-	-	-	-	-	-	-
Volvo V40	-	-	-	-	-	-	-	-
Volvo V70	0.74	0.87 - 0.62	-	-	0.86	0.99 - 0.73	-	-

For each make and model, PPI and API show respectively the ratio of driver injury risk and imposed driver injury risk of that make and model to the average risk in fleet in similar crashes. Therefore, values greater than 1 mean higher risk of injury than average (undesirable) and values less than 1 mean lower risk of injury than average (desirable)

- : Not Significant

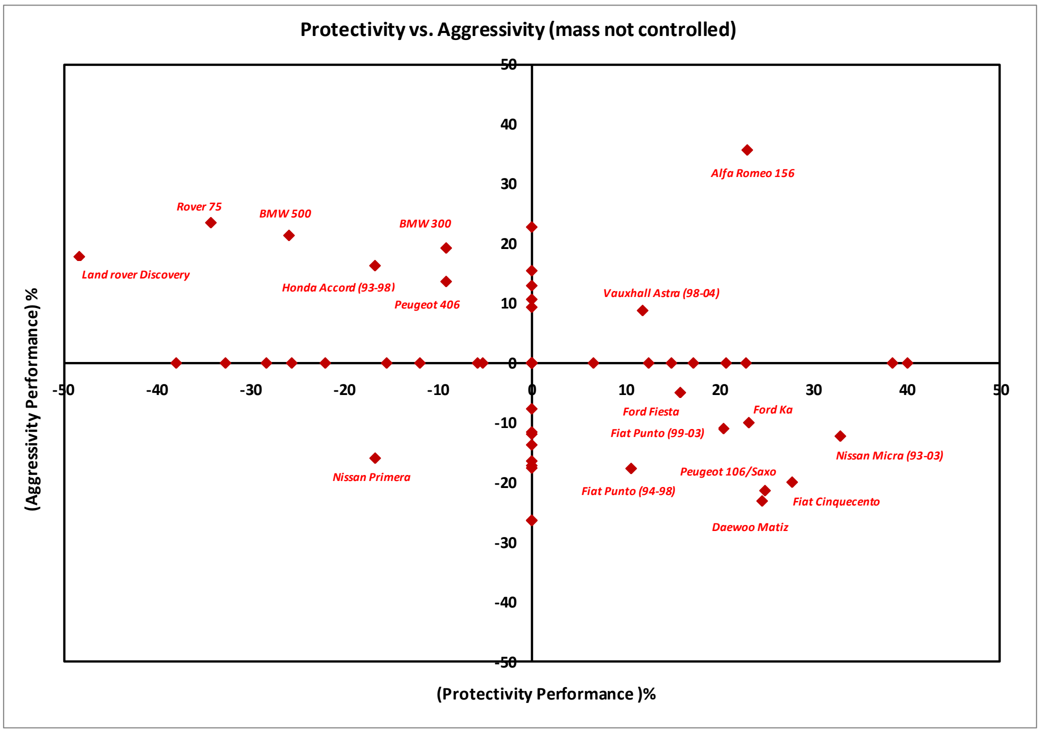

Figure 2 – Cross-comparison of makes and models by PPI and API (before controlling for vehicles mass)

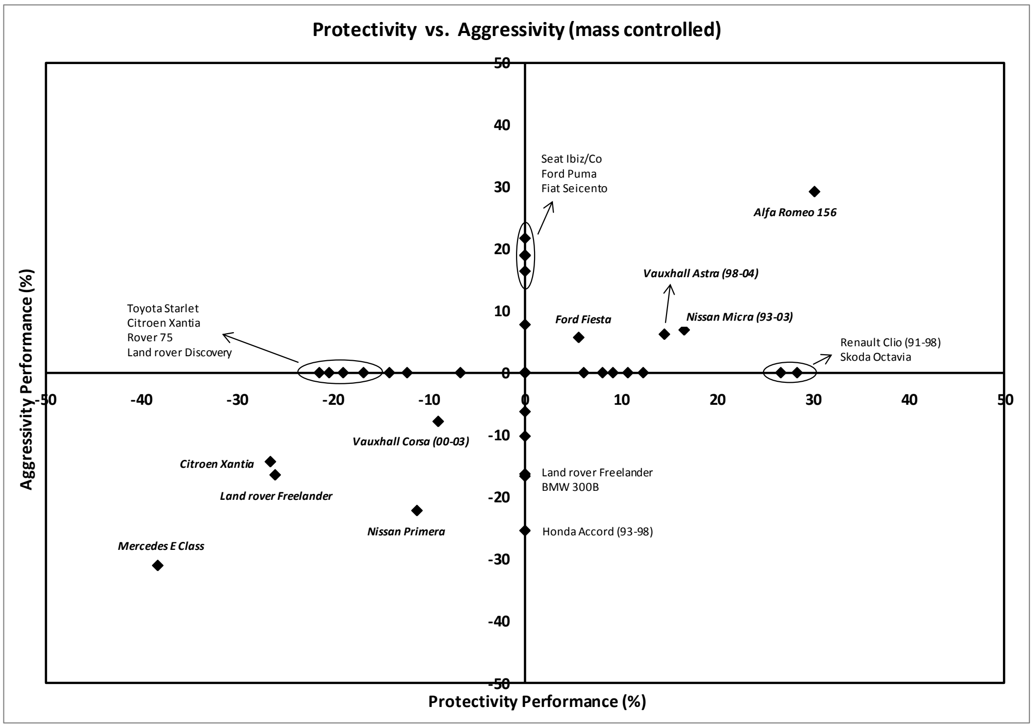

Figure 3 – Cross-comparison of makes and models by PPI and API (after controlling for vehicles mass)

Figures 2 suggests that a general trade-off exists between protectivity performance and aggressivity performance and design advantages for one aspect have resulted in design disadvantages for the other aspect. However, there is no such a trend in Figure 2 when the effect of mass proportion is controlled. This confirms that the trade-off in vehicle design between protectivity performance and aggressivity performance is imposed by vehicles mass.

