Abstract

Exposure control or case-control methodologies are common techniques for estimating crash risks, however they require either observational data on control cases or exogenous exposure data, such as vehicle-kilometres travelled. This study proposes an alternative methodology for estimating crash risk of road user groups, whilst controlling for exposure under a variety of roadway, traffic and environmental factors by using readily available police-reported crash data. In particular, the proposed method employs a combination of a log-linear model and quasi-induced exposure technique to identify significant interactions among a range of roadway, environmental and traffic conditions to estimate associated crash risks. The proposed methodology is illustrated using a set of police-reported crash data from January 2004 to June 2009 on roadways in Queensland, Australia. Exposure-controlled crash risks of motorcyclists—involved in multi-vehicle crashes at intersections—were estimated under various combinations of variables like posted speed limit, intersection control type, intersection configuration, and lighting condition. Results show that the crash risk of motorcycles at three-legged intersections is high if the posted speed limits along the approaches are greater than 60 km/h. The crash risk at three-legged intersections is also high when they are unsignalized. Dark lighting conditions appear to increase the crash risk of motorcycles at signalized intersections, but the problem of night time conspicuity of motorcyclists at intersections is lessened on approaches with lower speed limits. This study demonstrates that this combined methodology is a promising tool for gaining new insights into the crash risks of road user groups, and is transferrable to other road users.

1. Introduction

Estimating the relative and absolute crash risks of various road user groups has been of significant practical and academic interest among researchers and road safety professionals for many years. Accurate estimation of the
crash risk under various road traffic conditions is critically important for informing injury prevention and mitigation schemes targeted to specific road user groups such as motorcyclists and older drivers. Of prime interest when estimating the crash risk of a road user group is the likelihood of crash involvement. Based on epidemiologic principles, the crash risk or likelihood of crash involvement depends on exposure. The lack of readily available and reliable exposure information—particularly under various combinations of risk—challenges the profession’s ability to accurately estimate crash risks.

Common techniques for estimating risks are descriptive comparisons, regression models, case-control techniques, and exposure-controlled methodology. In descriptive comparisons, crash data are usually compared under various circumstances of roadway, traffic and environmental factors through frequency counting and categorical data analysis. For example, Ponnaluri (2012) examined various crash circumstances and contributing factors of fatal and injury road traffic crashes of Andhra Pradesh and compared them to national crash records of India using descriptive analyses. The crash characteristics of motorcycle and moped riders in Queensland, Australia were compared across various roadway, traffic and environmental strata using descriptive analyses to develop insights into the crash risk of these two road user groups (Blackman & Haworth, 2013). Descriptive analysis techniques are relatively simple to apply and are helpful for obtaining a quick understanding of the factors associated with crash risk; however, this method suffers from serious limitations due to lack of exposure data or control cases for estimating crash risks.

Road traffic crashes are sometimes analysed after grouping the data into case and control categories so that comparisons can be made between them. For instance, Langford and Kopel (2006) examined the crash risk of older drivers in various crash aspects like posted speed limit, location type, intersection type, number of vehicles involved in a crash, restraint use, and drink driving after categorizing data into young, middle and older age groups, where older drivers were cases and young and middle aged drivers served as controls. The lack of exposure data or proper control cases again raises questions about the reliability of the estimated crash risks. Moreover, this type of disaggregate-analysis approach—in contrast to a regression modelling approach—might not control for other exogenous variables while comparing crash risks between case and control categories in a particular circumstance, e.g. night time crashes. Crash risk comparisons using a proper case-control methodology requires a careful selection of cases and controls, and extensive data collection efforts. Hijar et al. (2000) applied a case-control technique to examine the risk factors and crash risks of being involved in highway traffic crashes, where drivers involved in a crash during a trip in a highway were cases and drivers who completed the trip without any crash were controls. The data for this study was collected through interviews of drivers, through observations, and collecting witnesses for those who died in the crash. Similar extensive case-control data collection and comparison of crash risks under different combination of roadway, traffic and environmental factors is evident in motorcycle safety literature (e.g., Hurt et al., 1981).

Regression models in the form of count modelling are often used to identify influential factors and associated crash risks by developing a relationship between crash frequency and explanatory variables such as roadway and traffic conditions using Poisson or Negative Binomial regression models and their slight variants. For example, Oh et al. (2006) applied a negative binomial model to examine risk factors associated with rail road crashes. Haque et al. (2010) applied Bayesian hierarchical Poisson models to investigate risk factors and magnitude of crash risks associated with motorcycle crashes at signalized intersections. These models usually use traffic volumes in a transportation entity (e.g. intersection, road segment) as an exposure variable, and thus provide crash risks under different influential factors after accounting for exposure of crashes. These models, however, require a system-wide data collection for the exposure variable as well as other roadway geometric and traffic variables, since a transportation entity is an observation unit in count models.

