Accident Analysis and Prevention 70 (2014) 258–266

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Safety performance models for urban intersections in Brazil

Heloisa Barbosa[a], Flávio Cunto[b,*], Bárbara Bezerra[c], Christine Nodari[d], Maria Alice Jacques[e]

[a] Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[b] Universidade Federal do Ceará, Fortaleza, Brazil
[c] Universidade Estadual Paulista, Bauru, Brazil
[d] Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[e] Universidade de Brasília, Brasília, Brazil

ARTICLE INFO

Article history:
Received 19 November 2013
Received in revised form 9 April 2014
Accepted 14 April 2014

Keywords:
Safety performance models
Road safety
Generalized linear models
Observational road safety studies
Generalized Estimating Equations

ABSTRACT

This paper presents a modeling effort for developing safety performance models (SPM) for urban intersections for three major Brazilian cities. The proposed methodology for calibrating SPM has been divided into the following steps: defining the safety study objective, choosing predictive variables and sample size, data acquisition, defining model expression and model parameters and model evaluation. Among the predictive variables explored in the calibration phase were exposure variables (AADT), number of lanes, number of approaches and central median status. SPMs were obtained for three cities: Fortaleza, Belo Horizonte and Brasília. The SPM developed for signalized intersections in Fortaleza and Belo Horizonte had the same structure and the most significant independent variables, which were AADT entering the intersection and number of lanes, and in addition, the coefficient of the best models were in the same range of values. For Brasília, because of the sample size, the signalized and unsignalized intersections were grouped, and the AADT was split in minor and major approaches, which were the most significant variables. This paper also evaluated SPM transferability to other jurisdiction. The SPM for signalized intersections from Fortaleza and Belo Horizonte have been recalibrated (in terms of the C_x) to the city of Porto Alegre. The models were adjusted following the Highway Safety Manual (HSM) calibration procedure and yielded C_x of 0.65 and 2.06 for Fortaleza and Belo Horizonte SPM respectively. This paper showed the experience and future challenges toward the initiatives on development of SPMs in Brazil, that can serve as a guide for other countries that are in the same stage in this subject.

1. Introduction

One of the primary goals for transportation researchers and practitioners is to ensure adequate safety performance of the various transportation components for all road users given the resources available. Historically, safety has been defined and measured in terms of the observed number of crashes in part because of the intuitive and logical link between these two items. This type of approach relies heavily on the reliability and overall quality of accident data systems as well as on statistical models aimed at estimating the expected number of crashes as a function of geometric and operational attributes of the traffic system components, also known as safety performance models (SPMs), safety performance functions (SPFs) or accident prediction models (APMs).

The occurrence of accidents in a given location is known to have a strong random component. In this context, SPM can contribute to measuring real safety performance by attenuating the effect of randomness in the observed crash frequency (Hauer, 2002). SPMs are developed based on police-reported crashes and geometric and operational road attributes as covariates. These models have the potential to improve road safety by comparing alternative road projects regarding expected relative safety performance (AASHTO, 2010). Thus, it is possible to explicitly include road safety criterion in the decision-making process for selecting those projects that are expected to have fewer accidents during a given operation period.

A great deal of the recent worldwide interest in developing SPM can be credited to the release of the first edition of the Highway Safety Manual – HSM (AASHTO, 2010). The HSM devotes much of its content to justification, premises, development and application of SPM in the transportation systems planning process at the strategic, tactical and operational level.

Though SPMs have been explored for more than two decades in countries such as Canada, the USA, England and Sweden, in Brazil,

* Corresponding author. Tel.: +55 8533669488x219; fax: +55 8533669488x201.
E-mail addresses: flaviocunto@det.ufc.br, fjcc07@gmail.com (F. Cunto).

http://dx.doi.org/10.1016/j.aap.2014.04.008
0001-4575/

this methodology is still in its infancy. Few scientific papers dealing with this topic come from under developing countries. Early modeling efforts for the Brazilian environment can be accredited to the doctoral research of Cardoso (2006) and summarized in the later work of Cardoso and Goldner (2007), in which urban arterial segments in the city of Porto Alegre were analyzed.

It is believed that this discrepancy can be attributed in part to problems with the availability and quality of information on traffic accidents associated with the relative scarcity of procedures for calibration and validation of such models nationwide. This situation, along with the increase in crash frequencies in Brazilian urban areas, has resulted in a joint research effort started in 2009 and sponsored by the National Council of Technological and Scientific Development (CNPq). The research group consists of six Brazilian Universities and four international institutions, namely University of Waterloo and Ryerson University from Canada, Lund University from Sweden and *Universidade do Minho* from Portugal.

This paper presents the modeling effort and initial results for the development of safety performance models for urban streets in Brazil that could be useful for worldwide practitioners in the same situation. Given the limited data available in most cities, this project also aimed at generating alternative solutions to obtain data and, at the same time, guide future developers on the best approach to define the functional form of the models, techniques to be adopted in the calibration process and investigating model spatial and temporal transferability.

