Accident Analysis and Prevention 72 (2014) 116–126

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

# Accounting for heterogeneity among treatment sites and time trends in developing crash modification functions

Emanuele Sacchi*, Tarek Sayed

*The University of British Columbia, Department of Civil Engineering, 2002, 6250 Applied Science Lane, Vancouver, B.C., Canada V6T 1Z4*

ARTICLE INFO

*Article history:*
Received 27 March 2014
Received in revised form 12 June 2014
Accepted 13 June 2014
Available online 14 July 2014

*Keywords:*
Crash modification functions
Full Bayes
Treatment sites heterogeneity
Fixed and random effects
Distributed lag models

ABSTRACT

Collision modification factors (CMFs) are commonly used to quantify the impact of safety countermeasures. The CMFs obtained from observational before–after (BA) studies are usually estimated by averaging the safety impact (i.e., index of effectiveness) for a group of treatment sites. The heterogeneity among the treatment locations, in terms of their characteristics, and the effect of this heterogeneity on safety treatment effectiveness are usually ignored. This is in contrast to treatment evaluations in other fields like medical statistics where variations in the magnitude (or in the direction) of response to the same treatment given to different patients are considered.

This paper introduces an approach for estimating a CMFunction from BA safety studies that account for variable treatment location characteristics (heterogeneity). The treatment sites heterogeneity was incorporated into the CMFunction using fixed-effects and random-effects regression models. In addition to heterogeneity, the paper also advocates the use of CMFunctions with a time variable to acknowledge that the safety treatment (intervention) effects do not occur instantaneously but are spread over future time. This is achieved using non-linear intervention (Koyck) models, developed within a hierarchical full Bayes (FB) context. To demonstrate the approach, a case study is presented to evaluate the safety effectiveness of the "Signal Head Upgrade Program" recently implemented in the city of Surrey (British Columbia, Canada), where signal visibility was improved at several urban signalized intersections. The results demonstrated the importance of considering treatment sites heterogeneity and time trends when developing CMFunctions.



## 1. Introduction

An important component of any transportation planning or road design project is the explicit evaluation of the safety performance. This explicit safety evaluation facilitates the quantification of safety impacts resulting from changes in road design and traffic operation parameters (Ng and Sayed, 2004). Quantifying these safety impacts allows decision makers the opportunity to analyze the safety benefits in relation to the cost of the project, leading to the justification and rationalization of road infrastructure investments (Sayed and deLeur, 2005). The ability to quantify the safety impacts can be achieved by utilizing collision/crash modification factors (CMFs). CMFs are multiplicative factors usually represented as single values. A CMF of 1.0 indicates no impact on safety, a CMF greater than 1.0 indicates a negative impact on safety, and a CMF less than 1.0 indicates a positive impact on safety. They are generally based on results from safety evaluation studies such as time series or cross-sectional analysis. A selection of the commonly used CMFs is contained in the 2010's release of the (HSM) (2010) or online at the "Crash Modification Factors Clearinghouse (2014)" repository.

Different techniques are available in the literature to estimate CMFs with the most common being observational before–after (BA) studies or cross-sectional analysis (Gross et al., 2010). CMFs derived from BA studies are based on the change in safety performance due to the treatment implemented. A reliable BA evaluation process should ensure that a change in safety has been caused by the treatment and not by other confounding factors. In road safety evaluation, there are normally several confounding factors that can impact the reliability of the result, such as the regression-to-the-mean (RTM) phenomenon, unrelated effects, and trends. In this respect, the well-known empirical Bayes (EB) method combined with the use of comparison group sites is considered to be able to account for these potential biases (Hauer, 1997; Sayed et al., 2004).

Alternatively, CMFs can be derived from cross-sectional analysis based on a single time period. This type of study compares the collision frequency of a group of locations having a specific component

* Corresponding author. Tel.: +604 822 8785.
*E-mail addresses:* esacchi@mail.ubc.ca (E. Sacchi), tsayed@civil.ubc.ca (T. Sayed).

http://dx.doi.org/10.1016/j.aap.2014.06.016
0001-4575/

of interest (i.e., the treatment for BA studies) to the collision frequency of a group of locations with similar characteristics, yet these locations lack the presence of this specific component. Differences in collision frequency between the two groups can be attributed to the presence of the specific component (treatment). An observational BA study is perceived by many researchers to be the best way to estimate the safety effect of changes in location or traffic characteristics. The reason for the superiority of a BA study is that it is a longitudinal analysis meaning that it bases its results on actual changes that have occurred in one data set over a period of time extending from the before condition to the after condition (Sawalha and Sayed, 2001). Problems with the cross-sectional approach include inappropriate functional forms, potential correlation that might exist among variables in the model such that it is difficult to separate their individual effects on safety, and other unforeseen factors whose inclusion in the model was not possible (Sawalha and Sayed, 2001; Gross et al., 2010). One advantage of cross-sectional studies is related to the fact that the measured variables (e.g., geometric characteristics) can be controlled for the regression model and a CMF or a collision/crash modification function (CMFunction) for a feature can be inferred from the model form and the coefficients (Chen and Persaud, 2014). In this context, CMFs and CMFunctions are believed to be reliable if the results are consistent with the ones from longitudinal studies; this is because regressions, based on cross-sectional data, can fail to capture a causal relationship (Hauer, 2010).