Estimating crash risks by normalizing with exogenous exposure measures is one of the most appealing methods of risk assessment. In exposure-controlled methods, crash risks of different road users are usually obtained by dividing the crash frequency by the Vehicle Miles Traveled (VMT) which is one of the most common and reliable measures of exposure. After controlling for the VMT obtained from the National Personal
Transportation Survey (NPTS) in US, Kweon and Kockelmen (2003) compared the injury risk of various driver/vehicle groups categorized by driver age, gender, vehicle type and crash involvement type. Washington et al. (1999) compared the fatal crash rates between south-eastern and other states in US, where VMT by functional class of roads or VMT by region was used as an exposure measure. Langford et al. (2006) examined and compared the crash risk of different driver age groups through the annual crash rates per million VMT, where the denominator was obtained from the annual Dutch travel survey data. Crash risks using these methods are generally reliable and of high quality, however gathering exogenous exposure data poses a significant challenge.

In summary, exposure-controlled or case-control methodologies are reliable techniques for estimating crash risks; however, they require either observational data for control cases or exogenous exposure data such as vehicle-kilometres travelled on road segments and entering and conflicting vehicular flows at intersections. These data often are not readily available and often require extensive system-wide data collection efforts. This study proposes an alternative methodology for estimating crash risk of road user groups whilst controlling for exposure under a variety of roadway, traffic, and environmental factors by using readily available police-reported crash data. The proposed method has been illustrated using a set of motorcycle crash data in Queensland, Australia. In a companion paper (Haque et al., 2012), motorcycle crashes in Singapore have been analysed using the proposed method and various aspects of motorcycle safety are mainly discussed, leaving the focus here on the methodological aspects and robustness of the proposed method. The illustration using the Queensland data in this paper has been focused on the methodological procedures and attempted to keep straightforward so that it can be easily understood by the road safety professionals or researchers and can easily be applied to other road user groups or road facilities.

2. Proposed method

The proposed method consists of three steps: 1) identifying significant risk factors of a road user group using a log-linear model, 2) measuring the relative exposure of that road user on the identified risky circumstances using the quasi-induced exposure technique, and 3) estimating relative crash risks by combining the results.

2.1 Log-linear modelling

A contingency table analysis in the form of a log-linear model is helpful for identifying significant interactions among roadway, traffic and environmental variables associated with crash frequencies of a road user group. The general log-linear model seeks to explain or fit cell frequencies with an additive model incorporating main effects as well as interactions between variables (Agresti, 1990). An explicit advantage of log-linear models is that they do not require identification of dependent and independent variables. A log-linear model provides measures of the magnitude, direction, and statistical significance of main effects as well as interactions among a set of categorical variables.

For example, consider a three-way frequency table with variables R, T, and E representing respectively roadway, traffic, and environment related factors affecting crashes of a road user. Let, $f_{ijk}$ be the observed frequency and $m_{ijk}$ be the expected frequency for cell $ijk$, where $i, j, k$ designates categories for roadway, traffic, and environmental related variables respectively. The saturated log-linear model is as follows

$$\log_e(m_{ijk}) = \lambda_iR + \lambda_jT + \lambda_kE + \lambda_{ij}RT + \lambda_{ik}RE + \lambda_{jk}TE + \lambda_{ijk}^{RTE}$$  \hspace{1cm} (1)

where $\lambda$ is the overall effect; $\lambda_iR$, $\lambda_jT$ and $\lambda_kE$ are respectively main effects for roadway, traffic, and environmental factors; $\lambda_{ij}RT$, $\lambda_{ik}RE$ and $\lambda_{jk}TE$ are respectively two-way interaction effects for roadway and traffic, roadway and environment, and traffic and environment related variables; and $\lambda_{ijk}^{RTE}$ is the three-way interaction effect for roadway, traffic, and environmental factors.
A parsimonious model is obtained from the saturated model in Eq.1 by deleting insignificant variables using the backward elimination technique. However a hierarchical structure of the model is maintained during the elimination process so that interactive effects can be interpreted without additional mathematical complexities (Agresti, 1990). To measure the goodness of fit, both likelihood ratio $G^2$ (also known as deviance) and Pearson $\chi^2$ are used. These two statistics are used, since the same conclusion from both statistics assures the adequacy of sample size for the model of interest (Goodman, 1984). The significance of a parameter is tested by the partial association test, which calculates the difference in deviance or chi-square values with and without the parameter in question and computes the statistical significance of the ‘larger’ versus the ‘reduced’ model.

