Accident Analysis and Prevention 98 (2017) 108–117

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Prediction of secondary crash frequency on highway networks

Afrid A. Sarker[a,c], Rajesh Paleti[b], Sabyasachee Mishra[a,c,*], Mihalis M. Golias[a,c], Philip B Freeze[d]

[a] Department of Civil Engineering, University of Memphis, 3815 Central Avenue, Memphis, TN 38152, United States
[b] Civil & Environmental Engineering, Old Dominion University, 135 Kaufman Hall, Norfolk, VA 23529, United States
[c] Intermodal Freight Transportation Institute, University of Memphis, Memphis, TN 38152, United States
[d] Tennessee Department of Transportation, Nashville, TN 37243, United States

ARTICLE INFO

Article history:
Received 14 December 2015
Received in revised form 23 August 2016
Accepted 19 September 2016
Available online 5 October 2016

Keywords:
Secondary crash frequency
Count models
Prediction model
GORP
Segment characteristics
Freeways

ABSTRACT

Secondary crash (SC) occurrences are major contributors to traffic delay and reduced safety, particularly in urban areas. National, state, and local agencies are investing substantial amount of resources to identify and mitigate secondary crashes to reduce congestion, related fatalities, injuries, and property damages. Though a relatively small portion of all crashes are secondary, determining the primary contributing factors for their occurrence is crucial. The non-recurring nature of SCs makes it imperative to predict their occurrences for effective incident management. In this context, the objective of this study is to develop prediction models to better understand causal factors inducing SCs. Given the count nature of secondary crash frequency data, the authors used count modeling methods including the standard Poisson and Negative Binomial (NB) models and their generalized variants to analyze secondary crash occurrences. Specifically, Generalized Ordered Response Probit (GORP) framework that subsumes standard count models as special cases and provides additional flexibility thus improving predictive accuracy were used in this study. The models developed account for possible effects of geometric design features, traffic composition and exposure, land use and other segment related attributes on frequency of SCs on freeways. The models were estimated using data from Shelby County, TN and results show that annual average daily traffic (AADT), traffic composition, land use, number of lanes, right side shoulder width, posted speed limits and ramp indicator are among key variables that effect SC occurrences. Also, the elasticity effects of these different factors were also computed to quantify their magnitude of impact.

1. Introduction

Urban areas are experiencing increasing traffic congestion and significant portion is non-recurring in nature. Secondary crashes (SCs) contribute up to 50% of nationwide urban congestion (Chimba et al., 2014; Kwon et al., 2006; Ozbay and Kachroo, 1999; Skabardonis et al., 1998). SCs are defined as crashes that occur in close proximity of the primary incident's location as a result of either queuing or driver distraction (Margiotta et al., 2012). Earlier studies suggest that up to 15% of reported crashes have occurred partly or entirely as the result of a primary crash (PC) (Raub, 1997). Though, SCs make up a relatively small portion of all the crashes, it is essential to identify the contributing factors and their characteristics. Reducing the occurrence of SCs is a major concern for traffic incident management (TIM) agencies, especially when dispatching rescue vehicles to manage incidents (Dunn and Latoski, 2003; Owens et al., 2010). Limiting the impact of nonrecurring events, such as SCs and disabled vehicles, through effective incident management is one of the objectives of emergency response professionals (Raub and Schofer, 1997) and "reduction of SCs" is considered as one of the performance measures. Understanding the characteristics of secondary crashes can help decision-makers to select better traffic operation practices and safety programs. If SCs can be predicted in real time, it would enable TIM agencies to contain the effects of the primary incident and disseminate the information to the users accordingly to prevent the occurrence of SCs. Such prediction models can be a first step towards alleviating SC effects on congestion, delay, fuel consumption and emissions.

Past researchers have developed econometric models to predict SC occurrence based on primary crash factors, crash severity, incident type, and driver characteristics (Karlaftis et al., 1999; Khattak et al., 2009, 2011, 2010; Zhang and Khattak, 2010). Several studies have focused on identifying contributing factors of SCs identifying

* Corresponding author at: Department of Civil Engineering, University of Memphis, 3815 Central Avenue, Memphis, TN 38152, United States.
E-mail addresses: aasarker@memphis.edu (A.A. Sarker), rpaleti@odu.edu (R. Paleti), smishra3@memphis.edu (S. Mishra), mgkolias@memphis.edu (M.M. Golias), Phillip.B.Freeze@tn.gov (P.B. Freeze).

http://dx.doi.org/10.1016/j.aap.2016.09.019
0001-4575/

peak hour during weekdays, clearance time, lane blockage duration, number of vehicles involved, vehicle type (car, tractor-trailer), and vehicle location as key factors associated with higher SC likelihood (Hirunyanitiwattana and Mattingly, 2006; Karlaftis et al., 1999). Alternatively, one could also model the total frequency of secondary crash occurrences as a function of roadway and traffic characteristics. Such a model would be useful tool to directly quantify the impact of different roadway improvements and policy interventions on "reduction of SCs". Given that segment and traffic characteristics play a crucial role in inducing SCs and they are also readily identifiable, they could be used as explanatory variables in SC frequency models enabling prediction of SC occurrences in real time. In this context, the primary objective of this paper is to develop a set of count data models to predict SC occurrences on freeways as a function of segment and traffic characteristics. This allows us to examine the characteristics of SCs on freeways and significance of their relation to geometric, land use and traffic characteristics. This paper aims to achieve the study objective by developing and comparing alternate count modeling methods for analyzing *SC frequency* on freeways using *only* segment and traffic characteristics that are readily available. The spatial unit of analysis considered in this study was a unit-mile long freeway segment. The study findings will help design effective incident management strategies and thus reduce the likelihood of SC occurrence.