According to the Figure 2, amongst the studied makes and models, Mercedes E Class, Land rover Freelander, Citroen Xantia, Nissan Primera, and Vauxhall Corsa are the makes and models designed successfully to have a significantly more desirable performance in both protectivity and aggressivity than the average make and model in fleet when the effect of mass is controlled. On the other hand, Alfa 156, Vauxhall Astra, Nissan Micra, and Ford Fiesta are the makes and models with a significantly worse performance in both aspects. It should be noted that these are not necessarily best makes and models in fleet in terms of secondary safety performance as not all the makes and models in fleet were included in the analysis because of lack of enough crash data for them.

SUMMARY AND CONCLUSIONS

A new methodology was introduced to estimate relative secondary safety performance of makes and models in fleet in terms of protectivity performance and aggressivity performance. The introduced methodology overcomes the disadvantage associated with the previous methodologies that use all-injury crash data by introducing two new independent index measures. A Protectivity Performance Index (PPI) was defined as the proportional difference in the absolute driver injury risk between a given make and model and the mean make and model in fleet when involved in similar crashes with similar vehicles, and an Aggressivity Performance Index (API) was defined as the proportional difference in the imposed absolute driver injury risk by a given make and model and that by the mean make and model in fleet when involved in similar crashes with similar vehicles. It was shown that unlike the previous methodologies, the introduced indices are independent of driver injury risk in the other vehicles in collision.

PPI and API were estimated for popular makes and models in 2000-2004 Great Britain vehicle fleet before and after controlling for the effect of mass proportion of the colliding vehicles, which contributes to crash severity. The results showed that when mass proportion is controlled, many makes and models in the dataset that previously showed a significantly different safety performance than the average do not have a significantly different performance. This confirms that vehicle mass is the main contributory factor to the risk of driver injury. Cross-comparison of protectivity and aggressivity performance before controlling for the effect of mass proportion showed a general trade-off between the two, however, controlling for the effect of mass proportion confirmed that this trade-off is imposed by vehicle mass. It was found that even when the effect of mass is controlled, there are a few makes and models that are designed successfully in favour of both protectivity and aggressivity performance or in favour of one aspect without a negative impact on the other aspect.

Although lack of sufficient data for some makes and models prevented the study from including a wider range of makes and models into the analysis, the introduced methodology is transferable to any other two-car crash dataset. The most important factor regarding the introduced methodology which makes it potentially a general method for comparing secondary safety performance of different car models in a vehicle fleet is the fact that it estimates differences in the absolute crash injury risk (risk of injury when involved in a crash) without requiring non-injury crash data which is not usually available.

Although it was found that vehicle mass is the main contributing factor to driver injury risk, it is still not clear how much of this effect is due to the effect of velocity change (influenced by proportional mass of colliding vehicles). There is a correlation between vehicle mass and size and it is possible that some of the estimated effects of mass proportion is related to the effects of vehicle size. There is a substantial difference between the safety effects of mass and size. Vehicle mass has both protective and aggressive effects; however, vehicle size tends to have a protective effect only. More research is required to investigate partial effects of mass and size. It is also strongly recommended that given the substantially important effect of vehicle mass on risk of crash injury, an opportunity be provided to include mass data to the casualty data to provide the possibility to include more makes and models to the study and increase the reliability of the estimates.

ACKNOWLEDGEMENT

The author 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 author is also highly thankful to Dr Helena Titheridge and Professor Benjamin Heydecker for providing valuable guidance and suggestions for this study.

REFERENCES

Broughton, J. (1996a). The theoretical basis for comparing the accident record of car models. Accident Analysis and Prevention, 28(1), 89-99.

Broughton, J. (1996b). The British index for comparing the accident records of car models. Accident Analysis and Prevention, 28(1), 101-109.

Broughton, J., (1996c) Further aspects of the British index of secondary car safety. Accident Analysis and Prevention, 28(6), 791-798.

Department of Transport (DfT). (2006). Cars: make and model: The risk of driver injury in Great Britain: 2000-2004. Transport Statistics Bulletin, HMSO, London.

Hosmer, D. W., S., Lemeshow. (2000). Applied logistic regression, John Wiley and sons, second edition.

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

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

Newstead, S.V., M.H., Cameron, C.M. Le. (2000). Vehicle crashworthiness and aggressivity ratings and crashworthiness by year of vehicle manufacture: Victoria and NSW crashes during 1987-98, Queensland crashes during 1991-98. Monash University Accident Research Centre Reports. No. 171, p. 110.