Odds ratios are computed to interpret the effects of estimated parameters. Odds ratios provide the relative likelihood of occurrence of events for a given category in comparison with other categories. For instance, the conditional odds ratio for an interaction between roadway and traffic related factors at a fixed level $k$ of environmental factor, $E$ is estimated from the corresponding two-factor parameter estimate for roadway and traffic related variables, i.e., $\lambda_{11}^{RT}$ using the following equation:

$$\theta_{1(k)} = \frac{(m_{11k} \times m_{22k})}{(m_{21k} \times m_{12k})} = \exp(\lambda_{11}^{RT})$$

where $\theta_{1(k)}$ is the conditional odds ratio between roadway factor, $R$ and traffic factor, $T$ at a fixed level $k$ of environmental factor, $E$.

2.2 Quasi-induced exposure technique

Exposure refers to the extent to which a particular road user is exposed to the environment resulting in crashes. In this study, the quasi-induced exposure technique (Carr, 1969; DeYoung et al., 1997; Stamatiadis & Deacon, 1997) is used as an indirect measure of exposure. The strength of this method is that it makes use of crash data to estimate the relative exposure of different road user groups. The basic assumption of the quasi-induced exposure technique is that at-fault drivers in multi-vehicle crashes will choose their not-at-fault victims randomly from all vehicles present. Hence, not at fault drivers/riders can be considered a random sample of the total population of drivers. The distribution of not-at-fault drivers/riders of a driver/rider group reflects the exposure of that group under a given set of roadway, traffic, and environmental conditions.

Suppose, $NF_g$ and $NF_{all}$ denote the frequencies of the not-at-fault crash involvement of a road user group, $g$ and the entire population, respectively, under a given set of roadway, traffic and environmental conditions. Hence, the relative exposure of the road user group $g$, $RE_g$ is the ratio of the not-at-fault crash involvement of that road user group to the entire population,

$$RE_g = \frac{NF_g}{NF_{all}}$$

Similar to Eq.2, the conditional odds ratio of relative exposure estimates $\theta'_{1(k)}$ between roadway factor, $R$ and traffic factor, $T$ at a fixed level $k$ of environmental factor, $E$ is calculated as,

$$\theta'_{1(k)} = \frac{[(RE_g)_{11k} \times (RE_g)_{22k}]}{[(RE_g)_{21k} \times (RE_g)_{12k}]}$$

2.3 Relative risk index

Estimates of the log-linear model and the quasi-induced exposure technique are combined to measure the exposure-controlled crash risk of the road user group of interest. In particular, the Relative Risk Index ($RRI$) measure is computed by combining the odds estimates of risk factor and exposure. Mathematically, $RRI$ is the ratio of odds ratios of expected crash frequencies from a log-linear model and exposure estimates from the quasi-induced exposure technique under a set of certain characteristics, such that
\[ RRI = \frac{\theta_{ijk}}{\theta'_{ijk}} \]  

where, \( \theta_{ijk} \) is the odds ratio of expected frequencies from a log-linear model under a given set of roadway, traffic, and environmental factors; and \( \theta'_{ijk} \) is the odds ratio of exposure estimates for the same combination of roadway, traffic and environmental factors. Hence \( RRI \) provides an indication of crash risk of a road user group in various roadway, traffic and environmental circumstances after controlling for exposure. A \( RRI \) value of 1 in an interactive combination among roadway, traffic and environmental variables means the crash risk is equal to exposure, while a value less than 1 stands for a lower risk and greater than 1 implies a higher risk than the exposure on that interactive circumstance.

3. Dataset for illustration

The proposed method is illustrated using a set of traffic crashes on roadways in Queensland, Australia. In particular, the method is applied to motorcycle crashes involving at least one other motor vehicle at intersections. Crash risks of motorcyclists under various combinations of traffic, environmental and roadway factors were estimated. The road traffic crash data used in this study were extracted from the Department of Transport and Main Roads’ (TMR) road crash database (TMR, 2012). The TMR database includes all road traffic crashes reported to the Queensland Police Service (QPS), which result from the movement of at least one road vehicle on a public road or road-related area and involve death or injury to any person, or property damage. Included in this study are road traffic crashes that occurred at intersections from January 2004 to June 2009 and were captured in the TMR database. During this period there was a total of 50,614 crashes that occurred at intersections, of which 8% involved motorcycles. Motorcycle crashes involving at least one other motor vehicle amounted to a total of 2,829 crashes.

