Accident Analysis and Prevention 95 (2016) 250–265

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Development of safety performance functions for Spanish two-lane rural highways on flat terrain

Laura Garach, Juan de Oña*, Griselda López, Leticia Baena

TRYSE Research Group, Department of Civil Engineering, University of Granada, ETSI Caminos, Canales y Puertos, c/ Severo Ochoa, s/n, 18071 Granada, Spain

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 15 July 2016
Accepted 18 July 2016

Keywords:
Cumulative residuals
Safety performance functions
Two-lane rural highways
Flat terrain
Parsimonious models
Fully models

ABSTRACT

Over decades safety performance functions (SPF) have been developed as a tool for traffic safety in order to estimate the number of crashes in a specific road section. Despite the steady progression of methodological innovations in the crash analysis field, many fundamental issues have not been completely addressed. For instance: Is it better to use parsimonious or fully specified models? How should the goodness-of-fit of the models be assessed? Is it better to use a general model for the entire sample or specific models based on sample stratifications? This paper investigates the above issues by means of several SPFs developed using negative binomial regression models for two-lane rural highways in Spain. The models were based on crash data gathered over a 5-year period, using a broad number of explanatory variables related to exposure, geometry, design consistency and roadside features. Results show that the principle of parsimony could be too restrictive and that it provided simplistic models. Most previous studies apply conventional measurements (i.e., R^2, BIC, AIC, etc.) to assess the goodness-of-fit of models. Seldom do studies apply cumulative residual (CURE) analysis as a tool for model evaluation. This paper shows that CURE plots are essential tools for calibrating SPF, while also providing information for possible sample stratification. Previous authors suggest that sample segmentation increases the model accuracy. The results presented here confirm that finding, and show that the number of significant variables in the final models increases with sample stratification. This paper point out that fully models based on sample segmentation and on CURE may provide more useful insights about traffic crashes than general parsimonious models when developing SPF.

1. Introduction

According to the World Health Organization, approximately 1.24 million people are killed every year on the world's roads, and another 20–50 million sustain nonfatal injuries as a result of road crashes. All efforts to reduce traffic crashes are therefore well justified. In Europe, approximately 60% of road accident fatalities occur on two-lane rural roads (Cafiso et al., 2010). Two major factors usually play an important role in traffic accident occurrence: the first is related to the driver; and the second is related to the roadway design (Abdel-Aty and Radwan, 2000).

Safety Performance Functions (SPF) make it possible to predict the number of crashes that may take place on a given stretch of roadway with certain characteristics. For many years this type of model was developed using simple or multiple linear regression techniques. However, Miaou (1994) showed that Poisson regression models—or, in the case of overly dispersed data, Negative Binomial (NB) regression models—are more appropriate. Later research showed that, in general, the number of crashes used when calibrating the prediction models presents over-dispersion, with a greater dispersion than would be consistent under a Poisson model (Hauer et al., 2002). Most studies nowadays therefore assume that the number of crashes follows an NB distribution (Persaud et al., 1999; Cheng and Washington, 2008; Montella et al., 2008; Cafiso et al., 2010; FHWA, 2010; Montella, 2010; Camacho-Torregrosa et al., 2013).

Although substantial research has been conducted on the development of crash models, there are issues still on the forefront regarding: generalized models; unobserved heterogeneity; confounding variables; variables to be considered in models and how to add them (parsimonious vs. fully specified models); overfitting of models; measures used in assessing the goodness of fit; and the appropriateness of stratifying a sample to get better models.

The generalized models are used by authorities to study the safety of other locations in a given region that have characteris-

* Corresponding author.
E-mail addresses: lgarach@ugr.es (L. Garach), jdona@ugr.es (J. de Oña), griselda@ugr.es (G. López), lbaenar@ugr.es (L. Baena).

http://dx.doi.org/10.1016/j.aap.2016.07.021
0001-4575/

tics similar to those of the location used to build the model. Thus, models containing variables with highly significant parameters can predict accident frequencies at new locations not used in the model development. In addition, because explanatory variables that have statistically significant model parameters help explain the variability of the accident data, their inclusion in the model improves its fit with the data (Sawalha and Sayed, 2006).

As for the unobserved heterogeneity, the fact that crashes involve complex interactions among human, vehicle, roadway, traffic and environmental elements makes it impossible to take into account all factors influencing the likelihood of highway crashes. Crash databases contain a lot of information about road, vehicle and environment characteristics, yet many other elements remain unobserved, such as human behaviour, friction measurements, etc. These factors constitute unobserved heterogeneity and can introduce variation in the impact of the effect of observed variables on accident likelihood (Mannering et al., 2016). Unobserved heterogeneity can be defined as variations in the effect of variables across the sample population that are unknown to the researcher. If this issue is ignored and the effects of observable variables is held to be the same across all observations, the model may be misspecified and the estimated parameters might be biased, leading to erroneous predictions (Mannering et al., 2016). Although relatively recent research has explored unobserved heterogeneity, allowing new insights to be extracted from crash databases, the model-estimation process involved becomes considerably more complex; the result obtained from methods such as random parameter models may not be easily transferable to other datasets or different locations since the individual parameter vector associated with each observation is unique to that observation (Lord and Mannering, 2010; Mannering et al., 2016).

