Accident Analysis and Prevention 40 (2008) 835–839

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Brief communication

Modeling motor vehicle crashes for street racers using zero-inflated models

Zhuo Li [a,1], Stacey Knight [b,c], Lawrence J. Cook [b,c], Lisa K. Hyde [b,c], Richard Holubkov [b], Lenora M. Olson [b,c,*]

[a] Department of Family and Preventive Medicine, University of Utah, 375 Chipeta Way, Salt Lake City, UT 84108, USA

[b] Division of Pediatric Critical Care, Department of Pediatrics, University of Utah School of Medicine, 295 Chipeta Way, Salt Lake City, UT 84108, USA

[c] Intermountain Injury Control Research Center, Department of Pediatrics, University of Utah School of Medicine, 295 Chipeta Way, Salt Lake City, UT 84108, USA

Received 5 May 2007; accepted 7 September 2007

Abstract

Motor vehicle crashes are a leading cause of death for young people in the United States. Assessing which drivers are at a high risk of experiencing a crash is important for the implementation of traffic regulations. Illegal street racing has been associated with a high rate of motor vehicle crashes. In this study, we link Utah statewide driver license citations and motor vehicle crash data to evaluate the rate of crashes for drivers with street racing citations relative to other drivers. Using a zero-inflated negative binomial model we found that drivers with no citations are approximately three times more likely to be at zero risk of a crash compared to drivers with street racing citations. Moreover, among drivers at non-negligible risk of crash, cited street racers are more likely to be involved in a crash compared to drivers without citations or those cited for violations other than street racing. However, drivers with increased numbers of non-street-racing citations experience crash risks approaching those of the cited street racers.

Keywords: Motor vehicle crash; Street racing; Zero-inflated model; Citation; Crash risk

1. Introduction

Motor vehicle crashes are among the leading causes of death in the United States, especially for those between the ages of 3–33 years (NHTSA, 2005). A growing area of interest and research in motor vehicle safety is illegal street racing which may be associated with an increased risk of experiencing a motor vehicle crash. Illegal street racing can refer to either an illegally organized event with spectators and marked distances on public roadways, or illegal spontaneous racing with another car on public roadways. While street racing is a contributing factor to less than 1% of fatal crashes in the United States, the morbidity and mortality associated with street racing appears to be an increasing problem in many cities and towns (Knight et al., 2004). According to nation-wide statistics, 49 people are injured for every 1000 who participate in illegal street racing (NHTSA, 2005). However, most studies examining street racing have not assessed a driver's risk of being involved in a motor vehicle crash. We used Utah statewide driver license citations and motor vehicle crash data to compare drivers with street racing citations to drivers with no citations and drivers with non-street racing citations in order to examine the influence of street racing on the rates of motor vehicle crashes. The ability to determine if a driver's behavior is associated with an increased risk of a motor vehicle crash is important for the implementation and modification of traffic rules, regulations, and prevention activities.

2. Methods

This study was conducted using linked data from the Utah statewide driver license citation database and the Utah statewide

* Corresponding author at: University of Utah School of Medicine, Department of Pediatrics, 295 Chipeta Way, Salt Lake City, UT 84108, USA. Tel.: +1 801 585 9160; fax: +1 801 581 8686.

E-mail address: Lenora.Olson@hsc.utah.edu (L.M. Olson).

[1] Present address: Biostatistics Division, Mayo Clinic, 200 First Street SW, Rochester, MN 55902, USA.

0001-4575/$ – see front matter
doi:10.1016/j.aap.2007.09.022

motor vehicle crash data from 1992 to 2002. The University of Utah School Of Medicine Institutional Review Board approved this study.

2.1. Data

The driver license citation data were obtained from the Utah Department of Public Safety, Drivers License Division. This file contains demographic information and driving citations for all Utah licensees during the study period. The motor vehicle crash data were obtained from the Utah Department of Transportation, Division of Traffic and Safety. This file contains information on reported motor vehicle crashes occurring on public roadways with at least one injury or fatality or at least $1000 in property damage.