The crash dataset contains numerous variables including crash location, crash date, collision type, contributing crash factors, roadway geometrics, traffic features and environmental conditions. The analysis started with all variables related to roadway, traffic and environmental factors, however several variables like road surface condition, atmospheric condition, horizontal curvature and vertical road alignment resulted in sparse cell counts in the contingency table of multi-vehicle motorcycle crashes at intersections when cross-classified with other variables. These insignificant variables were excluded from further analysis. The contingency table for multi-vehicle motorcycle crashes at intersections was formed by cross-classifying variables like posted speed limit, intersection type, lighting condition and traffic control type. The univariate distributions or one-dimension frequencies of multi-vehicle motorcycle crashes across these variables are shown in Table 1. Posted speed limit along the approaches of the intersection had three categories: less than 60 km/h, 60 km/h and more than 60 km/h. The distribution of motorcycle crashes in these categories was respectively about 17%, 68% and 14%. Intersection configuration based on the geometric arrangement of approaches had three categories including three-legged, roundabout and four-legged, with the corresponding distribution of motorcycle crashes respectively about 51%, 15% and 34%. Lighting condition at the time of a crash had two categories: daylight and dark including dawn and dusk, with the latter accounting for about 25% of multi-vehicle motorcycle crashes at intersections. Traffic control type at the intersection was classified as signalized and unsignalized, with the former representing about 24% of motorcycle crashes involving at least one other vehicle at intersections.

Collision records of all vehicles at intersections were extracted from the same dataset to estimate the relative exposure of motorcyclists under various combinations of roadway, traffic and environmental factors using the quasi-induced exposure technique. The randomness assumption of the quasi-induced exposure technique heavily relies on the fault of crash involvement and subjected to errors if there is any bias in assigning the fault in a crash. A bias can originate from the assignment of fault to a driver/vehicle unit if there are certain contributing human factors (e.g., alcohol impairment) to the crash occurrence. Hence, several researchers e.g., Stamatiadis and Deacon (1997), Chin and Haque (2012) have advocated using ‘clean’ crash records to minimize the bias on the
at-fault determination. It is useful to mention at the outset that this study utilizes the fault data contained in the police crash records. Generally, the Police in Queensland assign unit number one to the road user that they consider most at fault for the crash (Watson, 2004). The dataset is filtered for a few conditions so as to reduce or minimize bias of the fault assignment. First, the fault assignment in crashes involving more than two vehicles might be ambiguous, hence only two-vehicle crashes are used for relative exposure estimation. In the dataset of this study, crashes involving only two vehicles represented about 70% of total intersection crashes. Second, the dataset is screened to include ‘clean’ crash records only, whereby a crash is excluded if cited for contributing factors like alcohol, drug or using a mobile phone whilst driving. A citation for driving under the influence of alcohol or using a mobile phone while driving might always be considered as at-fault and hence may not truly reflect hazardous driving conditions. After filtering the data for clean records and missing information on the vehicle types, about 95.4% of two-vehicle crashes at intersections were retained. In the ‘clean’ dataset, there were 33,954 driver/vehicle units involved in two-vehicle crashes at intersections as a not-at-fault party.

### Table 1: Frequencies of multi-vehicle motorcycle crashes at intersections

| Variable                | Categories                     | Frequency | %   |
|-------------------------|--------------------------------|-----------|-----|
| Speed Limit             | Less than 60 km/h              | 488       | 17.25 |
|                         | 60 km/h                        | 1,934     | 68.36 |
|                         | More than 60 km/h              | 407       | 14.39 |
| Intersection Configuration | 3-legged                       | 1,440     | 50.9 |
|                         | Roundabout                     | 413       | 14.6 |
|                         | 4-legged                       | 976       | 34.5 |
| Lighting Condition      | Dark including dawn and dusk   | 694       | 24.53 |
|                         | Daylight                       | 2,135     | 75.47 |
| Traffic Control Type    | Signalized                     | 673       | 23.79 |
|                         | Unsignalized                   | 2,156     | 76.21 |

