Accident Analysis and Prevention 94 (2016) 143–152

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Exploring the transferability of safety performance functions

Ahmed Farid*, Mohamed Abdel-Aty, Jaeyoung Lee, Naveen Eluru, Jung-Han Wang

Department of Civil, Environmental and Construction Engineering, University of Central Florida, Orlando, FL 32816-2450, USA

ARTICLE INFO

Article history:
Received 16 November 2015
Received in revised form 18 April 2016
Accepted 27 April 2016
Available online 17 June 2016

Keywords:
Transferability of safety performance functions
Rural divided multilane highway segments
Negative binomial models
Highway safety manual
Calibration factor
Empirical bayes method

ABSTRACT

Safety performance functions (SPFs), by predicting the number of crashes on roadway facilities, have been a vital tool in the highway safety area. The SPFs are typically applied for identifying hot spots in network screening and evaluating the effectiveness of road safety countermeasures. The Highway Safety Manual (HSM) provides a series of SPFs for several crash types by various roadway facilities. The SPFs, provided in the HSM, were developed using data from multiple states. In regions without local jurisdiction based SPFs it is common practice to adopt national SPFs for crash prediction. There has been little research to examine the viability of such national level models for local jurisdictions. Towards understanding the influence of SPF transferability, we examine the rural divided multilane highway models from Florida, Ohio, and California. Traffic, roadway geometry and crash data from the three states are employed to estimate single-state SPFs, two-state SPFs and three-state SPFs. The SPFs are estimated using the negative binomial model formulation for several crash types and severities. To evaluate transferability of models, we estimate a transfer index that allows us to understand which models transfer adequately to other regions. The results indicate that models from Florida and California seem to be more transferable compared to models from Ohio. More importantly, we observe that the transfer index increases when we used pooled data (from two or three states). Finally, to assist in model transferability, we propose a Modified Empirical Bayes (MEB) measure that provides segment specific calibration factors for transferring SPFs to local jurisdictions. The proposed measure is shown to outperform the HSM calibration factor for transferring SPFs.

1. Introduction

According to the Local and Rural Road Safety Program of the Federal Highway Administration (2012), 54% of traffic fatalities in the US occurred on rural roads in 2012. Hence, it is not surprising that safety on rural roads is identified as an area of critical importance in enhancing traffic safety. Safety Performance Functions (SPF) form a crucial part of improving traffic safety by allowing us to predict crash frequency and identify hot spots. SPFs are employed to predict the number of crashes of any type (vehicular, pedestrian or bicyclist) or severity level (fatal (K), incapacitating injury (A), non-incapacitating injury (B), possible injury (C) and property damage only (O)). The SPFs are developed by regressing crash frequency data, with traffic volume and geometric characteristics of segments or intersections. Ordinary linear regression models are inappropriate since crash frequencies are non-negative (Miaou and Lum, 1993; Miaou, 1994; Kim et al., 2005; Garber and Wu, 2001). Generalized linear regression models have been used in recent studies (Sawalha and Sayed, 2006; Taylor et al., 2002; Harnen et al., 2004; Donnell and Mason, 2006). Furthermore, the mathematically appropriate frameworks for SPFs are count modeling approaches such as Poisson and Negative Binomial (NB) models. However, due to the assumption of equal mean and variance, Poisson models are quite restrictive for crash frequency modeling. That is, the variance of the observed crash frequency per site is usually greater than the mean (what is referred to as over-dispersion). The NB framework, by allowing for over-dispersion, is more flexible and has been used as the pillar for SPF development in the safety literature (Miaou and Lum, 1993; Miaou, 1994; Donnell and Mason, 2006; Lord et al., 2005).

The Highway Safety Manual (HSM), of the American Association of State Highway and Transportation Officials (2010), includes several NB based SPFs for different types of roadway segments. The HSM's default NB SPFs for divided rural multilane highway segments are developed based on pooled data from California and Texas (Lord et al., 2008). The rationale behind providing the HSM's default SPFs is to employ these SPFs for a particular jurisdiction if the jurisdiction specific SPFs are not available. To elaborate, if a region does not have localized SPFs, the default SPFs from the HSM would be considered as the SPFs and used for crash fre-

* Corresponding author.
E-mail address: Ahmedtf91@knights.ucf.edu (A. Farid).

http://dx.doi.org/10.1016/j.aap.2016.04.031
0001-4575/

quency prediction. Some jurisdictions might not have the expertise or resources to process the data and estimate SPFs. Hence, if the transferred SPFs are shown to be efficient in predicting crashes, the jurisdiction could simply apply the SPFs instead of developing local ones thereby cutting costs and time. For customizing the SPF to a study region, the SPF is calibrated using an aggregate correction measure based on the ratio of total observed crashes to total predicted crashes. The correction factor can be viewed as correcting the intercept in the SPF to correspond to the study region. In past studies, the HSM's default SPFs were applied to rural divided multilane highway segments in specific states in the US and abroad. The observed and predicted crash frequencies were compared to assess the HSM's SPF prediction accuracy. Furthermore, jurisdiction specific SPF's are compared with those of the HSM in terms of goodness of fit measures.

