Journal of Safety Research xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of Safety Research

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

Can post encroachment time substitute intersection characteristics in crash prediction models?

Lakshmi N. Peesapati, [a,*] Michael P. Hunter, [b] Michael O. Rodgers [b]

[a] 15634 Main Street Bellevue, WA 98008, United States
[b] Department of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA 30332, United States

A R T I C L E I N F O

Article history:
Received 29 March 2017
Received in revised form 21 December 2017
Accepted 8 May 2018
Available online xxxx

A B S T R A C T

Introduction: Transportation safety analyses have traditionally relied on crash data. The limitations of these crash data in terms of timeliness and efficiency are well understood and many studies have explored the feasibility of using alternative surrogate measures for evaluation of road safety. Surrogate safety measures have the potential to estimate crash frequency, while requiring reduced data collection efforts relative to crash data based measures. Traditional crash prediction models use factors such as traffic volume, sight distance, and grade to make risk and exposure estimates that are combined with observed crashes, generally using an Empirical Bayes method, to obtain a final crash estimate. Many surrogate measures have the notable advantage of not directly requiring historical crash data from a site to estimate safety. Post Encroachment Time (PET) is one such measure and represents the time difference between a vehicle leaving the area of encroachment and a conflicting vehicle entering the same area. The exact relationship between surrogate measures, such as PET, and crashes in an ongoing research area. *Method:* This paper studies the use of PET to estimate crashes between left-turning vehicles and opposing through vehicles for its ability to predict opposing left-turn crashes. By definition, a PET value of 0 implies the occurrence of a crash and the closer the value of PET is to 0, the higher the conflict risk. *Results:* This study shows that a model combining PET and traffic volume characteristic (AADT or conflicting volume) has better predictive power than PET alone. Further, it was found that PET may be capturing the impact of certain other intersection characteristics on safety as inclusion of other intersection characteristics such as sight distance, grade, and other parameters result in only marginal impacts on predictive capacity that do not justify the increased model complexity.

1. Introduction

The use of crash data for transportation safety analysis has several limitations with respect to both timeliness and accuracy. Crashes are rare events, typically requiring three or more years of crash data to evaluate safety (Nicholson, 1985). The use of surrogate safety measures potentially allows for an earlier safety assessment relative to crash data based analysis. Over the past several decades researchers have sought to identify indicators of traffic conflicts as surrogates for safety. An indirect safety measurement technique that has been in practice since the 1960s is the Traffic Conflict Technique (TCT) and much of the literature available to date focuses on the use of traffic conflicts as surrogate safety measures (Hyden, 1987; Parker Jr. & Zegeer, 1989; Perkins & Harris, 1967). However, this technique is inherently subjective in nature and achieving consistency between observers is a challenge. A white paper on surrogate safety measures (Tarko, Davis, Saunier, Sayed, & Washington, 2009) also proposes that a desirable property for an effective surrogate measure is its ability to be observable or measurable in the traffic system. One observable measure that allows for consistency between observers as well as locations is post-encroachment time (PET). PET is the difference between the time when the first vehicle ends encroachment over the area of conflict and the second vehicle enters the area of conflict. PET requires only two time stamps to compute and it enjoys the advantage of having a definite boundary to differentiate a crash from a non-crash event. A PET value of 0 implies a crash, while non-zero PET values indicate crash proximity. Though it does not describe the initial stage of the conflict nor the action taken by the drivers involved, it shows the resulting event in the final stage and provides a measure of relative closeness to a collision. The current study evaluates the effectiveness of PET as a surrogate safety measure for potential left-turn to opposing through vehicle conflicts.

Crash based safety evaluation is the primary approach for safety analyses and the literature presents many statistical methods to model crash frequency. The generalized linear modeling (GLM) approach is currently the most frequently used technique to model crash counts (Lord et al., 2005). The GLM approach suggests that the actual

* Corresponding author.
E-mail addresses: plnarasimham@gmail.com (L.N. Peesapati), michael.hunter@ce.gatech.edu (M.P. Hunter), michael.rodgers@ce.gatech.edu (M.O. Rodgers).

https://doi.org/10.1016/j.jsr.2018.05.002
0022-4375/

estimated crashes at a location can be assumed to follow a separate distribution, and the mean number of crashes of this distribution is assumed to be related to the model of covariates through a link function. A GLM typically assumes this model of covariates to be of a linear nature. In safety studies, crashes are commonly assumed to follow a Poisson distribution (Fridstrom et al., 1995; Nelder et al., 1972) or a Negative Binomial (NB) distribution (Hauer et al., 1988). Typically, a link function would be log, logit, inverse, or identity.

Parametric modeling and regression analysis techniques are popular even in studies relating to surrogate measures. Though earlier studies such as Parker Jr. and Zegeer (1989) found that the relationship between traffic conflicts and crashes is linear and statistically significant, the exact relationship between surrogate measures and crashes is yet to be established consistently. Djikstra et al. (2010) conducted a study in the Netherlands where they modeled 300 km^2 of road network in PARAMICS. Conflicts were identified from the simulated model and GLM approach was used to develop models to predict crash frequency. A study similar to this paper was conducted by El-Basyouny and Sayed (2013) where they proposed a two-phase approach – one for predicting conflicts based on the intersection characteristics and the second to predict collisions based on predicted conflicts. They found that a NB model showed a significant proportional relationship between conflicts and collisions. However, this study has a limitation in the variety of intersection characteristics considered. Another study (Shahdah, Saccomanno, & Persaud, 2014) developed a new methodology to estimate crash modification factors using conflicts. Such approaches were also used in studies such as (Boonsiripant et al., 2011; Gettman, Pu, Sayed, & Shelby, 2008; Songchitruksa & Tarko, 2006).