### 4. Results and discussion

Log-linear model goodness of fit statistics accommodating different levels of interaction among explanatory variables are shown in Table 2. These statistics assist in the retention of an appropriate number of interactions among explanatory variables influencing motorcycle crashes at intersections. The deviance, $G^2$ and Pearson $\chi^2$ values for a log-linear model with two-way interactions are 23.6 ($p$-value = 0.10) and 24.1 ($p$-value = 0.09) respectively, suggesting that removing three- and four-way interactions from the saturated model does not significantly impact model goodness of fit. On the hand, a log-linear model with main effects yields the deviance, $G^2$ and Pearson $\chi^2$ values 604.9 and 511.2 respectively, with corresponding $p$-values less than 0.001, indicating that two-way interactions have a significant impact on the model fit and should be included. A parsimonious model is identified by removing insignificant variables one by one and is shown in Table 3. The model contains all four main effects and five two-way interactions, and yields deviance, $G^2$ and Pearson $\chi^2$ values of 24.3 and 24.9 respectively with 18 degrees of freedom, and $p$-values of 0.14 and 0.13, suggesting a good fit to the data.

Table 3 presents the parameter estimates of the parsimonious log-linear model of multi-vehicle motorcycle crashes at intersections. Estimates of relative exposures of motorcycles and relative risk indices under the same combination of explanatory variables are also reported in Table 3. The ‘baseline’ case of the model assumes a motorcycle crash that happens on an approach to a four-legged unsignalized intersection with a posted speed limit of 60 km/h during the daytime. Five significant two-way interactions include speed limit x intersection configuration, lighting condition x intersection configuration, traffic control type x intersection configuration, speed limit x lighting condition, and speed limit x traffic control type.
configuration, speed limit x lighting condition, speed limit x traffic control type, intersection configuration x traffic control type and traffic control type x lighting condition.

Table 2: Significance of different level of interactions in the log-linear model

| Level of interactions | Degrees of Freedom (df) | Deviance, $G^2$ | $p$-value | Pearson Chi-square | $p$-value |
|-----------------------|-------------------------|----------------|-----------|-------------------|-----------|
| All interactions      | 0                       | 0.000          | -         | 0.000             | -         |
| 3-way interactions    | 4                       | 2.560          | 0.6340    | 2.352             | 0.6713    |
| 2-way interactions    | 16                      | 23.590         | 0.0989    | 24.061            | 0.0882    |
| Main effects only     | 29                      | 604.895        | <0.001    | 581.203           | <0.001    |
| Parsimonious          | 18                      | 24.342         | 0.1441    | 24.936            | 0.1267    |

The two-way interaction between speed limit and intersection configuration is statistically significant ($\chi^2 = 16.3$, $p$-value = 0.003) and thus influences multi-vehicle motorcycle crashes at intersections. Compared to baseline, the relative crash risk of motorcycles at three-legged intersections is about 40% lower on approaches with a speed limit less than 60 km/h but about 14% higher on approaches with a speed limit more than 60 km/h. The odds of expected frequencies suggest that the crash risk on high speed approaches at three-legged intersections is about 28% higher; however, after normalizing by the exposure estimates the corresponding risk is found to be only 14% higher. Earlier research has suggested that motorcycle crashes are more likely to occur at intersections where other vehicles violating the right-of-way of motorcycles (Hurt et al., 1981), and this problem is more common at uncontrolled T intersections (e.g., Clarke et al., 2007; Pai & Saleh, 2008). The result of this study suggests that the likelihood of right-of-way violation motorcycle crashes at three-legged intersections might escalate if its approaches have a speed limit higher than 60 km/h. On the other hand, the at-fault crash involvement of motorcyclists might also be higher along roadways with a high speed limit (Haque et al., 2009).

The interaction between speed limit and lighting condition suggests that the relative risk of intersection motorcycle crashes during dark lighting condition is about 13% lower if the speed limit along the approaches is less than 60 km/h. Reduced conspicuity of motorcycles has been reported as one of the significant problems for right angle crash vulnerability of motorcycles at intersections (Haque & Chin, 2010). This problem appears to be less for intersections with approaches having a lower speed limit than 60 km/h.

There has been a significant ($\chi^2 = 37.9$, $p$-value < 0.001) interaction between speed limit and traffic control type in influencing motorcycle crashes at intersections. Results show that the relative risk of motorcycle crashes at signalized intersections is 65% lower if a crash occurs along an approach with a speed limit less than 60 km/h. Earlier research has documented a variety of issues related to motorcycle cycle safety at signalized intersections, including over-exposure of motorcyclists (Haque et al., 2008), vulnerability to right-turning vehicles due to a less perception of motorcycles by drivers turning right on a permitted green cycle (Haque & Chin, 2010), and crash involvement with vehicles turning left using an uncontrolled left-turn lane at signalized intersections (Haque et al., 2010). As expected, the effects of these hazardous situations are supposed to be less if the speed limits of intersection approaches are less than 60 km/h.