2. Road safety modeling with observational data

From the engineering perspective, methodologies for safety assessments are heavily influenced by the way safety is defined and measured. Traditionally, the level of safety of a given entity has been defined as "the number of crashes by type and severity, expected to occur on the entity during a specific period" (Hauer, 2002). Representing safety throughout crash events is therefore the natural domain of observational studies.

Observational studies can be viewed as a passive learning process in which the knowledge comes from meticulous analysis of the outcome of events that have not been formally designed to address the problem. According to Davis (2004), an underlying assumption of these studies is that crashes are individually unpredictable, although groups of crashes observed on a given location can produce predictable statistical patterns. Basically, these groups of crashes may be related to a single time period (number of crash related to one or more years are considered globally) or may include longitudinal data. In the latter situation, data belonging to different years for a given location can be treated and analyzed as time series events. In addition, the safety condition of different locations may be related to each other. The presence of temporal or spatial correlations in the database imposes specific statistical considerations for model development (Lord and Persaud, 2000; Wang and Abdel-Aty, 2006).

Under this paradigm, several methods for linking accidents and their consequences to human, vehicle, roadway and environmental attributes have been proposed over the last two decades. These include the use of contingency tables, linear multivariate regression models, logistic models, hierarchical loglinear models, induced exposure models and generalized linear models, among others. A broad review of these methodological alternatives for global crash-frequency data can be found in Lord and Mannering (2010). In addition, the work of Savolainen et al. (2011) presents a general review on methodological alternatives for specific analysis of crash-injury severities.

Because of the relative ability to address some aspects of the inherent stochastic rare random nature of crashes, such as the regression to the mean phenomenon and the crash frequency over dispersion, the generalized linear modeling approach (GLM) has recently become widely applied (Hauer, 2004; Sawalha and Sayed, 2006; Hadayeghi et al., 2007; AASHTO, 2010). Furthermore, a procedure derived from the GLM, called Generalized Estimating Equations (GEE), has been successfully applied in the presence of longitudinal and/or spatial correlations in databases (Lord and Persaud, 2000; Wang and Abdel-Aty, 2006). The most commonly found general expression for SPM can be written as follows (Hakkert and Hochermanand, 1996; Sawalha and Sayed, 2006; AASHTO, 2010):

$$Y = \left[\prod_i (A_i)^{\beta_i} \right] \cdot e^{\Sigma_j(\gamma_j B_j)} \tag{1}$$

where Y, expected number of crashes over a specific time interval (year); A and B, predictive variables; α, β_i and γ_j, model's coefficients.

Initially, SPMs were developed assuming the error structure to be compatible with the Poisson distribution; however, other studies have produced better results by assuming the negative binomial distribution (also known as Poisson-gamma) error structure in cases where crash frequency presents a considerable dispersion (variance greater than average) among similar entities (Bonneson and McCoy, 1993; Persaud and Mucsi, 1995). For all distributions, one important aspect that imposes difficulties for the modeling procedure is the presence of many records of zero crashes in the database. Different approaches have been adopted to address this situation (Shankar et al., 1997; Lee and Mannering, 2002; Kumara and Chin, 2003). However, some of them present limitations when considered under the traffic engineering point of view (Lord et al., 2005, 2007).

SPM expressions have been used as one of the most important component of a methodology to improve crash estimations known as the Empirical Bayes method (EB method), which applies concepts of conditional probability to both the reference population (represented by SPMs) and specific sites to produce a weighted value of the expected number of crashes. The EB estimate of crashes is given by (Hauer, 2002) the following equation:

$$E(m|x) = w \cdot E(m) + (1 - w) \cdot x \tag{2}$$

where $E(m|x)$, the expected number of crashes for entity m given that x crashes have been observed for the same entity; $E(m)$, crash estimate obtained from regression model developed using crash data of similar sites (SPM); w, weight assigned to $E(m)$ $(0 = w = 1)$; x, observational crash data for the site.

The expression that yields the best estimate of w is given by

$$w = \frac{1}{1 + \mathrm{VAR}(m)/E(m)} \tag{3}$$

where VAR(m) is the variance associated with the regression model developed. Basically, the weight w in Eqs. (2) and (3) is a function of the variability found in the data used to develop the crash prediction model. The lower the variation in these data, the higher the weight placed on the model estimates of crashes, i.e., higher level of confidence in the model results. The EB method has been largely adopted in hot-spot identification (network screening) and in before-after analysis (AASHTO, 2010).

Despite the considerably large body of research on SPM development, a general methodology for SPM calibration is currently not available mostly due to a general lack of consensus regarding strategic questions such as: (1) what would be the minimum sample size required to develop "acceptable" models?; (2) how should researchers select the most important variables to be considered in the model formulation?; (3) what is the most adequate model structure?; (4) how do researchers confirm the model usefulness