These two databases were probabilistically linked. Probabilistic linkage uses variables common in two databases to determine whether a pair of records refers to the same person and event based on the cumulative agreement and disagreement of field values. CODES 2000 linkage software (Strategic Matching, Inc., Morrisonville, NY) was used to probabilistically link the two databases. The matching variables were driver name, date of birth, gender, and driver license number. The resulting analysis data set contained 1,603,305 observations, with cases excluded if they had missing data on variables of interest (e.g., age at entry or gender).

Study variables and definitions:

Drivers were divided into three categories: drivers with no citations (non-cited drivers), drivers with citations not related to street racing (non-racing cited drivers), and drivers with street racing citations (racing cited drivers). Citation type was categorized based on the citation information obtained from the Driver License Division database. In addition to the classification of the citation types, the total citation events were also calculated by counting the number of times at least one citation was issued to a driver during the study period (e.g., if two citations were issued on the same date, it is counted as a single citation event). Covariates used in the study were (1) *driver's age at entry*, which is defined as the age of a Utah driver when he/she entered the study (1 January 1992, or date license first issued if later); (2) *driver's sex*; and (3) *licensing area* (*urban vs. rural*), categorized based on the county of residence from the drivers license file (USDA, 2003).

The main outcome of interest was the crash rate which was determined by dividing the total number of crashes during the study period by the number of driving days for each driver. The *driving days* were calculated as the difference in days between the time a driver entered the study (1 January 1992 or date license first issued if later) and the time the study ended (31 December 2002).

2.2. Statistical analysis

We first examined the distributions of the driver characteristics and the covariates using both univariate and bivariate analyses. Then, we assessed four different generalized linear models to fit the data: Poisson, negative binomial (NB), zero-inflated Poisson (ZIP), and zero-inflated negative binomial (ZINB). Traditionally, the Poisson distribution is appropriate for modeling count data, such as the number of motor vehicle crashes. However, since the number of people not experiencing a crash was far greater than people experiencing a crash, the motor vehicle crash data have more zero events than can be accounted for by the Poisson distribution. Zero-inflated models are used when excess zeros result in a bimodal distribution. This bimodal distribution is a mixture of a mass of extra zeros and a mass that has a Poisson or negative binomial distribution (Cheung, 2002). The mass of extra zeros can be thought of as two subgroups of drivers: (1) drivers whom due to some characteristic could not have crashed during the study period and (2) drivers that could have crashed but did not crash or did not have a reportable crash. Several motor vehicle crash studies have used zero-inflated models and negative binomial models (Kumara and Chin, 2003; Lee et al., 2002; Lee and Mannering, 2002; Qin et al., 2004).

The zero-inflated models have two parts: the zero-inflated and the Poisson or negative binomial depending if a ZINB or a ZIP model is used. The zero-inflated part estimates the probability that a group of drivers is in the "true zero state" by estimating how much more likely one group is to have an excess of drivers without crashes compared to another group. The Poisson or negative binomial coefficients estimate the rate of crashes, conditioned on the group of drivers not being in the "true zero state". These coefficients are reported using relative risk. The two parts of the zero-inflated model need to be considered together when assessing the association between the independent and dependent variable in the model.

Model goodness of fit was examined by the log likelihood using the Akaike information criterion (AIC) (Akaike, 1974; Burnham and Anderson, 2002), and the Bayesian information criterion (BIC) (McQuarrie and Tsai, 1998). The likelihood ratio test was used to compare the Poisson model and NB model (Agresti, 2002; Arminger and Clogg, 1995), and the ZIP and ZINB models. For the zero-inflated models, goodness of fit was assessed using the method of Vuong (Vuong, 1998).

The dependent variable in each model is the rate of crashes for a driver, defined as the total number of crashes for the driver divided by his/her driving days (crashes per 1000 days) during the study period. The independent variables modeled were citation type (categorized as none, one non-racing citation event, two non-racing citation events, three non-racing citation events, four or more non-racing citation events, at least one racing citation event), age group (categorized as 15–19 years, 20–24 years, 25 years or older), gender (female, male) and area where license was issued (categorized as rural or urban). All independent variables were chosen a priori. Programming was done using PROC NLMIXED in SAS 9.0. Statistical significance was set at $\alpha = 0.05$, and 95% confidence intervals were used.

3. Results

There were 1,603,305 drivers licensed and 597,894 crashes during the study period. The average number of crashes per 1000 driving days for all drivers was 0.16. Approximately