The interaction between traffic control type and lighting condition indicates that the relative crash risk of motorcycle at signalized intersections is almost doubled if a crash happens during dark lighting condition. As discussed in the last paragraph, signalized intersections are associated with a variety of motorcycle safety issues including over-exposure and vulnerability to right and left turning traffic. It is of no surprise that motorcycle hazards at signalized intersections are likely to escalate during dark lighting condition than daylight.

The interaction between intersection configuration and traffic control type has been found to have a signification ($\chi^2 = 499.7$, $p$-value < 0.001) association with multi-vehicle motorcycle crashes at intersections. The parameter estimates indicate that there is a negative association between three-legged configuration and signal control of an intersection. It implies that the relative crash risk of motorcycles at three-legged intersections is...
about 3.5 times higher if they are unsignalized rather than signalized. It appears that right-of-way violation related motorcycle crashes may be exacerbated at three-legged intersections if they are uncontrolled, as reported by earlier studies elsewhere (e.g., Clarke et al., 2007). Similar high risk of motorcycle crashes at uncontrolled three-legged intersections is also observed in the companion paper that has examined multi-vehicle motorcycle crashes in Singapore (Haque et al., 2012).

Table 3: Relative Crash Risks of Multi-Vehicle Motorcycle Crashes at Intersections

| Parameter                          | Estimate | S.E.  | Chi-Square | p-value | Odds Estimates | Relative Risk Index |
|------------------------------------|----------|-------|------------|---------|----------------|---------------------|
|                                    |          |       |            |         | Frequency      | Exposure            |
| Intercept                          | 5.570    | 0.054 | 10531.30   | <.0001  |                |                     |
| Speed Limit (< 60 Km/h)            | -1.000   | 0.103 | 93.59      | <.0001  | 0.37           | 1.23                | 0.30                |
| Speed Limit (> 60 km/h)            | -1.650   | 0.121 | 186.53     | <.0001  | 0.19           | 0.70                | 0.27                |
| Intersection Configuration (3-legged) | 0.853   | 0.062 | 190.33     | <.0001  | 2.35           | 1.48                | 1.59                |
| Intersection Configuration (Roundabout) | -0.181  | 0.079 | 5.26       | 0.022   | 0.83           | 2.81                | 0.30                |
| Lighting Condition (Dark Lighting) | -1.191   | 0.060 | 393.00     | <.0001  | 0.30           | 1.08                | 0.28                |
| Traffic Control Type (Signalized)  | -0.134   | 0.078 | 2.99       | 0.084   | 0.87           | 0.54                | 1.61                |
| Speed Limit x Intersection Configuration (< 60 km/h x 3-legged) | -0.234   | 0.119 | 3.84       | 0.050   | 0.79           | 1.32                | 0.60                |
| Speed Limit x Intersection Configuration (< 60 km/h x Roundabout) | -0.164   | 0.159 | 1.06       | 0.303   | 0.85           | 1.60                | 0.53                |
| Speed Limit x Intersection Configuration (> 60 km/h x 3-legged) | 0.248    | 0.128 | 3.72       | 0.054   | 1.28           | 1.12                | 1.14                |
| Speed Limit x Intersection Configuration (> 60 km/h x Roundabout) | -0.268   | 0.202 | 1.76       | 0.185   | 0.76           | 1.50                | 0.51                |
| Speed Limit x Lighting Condition (< 60 km/h x Dark) | -0.297   | 0.127 | 5.50       | 0.019   | 0.74           | 0.86                | 0.87                |
| Speed Limit x Lighting Condition (> 60 km/h x Dark) | -0.014   | 0.126 | 0.01       | 0.909   | 0.99           | 1.07                | 0.92                |
| Speed Limit x Traffic Control Type (< 60 km/h x Signalized) | -0.876   | 0.152 | 33.20      | <.0001  | 0.42           | 1.19                | 0.35                |
| Speed Limit x Traffic Control Type (> 60 km/h x Signalized) | -0.020   | 0.137 | 0.02       | 0.885   | 0.98           | 1.00                | 0.98                |
| Traffic Control Type x Lighting Condition (Signalized x Dark) | 0.451    | 0.098 | 21.06      | <.0001  | 1.57           | 0.77                | 2.03                |
| Intersection Configuration x Traffic Control Type (3-legged x Signalized) | -1.581   | 0.099 | 257.57     | <.0001  | 0.21           | 0.71                | 0.29                |
| Intersection Configuration x Traffic Control Type (Roundabout x Signalized) | -5.184   | 0.712 | 53.02      | <.0001  | 0.01           | 2.88                | 0.002               |

