Accident Analysis and Prevention 70 (2014) 131–139

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Methodology to develop crash modification functions for road safety treatments with fully specified and hierarchical models

Yongsheng Chen[a,*], Bhagwant Persaud[b,1]

[a] Office of Traffic Safety, City of Edmonton, Suite 200, 9304-41 Avenue NW, Edmonton, Alberta T6E 6G8, Canada
[b] Department of Civil Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B2K3, Canada

ARTICLE INFO

Article history:
Received 20 November 2013
Received in revised form 15 March 2014
Accepted 15 March 2014
Available online 12 April 2014

Keywords:
Crash modification function
Hierarchical model
Safety performance function
Crash prediction

ABSTRACT

Crash modification factors (CMFs) for road safety treatments are developed as multiplicative factors that are used to reflect the expected changes in safety performance associated with changes in highway design and/or the traffic control features. However, current CMFs have methodological drawbacks. For example, variability with application circumstance is not well understood, and, as important, correlation is not addressed when several CMFs are applied multiplicatively. These issues can be addressed by developing safety performance functions (SPFs) with components of crash modification functions (CM-Functions), an approach that includes all CMF related variables, along with others, while capturing quantitative and other effects of factors and accounting for cross-factor correlations. CM-Functions can capture the safety impact of factors through a continuous and quantitative approach, avoiding the problematic categorical analysis that is often used to capture CMF variability. There are two formulations to develop such SPFs with CM-Function components – fully specified models and hierarchical models. Based on sample datasets from two Canadian cities, both approaches are investigated in this paper. While both model formulations yielded promising results and reasonable CM-Functions, the hierarchical model was found to be more suitable in retaining homogeneity of first-level SPFs, while addressing CM-Functions in sub-level modeling. In addition, hierarchical models better capture the correlations between different impact factors.

1. Introduction

The three conventional model classes for safety performance functions (SPFs) are baseline models, general annual average daily traffic (AADT) models, and models with covariates (fully specified models) (Lord et al., 2008). Baseline and AADT-only models have AADT as only variable, but the former is only for specified base conditions, while the latter is for general, or average conditions of non-AADT variables. Neither should be applied directly without applying crash modification factors (CMFs) to adjust the model prediction for conditions other than those to which the model applies. Fully specified models, in principle, can be directly applied without the use of CMFs.

Since the development of fully specified models that capture all variables and interactions is still a challenge, the first edition of Highway Safety Manual (HSM) (AASHTO, 2010) did not propose fully specified models. Instead, it recommends a crash prediction algorithm as follows:

$$N_{predicted} = N_{spf\ x} \times (CMF_{1x} \times CMF_{2x} \times \ldots \times CMF_{yx}) \times C_x \quad (1)$$

where $N_{\text{predicted}}$ = predicted average crash frequency for a specific year for a site of type x; N_{spfx} = predicted average crash frequency determined for base conditions of the SPF developed for site type x; CMF_{yx} = Crash modification factors specific to SPF for site type x, and C_x = calibration factor to adjust SPF for local conditions for site type x.

A CMF is simply a multiplicative factor used to reflect the expected change in safety performance associated with the corresponding change in highway design and/or traffic control feature (AASHTO, 2010). Reliable CMFs must be methodologically and statistically valid (Harkey et al., 2008). However many CMFs currently applied were developed by 'naïve' before-after research studies with results that are questionable due to the failure to consider "regression to the mean" effects, and/or to insufficient data (Sayed and de Leur, 2008). This actually led to the exclusion of

---

* Corresponding author. Tel.: +1 780 495 0369.
E-mail addresses: yongsheng.chen@edmonton.ca, yongsheng.chen@ryerson.ca (Y. Chen), bpersaud@ryerson.ca (B. Persaud).
[1] Tel.: +1 416 979 5000x6464.

http://dx.doi.org/10.1016/j.aap.2014.03.012
0001-4575/

many potential CMFs from the first HSM edition (Federal Highway Administration, 2011).

On the other hand, some current CMFs are complex and cannot be quickly processed. For example, the HSM CMF for the shoulder width of multi-lane highway segments is calibrated based on three sub-categories of shoulder widths, and three sub-categories of shoulder types as well as several sub-categories of AADTs. In addition, adjacent sub-categories of shoulder widths are vaguely differentiated, with the result that similar shoulder widths may be estimated to have quite different CMFs. More importantly, one key issue that was not addressed by current CMF applications is whether the effect of a CMF when it is applied alone is different from its effect when applied multiplicatively with other CMFs.

Another unresolved key issue is whether a CMF is fixed or whether a crash modification function is more appropriate. Elvik (2009) concluded it would be better if it shows, instead, how the effect varies as a function of one or more characteristics that influence the size of the effect. This can only be achieved by using a crash modification function (Elvik, 2009).

These issues can, in principle, be addressed with an approach that includes CMF related variables, along with others, in safety models, and which investigates their impacts and correlations by a statistical method. This approach develops safety performance functions (SPFs) with components of crash modification functions that capture quantitative and other effects of CMF factors while accounting for cross-factor correlations.

