Safety Science 66 (2014) 47–53

Contents lists available at ScienceDirect

Safety Science

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

A segment level analysis of multi-vehicle motorcycle crashes in Ohio using Bayesian multi-level mixed effects models

Thomas Flask [a,1], William H. Schneider IV [a,*], Dominique Lord [b,2]

[a] Department of Civil Engineering, The University of Akron, 302 Buchtel Common, Akron, OH 44325-3905, United States
[b] Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843-3136, United States

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form 6 December 2013
Accepted 18 December 2013
Available online 27 February 2014

Keywords:
Hierarchical Bayes
Spatial random effects
Uncorrelated random effects
Negative binomial model
Conditional autoregressive distribution
Mixed effects

ABSTRACT

Multi-vehicle motorcycle crashes combine elements of design, behavior, and traffic. One challenge with working with motorcycle data are the inherit difficulties associated with missing data – such as motorcycle-specific: vehicle miles traveled (VMT) and average daily traffic (ADT). To address the challenges of the missing data, a random effects Bayesian negative binomial model is developed for the state of Ohio. In this study, the random effect terms improve the general model by describing the spatial correlation with fixed effects, the neighborhood criteria, and the uncorrelated heterogeneity for all the multi-vehicle motorcycle crashes that occurred on the 32,289 state-maintained roadway segments in Ohio. Some key findings from this study include regional data improves the goodness-of-fit, and further improvement of the models may be gained through a distance-based neighborhood specification of conditional autoregressive (CAR). In addition to the model improvement using the random effect terms, key variables such as smaller lane and shoulder widths, increases in the horizontal degree of curvature and increases in the maximum vertical grade will increase the prediction of a crash.

1. Introduction

Motorcycle crashes are known to compose a disproportionately high amount of the overall vehicle fatalities in the United States. Fatalities on motorcycles represented 4462 deaths of the 30,797 total fatal crashes in the United States in 2009 (FARS, 2012). In that year, when considering the number of fatalities per 100 million VMT (Vehicle Miles Traveled), the motorcycle fatality rate was over 18 times that of passenger car crashes, which was the second highest crash rate for a specific vehicle class (FARS, 2012).

Nationally, 10.1% of the vehicle crash fatalities were motorcycle crashes in 2009 (FARS, 2012). Despite the limited riding season in Ohio due to the weather, the amount of motorcycle involvement in fatalities in the same year was even higher at 16.3%, or 166 fatalities (FARS, 2012). In more recent years, the number of fatal crashes has decreased slightly due to a proactive approach, with 164 fatal crashes in 2010. At the same time, the number of motorcycle crashes (all severities) in Ohio increased from 4165 in 2009 to 4381 in 2010 (ODPS, 2012).

A motorcycle crash, as any other vehicle crash, is a complex event with many influential factors and characteristics. Since the mechanisms that lead to each crash may be dramatically different, it is natural to assume that the factors that are behind the two distinct crash types are different as well (Haque et al., 2012; Geedipally and Lord, 2010; Jonsson et al., 2007; Ivan, 2004; Savolainen and Mannering, 2007; Yau, 2004). Therefore, it is reasonable to separate single and multi-vehicle crashes.

Two common approaches to model motorcycle crashes are discrete outcome and negative binomial models (see Savolainen et al., 2011; Lord and Mannering, 2010, for a review of these models in highway safety research). Some common findings from discrete outcome models show that injury severity is significantly affected by factors such as helmet use, speeding, alcohol use, and operator age (Chang and Yeh, 2006; Savolainen and Mannering, 2007; Haque et al., 2009). While discrete outcome models perform well in estimating the impact of behavioral and crash characteristics on the type of crash that occurs, negative binomial models are more commonly used to estimate or predict the number of crashes based on information such as geometric, demographic, or infrastructural characteristics (Chin and Quddus, 2003; Haque et al., 2010; Harnen et al., 2003; Houston, 2007; Schneider et al., 2010).

More recently, negative binomial models have been improved by introducing random effects terms, which offer the prospect of including data and relationships that may be difficult to apply in a standard model configuration. These more advanced models are often referred to as multi-level models, as they introduce data

* Corresponding author. Tel.: +1 (330) 972 2426.
E-mail address: whs4@uakron.edu (W.H. Schneider IV).
[1] Tel.: +1 (330) 972 2426.
[2] Tel.: +1 (979) 458 3949.

http://dx.doi.org/10.1016/j.ssci.2013.12.006
0925-7535/

on multiple spatial, temporal, or conceptual levels. Some examples include multi-level random effects that estimate the impact of crashes occurring in the same intersection (Kim et al., 2007), corridor (Guo et al., 2010), region (Yannis et al., 2007), or year. In each case, the multi-level model improves both the model fit and the interpretation of the findings, showing which intersections or regions are more prone to crash occurrence than others. Not only can random effects be used with assigned groups of similar crashes (such as county or time) as shown previously, but they may also be used by comparing the crash frequency of nearby segments or regions (Eksler and Lassarre, 2008; Mitra, 2009; Quddus, 2008; Wang et al., 2009). The researcher may choose from a variety of methods to define which segments or regions are considered the neighbor of another, such as contiguity or a fixed distance. In this case, conditional autoregressive (CAR) random effects are shown to reduce the model error by adding the prior knowledge of neighboring regions and segments, leading to better parameter estimates. Additionally, CAR random effects are also frequently paired with an uncorrelated random effects term, which quantifies the model error that is not related to the nearby regions or segments, but rather unknown or unmeasured influences (Aguero-Valverde and Jovanis, 2008; Eksler and Lassarre, 2008; Guo et al., 2010; Mitra, 2009).