*Results of Type 3 analysis of variables in the parsimonious model

5. Conclusions

The study has proposed an alternative methodology for estimating crash risk of a road user group after controlling for exposure. It highlights the importance of considering interaction effects to better understand hazardous situations encountered by a particular road user group. The proposed methodology employs a
combination of a log-linear model and quasi-induced exposure to estimate crash risks under various combinations of roadway, traffic and environmental factors. The method is illustrated using a set of motorcycle crash data along roadways in Queensland, Australia and various interactive factors influencing multi-vehicle motorcycle crashes at intersections are identified and associated relative crash risks estimated.

The proposed technique for estimating crash risk has some distinct advantages over traditional techniques. First, it can provide exposure-controlled crash risk of a road user group by utilizing readily available police-reported crash data. Second, the method does not require the acquisition of accurate exogenous exposure data such as vehicle-km of travel for measuring normalized crash risk. Exposure-controlled crash-risks of a road user group like motorcyclists and older drivers can be estimated as a whole as well as in disaggregate circumstances, e.g. crash risk of motorcyclists in signalized intersections during night time. This method also provides the added benefit of providing exposure estimates in disaggregate conditions, while it might be a tremendous task for collecting such exposure measures exogenously. Third, while the traditional methods like descriptive comparisons and case-control comparison might be criticized in the statistical ground for not controlling explanatory variables systematically, the proposed method utilizes the full strength of a log-linear model in identifying significant interactions and associated odds estimates after controlling for all explanatory variables in the same model.

Relative exposures estimated using the quasi-induced exposure technique might be biased if the randomness assumption of not-at-fault victims is violated. To remove any bias on the fault assignment, a dataset is generally cleaned—by removing crashes where drivers’ received a citation like alcohol, drug and mobile phone use whilst driving—before estimating exposure. However, there might be inherent biases on the randomness assumption depending on the crash occurrence process or crash situation of a road user group. For example, a motorcyclist is more likely to be a victim of crashes at intersections than other roadway locations like highways and roadways away from intersections (Haque et al., 2009). The issue of non-randomness in this case might be removed or minimized by analysing intersection and highway crashes separately. In general, the randomness assumption of the quasi-induced exposure technique works better for disaggregated datasets. For instance, Stamatiadis and Deacon (1997) have demonstrated that the correlation between at-fault and not-at-fault victims is insignificant for disaggregated data. Applying a similar disaggregate analysis, Huang and Chin (2009) have confirmed the validity of randomness of not-at-fault victims on red light running related crashes. Since the proposed method mainly requires exposure estimates in disaggregated conditions by roadway, traffic and environmental factors, the exposure estimations here are unlikely to suffer from any bias from the violation of the randomness assumption.

In addition to the illustration of the proposed methodology, this study has identified several issues related to motorcycle safety at intersections. Posted speed limits along the approaches of an intersection have been found to have varied effects on motorcycles crashes involving at least one other motor vehicle. A speed limit greater than 60 km/h along the approaches of three-legged intersections appears to have a greater crash potential. On the other hand, a speed limit lower than 60 km/h along the approaches of intersections tend to reduce the effects of reduced conspicuity of motorcycles at night and vulnerability potentials at signalized intersections. Traffic control type is found to have significant associations with lighting condition and intersection configuration. The crash risk of motorcycles at signalized intersections is particularly high during dark lighting condition than daytime. Three-legged intersections appear to be more hazardous for multi-vehicle motorcycle crashes when they are unsignalized.

The proposed method of estimating exposure-controlled crash risk using traffic police crash data appears to be robust, as demonstrated to deduce insightful findings on motorcycle safety when applied to crash datasets from different jurisdictions like Queensland, Australia and Singapore. One of the interesting topics for the future research might be comparing the results between the proposed technique and a risk estimation method where crash risks are computed using exogenous exposure estimates. Applying this proposed technique to other road user groups like elderly drivers might be a good extension of this research. Another extension might be inclusion of driver/rider demographics and attributes inside the proposed model to investigate possible interactions of these
with roadway, traffic and environmental factors and estimate associated crash risks of a road user group under various circumstances and combinations of driver/rider attributes.

References

Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.

Blackman, R.A., & Haworth, N.L. (2013). Comparison of moped, scooter and motorcycle crash risk and crash severity. Accident Analysis & Prevention, 57, 1-9.

Carr, B.R. (1969). A statistical analysis of rural ontario traffic accidents using induced exposure data. Accident Analysis & Prevention, 1(4), 343-357.