This paper explores the use of PET both as a sole predictor of crashes and with a combination of other characteristics of an intersection using GLM techniques. Since a surrogate measure is an indicator of near-crashes and the actual outcome of vehicular interactions at a location, it can be expected that this measure by itself can predict crashes. It is also possible that PET would improve the current models by acting as an additional source of information that explains a part of the unexplained variance.

2. Model building

Generalized linear models (GLM) stem from the concept that linear models can be transformed to create a framework that closely resembles linear models but can accommodate a wide variety of non-normal outcome variables. Nelder and Wedderburn (1972) gives one of the first attempts at developing this framework. A GLM consists of three major components.

(i) Random component: This specifies the characteristic distribution of the response variable with respect to the predictors.
(ii) A linear predictor that is a linear function (η) of predictor variables on which the expected value of response (μ) depends.

$$\eta = \alpha + \beta_1 X_1 + \ldots\ldots + \beta_n X_n \quad (1)$$

(iii) A link function $g(\mu) = \eta$ which links the linear predictor of predictor variables to the expected value of response μ;

GLMs retain their linear character through this link function by which the response and predictor are related. Because the linear predictor is a linear function of explanatory variables, the linear assumption is preserved. However, it should be noted that retaining the linear component can be a limitation of this approach as well. Moreover, the distributions are restricted to certain families (e.g. exponential) and responses are constrained to be independent. While there are several distributions that can be used, the most commonly used methods to model crash counts use Poisson and Negative Binomial regression (Hauer et al., 1988), as crashes have a very small probability of occurrence and they can be classified as count data. The following section describes in detail these regression approaches.

This part of the analysis was performed using R software package (Venables et al., 2013). The function "glm" in R is specifically used to perform the generalized linear modeling analysis.

2.1. Poisson regression (Nelder & Wedderburn, 1972)

Poisson regression assumes that the observed counts are generated from a Poisson distribution. The Poisson distribution is often used to model count data and events that have a low probability of occurrence (e.g., telephone calls arriving in a system, vehicles arriving at a traffic signal, number of claim applications coming to an insurance company). The probability mass function of a Poisson distribution is:

$$P(Y = y) = \lambda^y e^{-\lambda}/y! \quad (2)$$

where

λ = mean number of events in a unit time
y = value of the random variable for which the probability is being estimated

The relation between GLM and Poisson regression is that the mean of the Poisson distribution λ is estimated from the linear predictor of explanatory variables using the link function. The most common link function is the log link, which is expressed as

$$\log(\lambda) = \eta = \alpha + \beta_1 X_1 + \ldots\ldots + \beta_n X_n \\ => \lambda = \exp(\alpha + \beta_1 X_1 + \ldots\ldots + \beta_n X_n) \quad (3)$$

where

$X_1, \ldots.., X_n$ are the explanatory variables and $\beta_1, \ldots\ldots., \beta_n$ are regression coefficients.

2.2. Negative binomial regression (Nelder & Wedderburn, 1972)

One of the properties of a Poisson process is that the mean of the distribution is equal to the variance. This property is often violated for crash counts (Hauer et al., 1988). Data are said to be under-dispersed if variance is less than the mean, and over-dispersed if variance is greater than the mean. Negative binomial regression is normally used in the case of over-dispersed data. Suppose that $Y \sim \text{Poisson}(\lambda)$ and that λ itself is a random variable with a Gamma distribution i.e., $\lambda \sim \text{Gamma}(\alpha, \beta)$ with mean $\alpha\beta$ and variance $\alpha\beta^2$. Therefore, a Negative Binomial distribution is also called as Poisson-Gamma distribution. The PDF of the distribution that λ follows is:

$$f(\lambda) = \left(1/\beta^\alpha\, \Gamma(\alpha)\right) \lambda^{\alpha-1} \exp(-\lambda/\beta) \quad (4)$$

It can be shown that in such a case, Y follows a negative binomial distribution with a mean $\alpha\beta$ and variance $\alpha\beta + \alpha\beta^2$. The negative binomial model is generally expressed in terms of parameters $\mu = \alpha\beta$ and an overdispersion parameter $K = 1/\alpha$. This makes

$$E(Y) = \mu \text{ and } Var(Y) = \mu + K\mu^2. \quad (5)$$

In terms of the parameters μ and K, the negative binomial distribution would be:

$$f(Y) = [\Gamma(1/K + y)/(\Gamma(1/K)y!)]\ [K\mu/(1 + K\mu)]^y\ [1/(1 + K\mu)]^{(1/K)} \quad (6)$$

Please cite this article as: Peesapati, L.N., et al., Can post encroachment time substitute intersection characteristics in crash prediction models?, *Journal of Safety Research* (2018), https://doi.org/10.1016/j.jsr.2018.05.002