Accident Analysis and Prevention 42 (2010) 1273–1282

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Investigating the effect of modeling single-vehicle and multi-vehicle crashes separately on confidence intervals of Poisson–gamma models

Srinivas Reddy Geedipally[a,*], Dominique Lord[b,1]

[a] Texas Transportation Institute, Texas A&M University, 3135 TAMU, College Station, TX 77843-3135, United States
[b] Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States

ARTICLE INFO

Article history:
Received 28 April 2009
Received in revised form 4 February 2010
Accepted 8 February 2010

Keywords:
Single-vehicle crashes
Multi-vehicle crashes
Confidence intervals

ABSTRACT

Crash prediction models still constitute one of the primary tools for estimating traffic safety. These statistical models play a vital role in various types of safety studies. With a few exceptions, they have often been employed to estimate the number of crashes per unit of time for an entire highway segment or intersection, without distinguishing the influence different sub-groups have on crash risk. The two most important sub-groups that have been identified in the literature are single- and multi-vehicle crashes. Recently, some researchers have noted that developing two distinct models for these two categories of crashes provides better predicting performance than developing models combining both crash categories together. Thus, there is a need to determine whether a significant difference exists for the computation of confidence intervals when a single model is applied rather than two distinct models for single- and multi-vehicle crashes. Building confidence intervals have many important applications in highway safety.

This paper investigates the effect of modeling single- and multi-vehicle (head-on and rear-end only) crashes separately versus modeling them together on the prediction of confidence intervals of Poisson–gamma models. Confidence intervals were calculated for total (all severities) crash models and fatal and severe injury crash models. The data used for the comparison analysis were collected on Texas multilane undivided highways for the years 1997–2001. This study shows that modeling single- and multi-vehicle crashes separately predicts larger confidence intervals than modeling them together as a single model. This difference is much larger for fatal and injury crash models than for models for all severity levels. Furthermore, it is found that the single- and multi-vehicle crashes are not independent. Thus, a joint (bivariate) model which accounts for correlation between single- and multi-vehicle crashes is developed and it predicts wider confidence intervals than a univariate model for all severities. Finally, the simulation results show that separate models predict values that are closer to the true confidence intervals, and thus this research supports previous studies that recommended modeling single- and multi-vehicle crashes separately for analyzing highway segments.

1. Introduction

Crash prediction models still constitute one of the primary tools for estimating traffic safety. These statistical models play a vital role in various types of safety studies, but they are most often used for estimating the safety performance of various transportation elements or entities (Persaud and Nguyen, 1998; Lord, 2000; Ivan et al., 2000; Lyon et al., 2003; Tarko et al., 2008; Lord and Mannering, in press). In this context, these models have been employed to estimate the number of crashes per unit of time for an entire highway segment or intersection, without distinguishing the influence different sub-groups have on crash risk. A few exceptions have been noted in the literature however. For instance, some researchers have developed distinct predictive models to estimate the safety performance as a function of different categories of vehicles, such as passenger vehicles and truck crashes (Jovanis and Chang, 1986; Miller et al., 1998; Lee and Abdel-Aty, 2005), different time periods (Cercarelli et al., 1992; Mensah and Hauer, 1998; Zador, 1985) or the number of vehicles involved in each crash, i.e., single-vehicle (SV) and multi-vehicle (MV) crashes (Qin et al., 2004; Lord et al., 2005; Griffith, 1999) among others.

Given the differences observed in the characteristics associated with SV and MV crashes, some transportation safety analysts have proposed that distinct crash prediction models should be developed for these two categories of crashes when the objective of the study consists of estimating the safety performance of highway segments (Mensah and Hauer, 1998; Ivan, 2004; Lord et al., 2005; Jonsson et al., 2007; Harwood et al., 2007; Bonneson et al.,

* Corresponding author. Tel.: +1 979 862 1651; fax: +1 979 845 6006.
E-mail addresses: srinivas8@tamu.edu, gsrinivas8@gmail.com (S.R. Geedipally), d-lord@tamu.edu (D. Lord).
[1] Tel.: +1 979 458 3949; fax: +1 979 845 6481.

0001-4575/$ – see front matter. Published by Elsevier Ltd.
doi:10.1016/j.aap.2010.02.004

2007). These researchers noted that developing two distinct models provides better predicting performance than developing models combining both crash categories together. In most cases, the motivation for separating models by the number of vehicles involved in the crash is based on shape of the functional form linking both crash types to the traffic flow variable that has been found to be very different from one another (Ivan, 2004; Lord et al., 2005).

Since these two categories of models have been used to estimate or predict the number of crashes on highway segments (Harwood et al., 2007; Bonneson et al., 2007; Lord et al., 2008), there is a need to determine whether there exist differences in the computation of confidence intervals when a single model is applied rather than two distinct models. Confidence intervals can play a major role in the selection of various highway design alternatives (Lord, 2008) and the identification of hazardous sites (Hauer, 1997) among others. The primary objective of this paper is to investigate if there is an important difference in the prediction of confidence intervals when a unique model is estimated compared to a distinct model for SV and MV crashes. The secondary objective is to examine if there is any difference in the prediction of confidence intervals when a bivariate negative binomial (BNB) model is used rather than a univariate negative binomial (UNB) model for analyzing SV and MV crashes. Confidence intervals were calculated for the Poisson mean, gamma mean and predicted response, respectively (Wood, 2005). Crash data collected on Texas undivided roads from 1997 to 2001 were used for this comparison analysis.