Studies on crash modification functions (denoted in this paper as “CM-Function”) have been sparse. Gross et al. (2012b) analyzed safety effectiveness of converting signalized intersections to roundabouts and their analysis indicated that the safety benefit of roundabouts for total crashes decreased as traffic volumes increase, a result that led to the development of a crash modification function. Another paper by Gross and others (2012a) raised issues associated with estimating the safety effects of multiple treatments, and argued that if multiple treatments are not independent, and the CMFs are simply multiplied to estimate the combined effect, the result may be an over- or underestimation of the combined treatment effect. As a solution, they developed a framework for investigating interrelationships between treatments, and a matrix was provided to help identify potential overlapping effects. This series of efforts led by Gross et al. (2012a, 2012b) investigated some key issues for current CMFs and explored CM-Functions with traffic volume as the only variable; thus, the results were limited and preliminary in nature and did not lead to well-established CM-functions. Elvik (2009) also did some exploratory research that related existing CMFs to certain factor values in developing a regression curve for a CM-Function. One drawback of this method, however, is that the final regression curve could lose its validity in the light of the fact that some individual CMFs may not be statistically significant. Moreover, each CMF in Elvik's work was generated through a temporal “before-after” approach, each for one pair of before-after CMF factor values; this method would therefore not be appropriate for CMFs for which the related factors could have a continuum values.

Another plausible method is to develop CM-Functions directly from observed data. Unfortunately, a literature search on this topic was not fruitful. This is not surprising, since CM-Functions have been functionally included as part of some SPFs, specifically, fully specified models with covariates. With a fully specified model, the number of crashes, $N_{predicted}$, is predicted by:

$$N_{predicted} = \alpha(AADT)^{\beta_0} \times \exp(\beta_1 x_1) \times \exp(\beta_2 x_2) \times \cdots \times \exp(\beta_m x_m) \quad (2)$$

where $x_1, x_2, \ldots, x_{\mathrm{m}}$ = covariates; $\alpha, \beta_0, \beta_1, \beta_2, \ldots, \beta_{\mathrm{m}}$ = coefficients, and AADT = annual average daily traffic.

Then, $\exp(\beta_1 x_1)$, ..., $\exp(\beta_m x_m)$ are technically CM-Functions respectively for each factor from x_1 to x_{m}.

Evidence of the advantages of using a fully specified model is provided by Lord et al. (2010), who compared this approach to using a baseline model multiplied by CMFs and concluded that the fully specified model produces much less variance. However, the fully specified model contradicts the requirements of ideal CM-Function in three aspects. First, a CM-Function component of a fully specified model is not easily matched with the relevant CMF of a calibrated HSM model due to discrepancy arising from the fact that majority of CMFs are discrete numbers with limited values, while a CM-Function is generally one continuous expression. Secondly, the structure of a fully specified model fails to accommodate the data heterogeneity of local jurisdictions. Furthermore, the fully specified model approach has issues relating to the correlation of variables and the challenge of modeling all interactions. Thus, while the fully specified model may give good or even better predictions, it may not be so useful for estimating the effect of a change in a design feature when designing a road, which is what the HSM predictive algorithm seeks to do.

While the fully specified model remains a plausible CM-Function concept, a new paradigm dealing with safety impact factors deserves exploration. A viable alternative is a hierarchical model that includes the first-level ADT-only SPF and some associated sub-models, respectively, with constant and/or shape parameters of the first-level model. In this, each sub-model is related to a function of one or multiple impact factors. The remainder of paper investigates these approaches by investigating CM-Function development from both fully specified and hierarchical models, with statistical analysis based on sample data from two jurisdictions.

2. Sample data

Two groups of data pertaining to same facility type – 4-leg signalized (4SG in the HSM terminology) intersections – were used. They were provided by the cities of Toronto, Ontario and Edmonton, Alberta, Canada. Table 1 provides summary statistics of these data.

Data heterogeneity is evident in Table 1. Toronto has a data item called “class” which categorizes intersection into 14 classes that basically depend on functional classification of the intersecting roads (arterial, collector, local, or other sub-categories). Edmonton does not have equivalent classes, but it does have an area classification separating urban and suburban settings.

Correlation between explanatory variables and AADTs was analyzed for all sample data, resulting in the correlation coefficient matrix shown in Table 2 and scatter-plot matrices in Fig. 1 (R Development Core Team, 2013), As seen in Table 2, the correlation coefficients between number of approaches with left-turn and right-turn lanes and the two AADT variables are all less than 0.5, suggesting “small strength of association” for these cases. The scatter-plot matrices (Fig. 1) show that each category of explanatory variables contains sufficiently wide ranges of AADTs and the median AADTs show no trend, providing evidence of mutual independence (King, 2013; R Development Core Team, 2013; Lund and Lund, 2013). Given these correlation analysis results, it seems reasonable to assume other explanatory variables can be introduced into models along with AADT variables.

3. Crash modification function developed through fully specified model

For the fully specified model that incorporates CM-Functions, the ideal structure should exactly match the relevant conventional