Ultimately, random effect terms may be used to reduce model error that is caused by unavailable or unrecorded data such as motorcycle specific vehicle miles traveled or ADT. The models within this paper will utilize the previous demonstration of the random effect terms used in other areas of traffic safety and apply these statistical methodologies directly to motorcycle specific models. This application will improve the key parameters that influence multi-vehicle motorcycle crashes. While the use of random effects in negative binomial models is relatively new in the field of safety, this application is new to motorcycle specific crashes.

2. Materials and methods

Three datasets (ODOT, 2011; ODPS, 2012; US Census Bureau, 2011) were used in this multi-level study. The first dataset is provided by the Ohio Department of Transportation (ODOT) and is composed of 32,289 interstate, US route, and state route segments. The segments are predominantly classified as principal and minor arterial routes. This dataset, shown in Table 1, includes the following: pavement type, lane width, shoulder width, number of lanes, median presence, horizontal and vertical curve related statistics, the overall vehicle ADT, and the length of the segment. Secondary information may be extracted from the initial dataset to calculate the number of horizontal curves per segment, horizontal curves per mile, maximum degree of curve, and percent of the segment that is a horizontal curve. Similarly, the number of vertical curves, vertical curves per mile, maximum grade, and the percent of the segment that is a vertical curve are extracted from the vertical curve data. In addition to the roadway segments, township information from the ODOT dataset, including the number of lane miles, area of the township, and the urban status of the township, is used to capture information about the 1459 townships in Ohio. Of the 1459 townships, 940 are considered urban townships in this study. A township is designated as urban if an incorporated city, based on the city boundaries provided by ODOT, is located inside the region (ODOT, 2011). All the variables listed above are considered as fixed effects parameters with the exception of the ADT and segment length, which are entered into the model as an offset as shown in order to measure exposure as a rate (see Miaou and Song, 2005 or Lord et al., 2005):

$$e_h = \frac{(\mathrm{ADT}_h * L_h)}{10^6} \quad (1)$$

where ADT_h is the ADT of segment h, L_h is the segment length, and e_h is the value of the offset for the segment. The offset accounts for the exposure of each segment to multi-vehicle motorcycle crashes. Although the data in this study are missing specific measures of motorcycle travel, these two means of exposure along with each random effects term reduce the error due to the lack of available data.

US Census data (US Census Bureau, 2011) were used to include demographic information that described the different regions of Ohio in a manner similar to Aguero-Valverde and Jovanis (2006). Knowledge about the household demographics – such as the percent of residents over age 65, percentage of residents under the poverty level, and the mean travel time to work – was included in the model. In addition to demographic information, the county population, number of motorcycle endorsements (motorcycle licenses), and number of registered motorcycles are used as measures of motorcycle and motor vehicle traffic and are compiled at a regional level.

The number of multi-vehicle motorcycle crashes that occur between 2006 and the summer of 2011 on each segment is determined by combining Ohio crash data as reported by the Ohio Department of Public Safety (ODPS) with the ODOT geographic locations of each roadway segment (ODPS, 2012; ODOT, 2011). A total of 3804 non-intersection related multi-vehicle motorcycle crashes are found to have occurred on state-maintained roadways from 2006 through the summer of 2011; this total includes 68 fatal crashes and 1163 injury crashes. Geographic coordinates, which are used to identify the segment on which each crash occurred, are available for 3379 crashes (ODPS, 2012). Segments with no geographic coordinates and those having unrealistic characteristics (such as excessive lane widths) are removed, and the remaining 3119 multi-vehicle motorcycle crashes are considered in this study.

3. Theory and calculation

Negative binomial modeling with Bayesian inference is commonly used to predict crashes. Within this practice, Bayesian inference differs from traditional statistics in that the parameters are estimated using prior knowledge, as shown in Bayes Theorem:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \quad (2)$$

where $p(\theta|y)$ represents the posterior density, $p(y|\theta)$ denotes the model likelihood, $p(\theta)$ is the known background information, and $p(y)$ represents the unconditional density of the data. Guo et al. (2010) and Congdon (2010) present a detailed description of Bayes' theorem, and the application of Bayesian negative binomial models may be found in examples such as Haque et al. (2010), Mitra (2009), Noland and Quddus (2004) or Quddus (2008).

Within the last few years, researchers such as Aguero-Valverde and Jovanis (2010) and Wang et al. (2009) have been introducing random effects into models so as to include information that may be either unavailable or may be difficult to express in the form of fixed effects. This form of modeling is particularly advantageous in studies such as this, since motorcycle-specific ADT and VMT are difficult to measure. In order to measure the impact of the random effects, three types of Bayesian negative binomial models are considered in this study:

- Uncorrelated heterogeneity model (UH model).
- County and township level random effects model (CT model).
- Spatially correlated random effects models (SC models).

In the UH model, only fixed effects and one random effects term for uncorrelated heterogeneity is specified. In the CT and SC models, full Bayesian negative binomial models are specified with