This paper is divided into five sections. The first section provides a discussion on the modeling prospects related to SV and MV crashes. The second section gives a brief description of data used in this study. The third section describes the methodology utilized for the comparison analysis. The fourth section presents the results of this analysis. The last section provides a summary of the work carried out in this research and recommendation for further work.

2. Background

Several researchers have examined the characteristics and the differences associated with SV and MV crashes (direct comparison). For instance, Ostrom and Eriksson (1993) were the first to examine crash characteristics as a function of the number of vehicles involved in a crash. They studied factors influencing crashes involving intoxicated drivers in northern Sweden. These authors reported that the driver's blood alcohol content was more significantly related to SV crash fatalities than those associated with MV crash fatalities.

Shankar et al. (1995) analyzed the safety effects of highway design features, weather and other seasonal variables on different crash types. Using data collected on a 61-km section near Seattle, WA, they analyzed several variables, such as the number of horizontal curves and their spacing, rainfall amount, snowing conditions, and their relationship to different types of SV and MV crashes. The authors concluded that models predicting crashes for different crash types had a greater explanatory power than a single model that combined all crash types together.

Mensah and Hauer (1998) examined the effects of aggregated and disaggregated traffic flow variables on the estimation of predictive models. They used data on two-lane rural highways published in Persaud and Mucsi (1995) for their analysis. Using exploratory data analyses and regression methods, they found that SV and MV crashes have significant different characteristics. They reported that using an aggregated model that combine both SV and MV crashes predicts fewer crashes than combining the output of two separate models for the same two categories of crashes.

Griffith (1999) studied SV and MV crashes caused by alcohol and drug-impaired drivers. This author found that SV run-off-the-road crashes on freeways resulted in a higher number of nonfatal injuries than for MV crashes involving an impaired driver. In contrast, MV crashes on freeways resulted in a higher number of fatal injuries than for SV run-off-the-road crashes.

Ivan et al. (1999) investigated differences in causality factors for SV and MV crashes on two-lane rural highways in Connecticut. They found that contributing factors were different for each category of crashes. For example, SV crashes were negatively associated with an increase in traffic intensity (exposure), shoulder width, sight distance, and level of service (LOS). On the other hand, MV crashes were positively associated with an increase in traffic intensity, shoulder width, truck percentage, and number of traffic signals.

In a subsequent study, Ivan et al. (2000) reported that the time-of-day differently influenced for both categories of crashes. SV crashes occurred mostly during the evening and at night, as expected, whereas MV crashes occur more frequently during daylight and evening peak periods. This was mainly attributed to the higher traffic intensity (or exposure). Driveway density had a mixed effect on SV crashes. Driveways at gas stations and minor road intersections were negatively associated with SV crashes, whereas driveways located adjacent to apartment complexes seemed to be associated with an increase in SV crashes. MV crashes increased for all types of driveways.

Ivan (2004) modeled crashes, also using data collected in Connecticut, according to the manner of collision (i.e., number of vehicles involved and their direction of travel). The study showed that the expected number of SV crashes decreases with the increase in traffic volume, whereas all types of MV crashes increased with the increase in traffic flow during the evening time period. During daytime, SV and MV opposite-direction crashes decreased with an increase in traffic flow, whereas MV same-direction and intersecting-direction crashes increased with an augmentation in traffic flow.

Qin et al. (2004) developed zero-inflated Poisson models for different crash types (see Lord et al., 2005, 2007 for a discussion about the application of such models in highway safety). These authors examined crashes occurring on highway segments in Michigan and concluded that crashes are differently associated with traffic flows for different crash types. They noted, for example, that aggregated crash prediction model ignores significant variation in highway crashes. For SV crashes, the marginal crash rate was found to be high for low traffic flow levels and small for high traffic flow volumes. For all MV crashes (except multi-vehicle intersecting-direction crashes), the marginal crash rate was small at low traffic volumes, but high for large traffic volumes, probably because this type of crash is more likely to occur under short headways.

Lord et al. (2005) evaluated several functional forms for SV and MV crashes, and one that combined both (referred to as *All Model*), as a function of traffic flow, vehicle density and volume over capacity (V/C) ratio on urban freeway segments in Montreal, Canada. Using regression methods, they recommended developing different predictive models for SV and MV crashes rather than developing a common model for both crash categories. The output of the SV model showed that crashes initially increase, peak and then decrease as density or V/C increases. On the other hand, the MV crash model and the model grouping both SV and MV crashes (*All Model*) showed that crashes increased with an increasing vehicle density or V/C ratio, as expected.

Recently, Jonsson et al. (2007) developed distinct models for SV and different types of MV crashes occurring at intersections on rural four-lane highways in California. These authors concluded that the SV crash model loses some of the observed covariate effects through the aggregation of collision types. Different crash type models exhibited dissimilar relationships with traffic flow and other covariates. Furthermore, Jonsson et al. (2007) noted that the distribution by crash severity varied for different crash types.