The current study contributes to growing literature on SPF transferability by considering a detailed and rigorous assessment of SPF transferability across multiple regions. Specifically, we consider transferability of jurisdiction specific SPFs of Florida (FL), Ohio (OH) and California (CA) and compare their predictive performances across the three states. The study effort has two main objectives. First, we explore the influence of pooling data from multiple states for SPF development and compare the transferability of the SPFs developed from the pooled data with those developed from single-state data. We consider two states and three states for pooling the data for SPF development and comparison. The comparison is undertaken based on a transfer index measure. Second, we propose a more disaggregate calibration measure that customizes SPFs from elsewhere for the study region without local SPFs. The predictive performance of the calibrated SPF from the new calibration parameter is compared to the predictive performance of the calibrated SPF using the HSM calibration approach. Even though Florida, Ohio and California are distant states that represent different regions of the nation, they are chosen for this study because pooling data from states having different features enriches the data. That is, the pooled data will better represent regions surrounding the three chosen states and hence the SPFs developed from the pooled data may be applicable to the surrounding regions. It should be noted that transferability research is not limited to microscopic SPFs. For instance, Hadayeghi et al. (2006) assessed the temporal transferability of macroscopic SPFs in Toronto. The concept of transferability is also applicable to travel demand modeling. In one recent study by Sikder et al. (2014), transferability of choice models, which are based on tours rather than trips, among San Francisco Bay Area's counties is evaluated.

2. Literature review

In traffic safety literature, examining transferability of SPFs is a relatively new research topic. The analysis approach involves applying the default HSM's SPFs to a specific jurisdiction with the HSM calibration factor (ratio of the sum of the observed number of crashes to that of crash frequencies predicted by the HSM's SPF). In some cases, the jurisdiction specific SPFs are also developed using the local data and compared with the HSM's SPFs multiplied by the calibration factor. The approach has been employed for studies of rural divided multilane highway segments, among other types of segments and intersections, in the states of Missouri, North Carolina, Oregon, and Alabama and internationally for Messina-Catania, Italy and Riyadh, Saudi Arabia (Sun et al., 2014; Salifu, 2004; Srinivasan and Carter, 2011; Xie et al., 2011; Cafiso et al., 2012; Al Kaaf and Abdel-Aty, 2015; Mehta and Lou, 2013 Salifu, 2004; Srinivasan and Carter, 2011; Xie et al., 2011; Cafiso et al., 2012; Al Kaaf and Abdel-Aty, 2015; Mehta and Lou, 2013). The same approach was chosen for roadway facilities other than rural divided multilane highway segments in Louisiana, Ohio and Regina, Saskatchewan, Canada (Sun et al., 2006; El-Dabaja and McAvoy, 2015; Young and Park, 2012).

Sun et al. (2014), employing KABCO crash data from 2009 to 2011, compiled from the Transportation Management System (TMS) of the Missouri Department of Transportation (DOT), employed the HSM's SPF. The HSM's negative binomial model that takes the following form is used to predict the number of crashes per segment.

$$N_{SPF} = \exp\left[A + B\ln(AADT) + \ln(L)\right] \quad (1)$$

In the model, L, represents the segment length and $AADT$ is the average annual daily traffic, both of which are exposure measures. The regression coefficients are represented by A and B. The model is applicable to segments conforming to the defined base conditions of the HSM only. Therefore, crash modification factors (CMFs) of the HSM are applied to account for deviations from the base conditions. The model, while simplistic with only exposure measures, is quite useful for identifying the hot spots (Salifu, 2004). For the case of Missouri, the calibration factor was calculated to be 0.98 indicating that the HSM's SPF marginally over-predicts the frequency of total crashes in rural divided multilane highway segments in Missouri. Srinivasan and Carter (2011) undertook a similar exercise for rural divided multilane highway segments in North Carolina using data between 2004 and 2008 from the accident analysis system of the North Carolina DOT. The analysis is also conducted for total crashes. The HSM's SPF marginally over-predicted total crashes by yielding a calibration factor of 0.97. Likewise, Xie et al. (2011) estimated KABCO crash frequencies in Oregon using data from 2004 to 2006. The CMFs are also taken into consideration. The prediction analyses arrived at a calibration factor of 0.78 for rural divided multilane highways indicating that the national model over-predicts crashes in Oregon by 22%.

In Alabama, Mehta and Lou (2013), expanded on earlier work by estimating the calibration factor as well as developing a local SPF. The authors employed records of KABCO crashes from the years 2006 to 2009. With the calibration factor, the HSM model was adapted to Alabama to yield a calibrated HSM version. Subsequently, the authors developed local SPFs with different functional forms. The predictions from the calibrated HSM were compared with predictions from the local SPF model using a validation dataset of 2000 segments. The comparison exercise included measures such as the mean absolute deviation (MAD), mean predicted bias (MPB), mean squared predicted error (MSPE), and Akaike's information criterion (AIC). The local SPF developed with additional parameters such as a dummy variable representing the year, the lane width, coefficient for the segment length (restricted to 1 in the HSM SPF) and the posted speed limit outperformed the calibrated HSM SPF.

The calibration of the HSM's SPFs has been undertaken internationally as well. Cafiso et al. (2012) conducted a study in the Messina-Catania region in Italy using KAB crash data from 2005 to 2008. The authors computed calibration factors for every year independently while taking into account the CMFs. The calibration factors range from 1.14 to 1.43. The average calibration factor is 1.26 indicating that the calibrated model under predicts crashes. The research team also developed a local SPF with the AADT, segment length, horizontal curvature and grades as variables. The predictions of the local model are compared to those of the calibrated HSM SPF in terms of the root mean square error. The model predictions are not to a large extent different as per the results. In another international study, using data from Riyadh, Saudi Arabia, Al Kaaf and Abdel-Aty (2015) calibrated HSM models of urban divided multilane roads for fatal and injury (FI) crash records using data from 2004 to 2009. The calibration factor was calculated to be 0.31 using the HSM's SPFs and CMFs indicating substantial over-prediction