A further issue that concerns researchers is that of confounding variables. In general, confounding variables are those that are not controlled in the model but may have a latent effect. A confounding factor can be defined as any variable—other than the cause of principal interest in a study—that can either (a) generate effects that may be mixed up with the effects of the causal variable, (b) distort the effects attributed to the causal variable, e.g. modifying their direction or strength, or (c) hide the effects of the causal variable (Elvik, 2011). Controlling for confounding factors is important in establishing causality, and poor control of confounding factors can seriously distort the findings of road safety studies and make them completely worthless (Elvik, 2008). However, the number of potentially confounding factors that are successfully controlled for is always limited due to the fact that most are unknown. Moreover, it is a fallacy to believe that if a model fits the data very neatly, this demonstrates that it includes all important factors and that factors not included in the model cannot have major effects (Elvik, 2011). Hence, this matter may be a limitation in most crash-frequency studies. In the models applied to all accidents, there is slight confounding owing to the mixture of different levels of accident severity (Elvik, 2011).

SPF are used for a variety of purposes. Most frequently they serve to estimate the expected crash frequencies from various roadway entities (highways, intersections, interstates, etc.) and to identify geometric, environmental, and operational factors that are associated with crashes. With respect to the selection of variables, the explanatory variables that are potentially relevant in SPF can be grouped in two main categories: (a) Variables describing exposure to crash risk; (b) Risk factors that influence the number of crashes expected to occur in a road.

In the first category, most studies include Annual Average Daily Traffic (AADT) and section length as exposure variables (Hadi et al., 1995; Anderson et al., 1999; Persaud et al., 1999; Pardillo and Llamas, 2003; Ng and Sayed, 2004; Pardillo et al., 2006; Dell'Acqua and Russo, 2008; Cafiso et al., 2010; Park and Abdel-Aty, 2015). Among the exposure variables, some authors moreover take into account the percentage of heavy vehicles (Fitzpatrick et al., 2000; Elvik, 2007; Ramírez et al., 2009; Montella, 2010; Hosseinpour et al., 2014).

In the second category, among risk factors that influence the number of crashes expected to occur on a highway, most authors consider explanatory variables included in one of the three following groups: geometric variables, consistency variables or context variables. A number of studies have attempted to quantify the effects of road geometric design variables and exposure variables on accident frequencies (Hadi et al., 1995; Persaud et al., 1999; Fitzpatrick et al., 2000; Anastasopoulos et al., 2008; Dell'Acqua and Russo, 2008; Cafiso et al., 2013; Park and Abdel-Aty, 2015). Some authors have looked into the influence of consistency variables—or a combination of geometric, environment and consistency variables—in SPF development for two-lane rural highways (Anderson et al., 1999; Ng and Sayed, 2004; Cafiso et al., 2010; De Oña et al., 2014). Others have developed consistency indexes that may be used as independent variables in SPF (Polus and Mattar-Habib, 2004; Camacho-Torregrosa, 2014; Garach et al., 2014). Some studies have attempted to relate crash frequency with environmental variables such as driveway density (Pardillo and Llamas, 2003; Pardillo et al., 2006; Cafiso et al., 2010).

Within this substantial body of research on SPF development, the vast majority of SPF studies include some kind of measure of exposure, such as AADT or segment length. Still, there is a lack of consensus regarding the number of variables that should be added in the model, and questions relating to parsimonious vs. fully specified models.

According to Sawalha and Sayed (2006), model generality requires that a model be developed in accordance with the principle of parsimony, which calls for explaining as much variability of the data as possible using the least number of explanatory variables. The notion behind the principle of parsimony is to avoid overfitting. If many variables are included in a model, a perfect fit could be obtained; but the developed model would not produce reliable predictions when applied to a different set of locations. In addition, as the data available to researchers is often limited, and many variables known to significantly affect the frequency of crashes may not be available, there is also a need to develop relatively simplistic models using only explanatory variables than can, in practice, be gathered and projected for use. Given these data limitations and the need to specify models with a few simplistic explanatory variables, parsimonious models are often estimated.

However, other authors disagree with the concept of parsimonious models. According to Mannering and Bhat (2014), the real problem with them is that models having a few simplistic explanatory variables might exclude significant explanatory variables; and the model-estimated parameter for the basic variables (like traffic volume) might be estimated with bias (omitted variables bias). The application of the model would be fundamentally flawed, because changes in the omitted variables cannot be captured and predicted crash frequencies will be incorrect. Mannering et al. (2016) indicated that if factors affecting the likelihood of an accident are not included (unobserved heterogeneity), these factors could introduce variation in the impact of the effect of observed variables on accident likelihood. Omission of important variables introduces bias in model parameters, and will lead to incorrect inference (Washington et al., 2010; Mitra and Washington, 2012).

Regarding model evaluation, many studies use statistical measures such as Akaike Information Criterion (AIC) or Pearson Chi-square statistics, among others. Few use cumulative residual analysis as a method to evaluate the calibrated prediction models. Hauer (2015) recommends analysing residual plots as an essential tool to calibrate crash models. Lord and Persaud (2000) applied cumulative residual analysis to evaluate prediction models showing