Chin, H.C., & Haque, M.M. (2012). Effectiveness of red light cameras on the right-angle crash involvement of motorcycles. Journal of Advanced Transportation, 46(1), 54-66.

Clarke, D.D., Ward, P., Bartle, C., & Truman, W. (2007). The role of motorcyclist and other driver behaviour in two types of serious accident in the UK. Accident Analysis & Prevention, 39(5), 974-981.

DeYoung, D.J., Peck, R.C., & Helander, C.J. (1997). Estimating the exposure and fatal crash rates of suspended/revoked and unlicensed drivers in California. Accident Analysis & Prevention, 29(1), 17-23.

Goodman, L.A. (1984). *The Analysis of Cross-classified Data Having Ordered Categories*. Cambridge: MA: Harvard University Press.

Haque, M.M., & Chin, H.C. (2010). Right-angle crash vulnerability of motorcycles at signalized intersections: A mixed logit analysis. Transportation Research Record: Journal of the Transportation Research Board, 2194, 82-90.

Haque, M.M., Chin, H.C., & Debnath, A.K. (2012). An investigation on multi-vehicle motorcycle crashes using log-linear models. Safety Science, 50(2), 352-362.

Haque, M.M., Chin, H.C., & Huang, H. (2009). Modeling fault among motorcyclists involved in crashes. Accident Analysis & Prevention, 41(2), 327-335.

Haque, M.M., Chin, H.C., & Huang, H., 2010. Applying Bayesian hierarchical models to examine motorcycle crashes at signalized intersections. Accident Analysis & Prevention, 42(1), 203-212.

Hague, M.M., Chin, H.C., & Huang, H.L. (2008). Examining exposure of motorcycles at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board, 2048, 60-65.

Hijar, M., Carrillo, C., Flores, M., Anaya, R., & Lopez, V. (2000). Risk factors in highway traffic accidents: a case control study. Accident Analysis & Prevention, 32(5), 703-709.

Huang, H.L., & Chin, H.C. (2009). Disaggregate propensity study on red light running crashes using quasi-induced exposure method. Journal of Transportation Engineering (ASCE), 135(3), 104-111.

Hurt, H.H., Ouellet, J.V., & Thom, D.R. (1981). Motorcycle accident cause factors and identification of countermeasures (No. DOT HS-5-01160). Los Angeles, California: Traffic Safety Center, University of Southern California.

Kweon, Y.-J., & Kockelman, K.M. (2003). Overall injury risk to different drivers: combining exposure, frequency, and severity models. Accident Analysis & Prevention, 35(4), 441-450.

Langford, J., & Koppel, S. (2006). Epidemiology of older driver crashes – Identifying older driver risk factors and exposure patterns. Transportation Research Part F: Traffic Psychology and Behaviour, 9(5), 309-321.

Langford, J., Methorst, R., & Hakamies-Blomqvist, L. (2006). Older drivers do not have a high crash risk—A replication of low mileage bias. Accident Analysis & Prevention, 38(3), 574-578.

Oh, J., Washington, S.P., & Nam, D. (2006). Accident prediction model for railway-highway interfaces. Accident Analysis & Prevention, 38(2), 346-356.

Pai, C.-W., & Saleh, W. (2008). Exploring motorcyclist injury severity in approach-turn collisions at T-junctions: Focusing on the effects of driver's failure to yield and junction control measures. Accident Analysis & Prevention, 40(2), 479-486.

Ponnaluri, R.V. (2012). Road traffic crashes and risk groups in India: Analysis, interpretations, and prevention strategies. IATSS Research, 35(2), 104-110.

Stamatiadis, N., & Deacon, J.A. (1997). Quasi-induced exposure: Methodology and insight. Accident Analysis & Prevention, 29(1), 37-52.

TMR. 2012. Data analysis unit road crash glossary: Data Analysis Unit, Department of Transport and Main Roads.

Washington, S., Metarko, J., Fomunung, I., Ross, R., Julian, F., & Moran, E. (1999). An inter-regional comparison: fatal crashes in the southeastern and non-southeastern United States: preliminary findings. Accident Analysis & Prevention, 31(1-2), 135-146.

Watson, B. (2004). The crash risk of disqualified/suspended and other unlicensed drivers. In Oliver, Dr. John, Williams, Dr. Paul, & Clayton, Dr. Andrew (Eds.), *17th International Conference on Alcohol, Drugs and Traffic Safety, 8 - 13 August, 2004*, Scottish Exhibition and Conference Centre, Glasgow, UK.