As safety countermeasures can depend on attributes of the treated sites, CMFunctions are considered a better option to describe the effect of a treatment related to one or more characteristics of the measure. Moreover, since observational BA studies are more accurate to evaluate a specific countermeasure, a methodology able to develop CMFunctions from BA study design would generate more reliable results in road safety research.

The CMFs obtained from observational BA studies are usually estimated from applying a safety treatment to a group of sites, calculating the safety impact (i.e., index of effectiveness) for each site and averaging these impacts to reach a single CMF value along with a measure of its uncertainty. The heterogeneity among the treatment locations in terms of their characteristics and the effect of this heterogeneity on safety treatment effectiveness are usually ignored. This is in contrast to treatment evaluations in fields like medical statistics, epidemiology, and biostatistics where variations in the magnitude (or in the direction) of response to the same treatment given to different patients are considered (Higgins and Thompson, 2002; Van Houwelingen et al., 2002). Therefore, this paper introduces an approach for estimating a CMFunctions from BA safety studies that account for variable treatment location characteristics (heterogeneity).

In addition to heterogeneity, the paper also advocates the use of CMFunctions with a time variable to acknowledge that the safety treatment (intervention) effects do not occur instantaneously but are spread over future time periods (El-Basyouny and Sayed, 2012a, 2012b, 2012c; Sacchi et al., 2014). This can be pursued with linear and non-linear intervention models, developed within a hierarchical full Bayes (FB) context.

## 2. Research goal and methodological approach

The main objective of this research is to demonstrate how to account for time trends and heterogeneity among treatment sites in developing CMFunctions from an observational BA study using FB non-linear intervention models.

For the inclusion of heterogeneity in treatment sites, the approach adopted is the use of meta-regressions. Meta-regression is similar in principle to simple/multiple regression, in which a dependent variable is predicted by means of one or more explanatory variables. The main difference is that in meta-regression the outcome variable (i.e., the treatment effectiveness for each single location in this context) is weighted by its own precision so that the resulting CMFunction represents an objective and statistically rigorous model that combines different CMFs. This method is based on meta-analysis which is a structured way of combining knowledge on treatment effectiveness from multiple BA studies. Recently, meta-analyses have been applied successfully to road safety problems (Elvik, 2005, 2009).

In general, the variability in the intervention effects evaluated in different studies is known as statistical heterogeneity and is a consequence of methodological or study location diversity. In this research, however, since the resulting indexes of effectiveness were derived from a single BA study, the methodological diversity was not present. However, heterogeneity among road sites can still be explored by conducting meta-regression. For this reason two kind of regression typologies were explored: fixed-effects and random-effects meta-regressions. The random-effects technique has the added advantage of allowing for extra-variability (residual heterogeneity) among intervention effects not modeled by the explanatory variables (Harbord and Higgins, 2008).

For time trends, the non-linear intervention model (dynamic regression) with FB estimates can identify the lagged effects of the treatment in order to measure its effectiveness over time in the form of CMFunction (El-Basyouny and Sayed, 2012a, 2012b; Sacchi et al., 2014).

To demonstrate the approach advocated in this paper, a case study is presented to evaluate the safety effectiveness of the "Signal Head Upgrade Program" recently implemented in the city of Surrey (British Columbia, Canada), where signal visibility was improved at several urban signalized intersections. These improvements included signal lens size upgrades, the installation of new backboards, reflective tapes added to existing backboards, and the installation of additional signal heads.

## 3. Developing CMFunctions from before–after data

### 3.1. The non-linear intervention (Koyck) model

Consider an observational BA study where collision data are available for a reasonable period of time before and after the intervention. In addition, a set of collision data for the same period of time is available for a comparison group similar to the treatment sites (time-series cross-sectional modeling). Let $Y_{it}$ denote the collision count recorded at site $i$ ($i$ = 1, 2, . . ., $n$) during year $t$ ($t$ = 1, 2, . . ., $m$). Using a hierarchical model, such as Poisson-Lognormal, with site-level random effects it is possible to write:

$$Y_{it}|\theta_{it} \sim \text{Poisson}(\theta_{it}), \quad (1)$$

$$\ln(\theta_{it}) = \ln(\mu_{it}) + \varepsilon_i, \quad (2)$$

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2), \quad (3)$$

where $\sigma_\varepsilon^2$ represents the extra-Poisson variation.

Then, assuming that $Y_{it}$ are independently distributed, it is possible to introduce the non-linear intervention models (El-Basyouny and Sayed, 2012a, 2012b). To introduce this model, the following notation is used: $T_i$ is a treatment indicator (equals 1 for treated sites, zero for comparison sites), $T_{0i}$ is the intervention year for the $i$th treated site and its matching comparison group, $I_{it}$ is a time indicator (equals 1 in the after period, 0 in the before period), $V_{1it}$ and $V_{2it}$ denote the annual average daily traffic (AADT) at the major and minor approaches respectively. Hence, the treatment effects can be modeled using distributed lags along with a first-order autoregressive (AR1) model as a proxy for the time effects (Judge et al., 1988;