Accident Analysis and Prevention 59 (2013) 296–300

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Using speeding detections and numbers of fatalities to estimate relative risk of a fatality for motorcyclists and car drivers

Richard Huggins*

Department of Mathematics and Statistics, The University of Melbourne, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 6 November 2012
Received in revised form 11 June 2013
Accepted 14 June 2013

Keywords:
Induced exposure
Motorcycle
Relative risk

ABSTRACT

Precise estimation of the relative risk of motorcyclists being involved in a fatal accident compared to car drivers is difficult. Simple estimates based on the proportions of licenced drivers or riders that are killed in a fatal accident are biased as they do not take into account the exposure to risk. However, exposure is difficult to quantify. Here we adapt the ideas behind the well known induced exposure methods and use available summary data on speeding detections and fatalities for motorcycle riders and car drivers to estimate the relative risk of a fatality for motorcyclists compared to car drivers under mild assumptions. The method is applied to data on motorcycle riders and car drivers in Victoria, Australia in 2010 and a small simulation study is conducted.

1. Introduction

Our initial motivation is the reporting of statistics relating to motorcycling in Victoria, Australia during 2010. According to a Victorian Transport Accident Commission media release on 18 May 2010, *"Independent research shows that riders are 38 times more likely than car occupants to be seriously injured or killed in a crash."* This statement caused much umbrage in the Victorian motorcycle community and motivated the current work. The justification of the above statement was Table 3.6 page 21 of Berry and Harrison (2008) that gave an estimate of the injury rate per 100 million kilometres travelled for cars and motorcycles, yielding a relative risk of 38.5 in comparing the injury rate of motorcycle riders and car drivers in Victoria. However, they note these figures should be used with caution due to the standard error of the estimated numbers of kilometres travelled. In examining injuries to cyclists in New Zealand, Tin Tin et al. (2010) use household travel surveys to assess travel times. As well as cyclists, they also give estimates of accident rates for motorcyclists and the risk ratios based on these estimates are even higher than those of Berry and Harrison (2008). Moreover, in their Table 6, Johnston et al. (2008) give risk multiples comparing deaths per kilometre for Australian motorcycle riders and drivers for 1998–2007 ranging between 25.4 and 34.6. They quote a Bureau of Infrastructure, Transport and Regional Economics Working Paper as the source of their figures on vehicle kilometres travelled. As noted in Kweon and Kockelman (2003), *"one cannot draw reliable conclusions on safety issues without exposure information."* The traditional method of estimating relative risk is to divide the numbers of fatalities say by a measures of exposure. Measures of exposure are a combination of the number of vehicles and the average time they each spend on the road, but as noted in Staminiadis and Deacon (1997) and evident in Berry and Harrison (2008), accurate estimates can be impossible to make. Apart from their variability, drawbacks of measures of exposure based on surveys is that they can be expensive, are rooted at the time at which the survey was conducted and again as noted in Staminiadis and Deacon (1997), differences such as large heavy trucks being over represented on highways and young drivers over represented on suburban streets during weekend evenings are not accurately measured.

Induced exposure (Thorpe, 1964), which is derived from accident involvement data, and quasi induced exposure avoid the direct estimation of exposure. As noted in Staminiadis and Deacon (1997), quasi-induced exposure analysis regards exposure as *"relative exposure of various classes of drivers/vehicles to situations conducive to multiple vehicle accidents"*. Induced exposure methods are well known in accident research (Cuthbert, 1994) and have been extensively studied and applied, e.g. Staminiadis and Deacon (1997), Redondo-Calderon et al. (2001) and many more, and the method has been compared with case–control studies (Lenguerrand et al., 2008). These methods are based on distinguishing between responsible and nonresponsible drivers in multi vehicle accidents.

Given the amount of summary or aggregate data on road safety and speeding offences currently available the development of statistical methods that could monitor relative risk without conducting surveys to estimate exposure would seem worthwhile. There may be two events related to the exposure of a class of road users. Here we consider being involved in a fatal accident and

* Tel.: +61 3 8344 6610; fax: +61 3 8344 4599.
E-mail address: r.huggins@ms.unimelb.edu.au

0001-4575/$ – see front matter
http://dx.doi.org/10.1016/j.aap.2013.06.020

being detected speeding by a camera. This is particularly suited to car/motorcycle comparisons as the two types of vehicle can be easily distinguished in the photographs. We use a type of induced exposure model to give an estimator of the relative risk of being involved in a fatal accident under an assumption on the relationship between exposure to speed detection and exposure to involvement in a fatal accident. This assumption is discussed in more detail in Section 5. The model proposed here is akin to a latent variable model, where the latent variables are the unobserved exposures of the populations to either risk.

The motivating data for the approach adopted here was collected in the state of Victoria, Australia in 2010. In this state there is an extensive network of cameras that are used to detect speeding drivers and riders and we use this information and information on fatalities to estimate relative risk. The Victoria Police submission to a Parliamentary Road Safety Committee Inquiry into Motorcycle Safety http://www.parliament.vic.gov.au/rsc/inquiries/article/1681 reports that in that year 49 motorcyclists, including pillions, were killed and 130 car drivers, excluding passengers, were killed. Moreover, during that year fixed and mobile speed cameras detected 1,298,485 speeding infringements for cars and 17,715 for motorcycles. These were total speeding infringements including those where the driver or rider could not be identified.

In Section 2 we give the model and develop an estimator of the relative risk. In Section 3 we conduct some simulations, in Section 4 we apply the method to the data described above and in Section 5 we give some discussion. Some technical assumptions and results are given in the appendices.