The rest of the paper is organized as follows. The next section discusses best practices and published research on prediction of SC occurrences. The third section, presents the proposed methodology followed by a case study in the fourth section. The final section concludes the paper by summarizing findings.

2. Literature review

In this section we present the relevant literature on different statistical models used in the past for analyzing SC occurrence along with their primary contributing factors. We also explore the application of count data models in prediction of crash frequency.

2.1. Statistical models for SC occurrence

Several past studies have focused on identifying contributing factors of SCs. One of the literature found that the peak hour during weekdays along with clearance time are associated with secondary incidents occurrence. A study by Karlaftis et al. developed a logit model to identify the relation between clearance time of primary incident and SC occurrence in which season, day of week, vehicle type (car, tractor-trailer) and vehicle location were found to be the most significant factors for higher secondary incident likelihood (Karlaftis et al., 1999). Zhan et al. developed a binary logit model to estimate the likelihood of SC occurrence. It was observed that longer the freeway lane blockage duration, higher the likelihood of SCs because of increased congestion and queue length (Zhan et al., 2009). The authors also concluded that SCs are more likely to occur during weekday morning, afternoon peaks and mid-day hours. Another past study found similar results concluding that incident type, lane blockage duration, number of lanes, time of day and number of vehicles involved are some of the key factors associated with SC occurrence (Zhan et al., 2008).

Khattak et al. developed several models for SC occurrence using logistic regression (Khattak et al., 2011). All the models were probit models with certain variations. Because of the presence of endogeneity, the authors used two stage least square (2SLS) method where SC occurrence is estimated using duration as the endogenous variable. The study also found that if the primary incident is a crash involving multiple vehicles and if it is occurring during peak hours on a roadway with high AADT, this primary incident is highly likely to induce a SC. In another study the authors focused primarily on the interdependence between SC occurrence and duration of primary incident and concluded that they are interdependent (Khattak et al., 2009). It suggested that secondary incidents are more likely to occur if the primary incident lasts long and simultaneously durations of primary incidents are expected to be longer if secondary incidents take place.

Yang et al. showed that more than half of SCs occurred from PC-induced queues lasting more than two hours, identifying associated congestion and longer clearance time as primary contributing factors (Yang et al., 2014). Results also revealed that the major contributing factor was "following too closely" along with improper lane change, distracted driving and unsafe speeds.

2.2. Application of count data models in predicting crash frequency

Crash frequency at any given location is a non-negative integer without any pre-specified upper bound. Such data is referred to as count data and parametric models such as the Poisson and Negative Binomial (NB) models were used for analyzing such data. These traditional count models have also been used for predicting SC frequency as a function of geometric and traffic characteristics of roadway segments (Khattak et al., 2011). However, these models have their own set of merits and limitations. For instance, the simple Poisson model is only suited for handling count data with the '*equi-dispersion*' property that implies that the mean and the variance are the same. To relax this restrictive assumption, researchers have used the Negative Binomial (NB) model that accounts for '*over-dispersion*' (variance > mean) that is often the case in crash frequency data. Another important characteristic of count data is over-representation of records with zero count outcome that is referred to as the '*excess zeroes*' problem. Researchers have used the hurdle and zero-inflated variants that assume that the data is generated from two different states – a zero state and normal count process state to account for the excess zeroes property (Lord et al., 2005).

Another aspect that is important to crash frequency analysis is *'unobserved heterogeneity'*. Crashes are rare and random events that depend on a wide array of factors including humans, vehicles, roadway, and weather conditions. It is very likely that the influence of different explanatory variables on crash frequency vary across crash locations and time periods due to the moderating influence of unobserved factors that are not controlled in the statistical model. Mannering et al. (2016) listed several standard explanatory variables used in safety research and the associated reasons for why there might be unobserved variation in the impact of these variables. Ignoring this unobserved heterogeneity and restricting the parameter effects to be the same across all locations can lead to biased estimates and thereby incorrect policy inferences. There are several methodological approaches available in the literature to address the unobserved heterogeneity problem. Mannering et al. (2016) provides an overview of these different approaches along with a detailed summary of their relative merits and limitations. A brief overview of these approaches is presented in this paper (Mannering et al., 2016). The most commonly used approach to address the unobserved heterogeneity problem is the random parameters count modeling approach. In this method, the parameters in the mean specification of standard count models are assumed to be realizations from multivariate random distributions (with normal being the most common distribution) and are integrated out during the evaluation of the log-likelihood function using simulation (Anastasopoulos and Mannering, 2009; Barua et al., 2016; Bullough et al., 2013; Venkataraman et al., 2011). While the random parameters approach is relatively straightforward and easier to estimate, other approaches may be better suited