2. Model and inference

The key to the induced exposure approach to estimating relative risk is the assumption that the rate of exposure to two types of risk are related, so that this quantity cancels out when we take ratios. These are the risk of single or multivehicle crashes. In our case the risks are of being involved in a fatal accident or being detected speeding by a speed camera. Let F_M and F_C be the number of motorcycle and car fatalities respectively and N_M and N_C the numbers of motorcycles and cars detected speeding in a given period.

Let E_M represent the population level exposure of motorcyclists to being in a fatal accident while riding, E_M^* the population level of exposure of motorcyclists to being detected speeding and similarly define E_C and E_C^* for car drivers. Let P_M, P_C, Q_M and Q_C be real parameters. These coefficients relate the risk exposure to the risk, with P_M and P_C being related to speeding, Q_M and Q_C to being fatally injured, where the M and C subscripts referring to motorcycles and cars respectively. For a random variables X, $E(X)$ denotes its expectation and $V(X)$ its variance. We suppose that

$$\begin{aligned} E(N_M) &= P_M E(E_M^*), \qquad & E(N_C) &= P_C E(E_C^*), \\ E(F_M) &= Q_M E(E_M), \qquad & E(F_C) &= Q_C E(E_C). \end{aligned} \tag{1}$$

The parameter Q_M relates the exposure of motorcyclists to the mean number of motorcycle fatalities and similarly Q_C relates the exposure of drivers to the mean number of driver fatalities. The quantity of interest is the relative risk

$$R = \frac{Q_M}{Q_C}. \tag{2}$$

This measures the increase in motorcycle fatalities compared to driver fatalities for the same level of exposure. In addition we suppose that

$$\frac{E(E_M)}{P_M E(E_M^*)} = \frac{E(E_C)}{P_C E(E_C^*)}. \tag{3}$$

This assumption is discussed in Section 5 and some further technical assumptions are given in Appendix A.

We estimate R by

$$\widehat{R} = \frac{F_M N_C}{F_C N_M}. \tag{4}$$

We call this the induced relative risk. Replacing each term by its expectation and using (3) yields

$$E(\widehat{R}) \approx \frac{Q_M E(E_M) P_C E(E_C^*)}{Q_C E(E_C) P_M E(E_M^*)} = \frac{Q_M}{Q_C}$$

so that under (3) the induced relative risk is an estimator of the relative risk. We justify this in Appendix B and further show that under further assumptions stated in Appendix A,

$$\widehat{V}\{\log(\widehat{R})\} \approx \frac{1}{F_M} + \frac{1}{N_C} + \frac{1}{F_C} + \frac{1}{N_M}, \tag{5}$$

which may be used to construct confidence intervals for R.

Remark. As noted by a referee, the probability a motorcycle is detected speeding and is recorded may be $\tilde{P}_M = kP_M$ for some k. For example in most developed countries motorcycles do not have front number plates and hence cannot be detected by front facing cameras. If only the number of infringements where fines were issued is reported then $k \neq 1$. We still suppose (3) holds for P_M, but not necessarily $\widetilde{P}_M$. This requires an adjustment to the estimator of R. Now we have $E(\widehat{R}) = R/k$, so that if $k < 1$, the resulting estimator will be positively biased. If we suppose that k is known, the estimator is now $\widetilde{R} = k\widehat{R}$ and the ends of the confidence intervals are now k times those computed using (5). If k is not known, it may be desirable to conduct an auxiliary experiment to estimate it. Typically, if the sample size is large, the variance of this estimator will be negligible compared with that of $\widehat{R}$. There is a similar adjustment if not all cars detected speeding are identified but for simplicity we omit this.

3. Simulations

To assess the validity of the formulae, a small simulation study was conducted. This part of the study is not meant to reflect reality, but merely to examine the performance of the estimators when the parameter values are known. In the reported simulations, to simulate a population of road users we took E_M = 3,000,000, $E_M^* = 10 \times E_M$, $E_C = 65 \times E_M$ and $E_C^* = 10 \times E_C$. These do not represent numbers of drivers but exposure of drivers, for example numbers of drivers times the average time spent driving. We took $P_M = P_C = 0.0006$, $Q_C = 6.12 \times 10^{-7}$ and $Q_M = R \times Q_C$ where initially we set $R = 27$ to match our application. This gave means $E(N_M)$ = 18,000, $E(N_C)$ = 1,170,000, $E(F_M) = 49.6$ and $E(F_C) = 119.4$. We simulated the observations from Poisson distributions with these means. After 10,000 simulated experiments the mean of the estimates was 27.23 suggesting there is little bias. The variance of the estimates of R was 0.03 and the mean of the estimated variance of $\log(\widehat{R})$ was 0.029 indicating the variance formula is appropriate. Moreover, the estimated coverage probability of the 95% confidence intervals was 0.947. Further simulations over a range of values of R from 0.5 to 30 gave similar results: The average bias over the range of values of R was 2.5% and the average coverage of the nominal 95% confidence intervals was 0.954. It was noted in the simulations that for smaller values of R the simulated value of F_M could be zero in which case the standard errors using (5) were not computable and that the confidence bounds for these smaller values were slightly conservative, with the coverage probabilities being around 0.96. This is due to small values of F_M inflating the variance estimates. See Eq. (5).

Following our remark above we also conducted simulations with $k = 0.5$, so that $\tilde{P}_M = 0.0003$ and $P_C = 0.0006$. In this case, without adjusting for k, the mean of the estimated values of R was 54.4