Accident Analysis and Prevention 71 (2014) 183–193

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Propensity scores-potential outcomes framework to incorporate severity probabilities in the Highway Safety Manual crash prediction algorithm

Lekshmi Sasidharan[a,*], Eric T. Donnell[b]

[a] Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, ETH Zurich, Wolfgang-Pauli Strasse 15, HIL F 41.1, 8093 Zurich, Switzerland
[b] The Pennsylvania State University, Department of Civil and Environmental Engineering, 212 Sackett Building, University Park, PA 16802, United States

ARTICLE INFO

Article history:
Received 5 March 2014
Received in revised form 2 May 2014
Accepted 26 May 2014

Keywords:
Severity probability
Highway Safety Manual
Propensity score
Potential outcome
Causal model
Lighting

ABSTRACT

Accurate estimation of the expected number of crashes at different severity levels for entities with and without countermeasures plays a vital role in selecting countermeasures in the framework of the safety management process. The current practice is to use the American Association of State Highway and Transportation Officials' *Highway Safety Manual* crash prediction algorithms, which combine safety performance functions and crash modification factors, to estimate the effects of safety countermeasures on different highway and street facility types. Many of these crash prediction algorithms are based solely on crash frequency, or assume that severity outcomes are unchanged when planning for, or implementing, safety countermeasures. Failing to account for the uncertainty associated with crash severity outcomes, and assuming crash severity distributions remain unchanged in safety performance evaluations, limits the utility of the *Highway Safety Manual* crash prediction algorithms in assessing the effect of safety countermeasures on crash severity. This study demonstrates the application of a propensity scores-potential outcomes framework to estimate the probability distribution for the occurrence of different crash severity levels by accounting for the uncertainties associated with them. The probability of fatal and severe injury crash occurrence at lighted and unlighted intersections is estimated in this paper using data from Minnesota. The results show that the expected probability of occurrence of fatal and severe injury crashes at a lighted intersection was 1 in 35 crashes and the estimated risk ratio indicates that the respective probabilities at an unlighted intersection was 1.14 times higher compared to lighted intersections. The results from the potential outcomes-propensity scores framework are compared to results obtained from traditional binary logit models, without application of propensity scores matching. Traditional binary logit analysis suggests that the probability of occurrence of severe injury crashes is higher at lighted intersections compared to unlighted intersections, which contradicts the findings obtained from the propensity scores-potential outcomes framework. This finding underscores the importance of having comparable treated and untreated entities in traffic safety countermeasure evaluations.

1. Introduction

Improvements in traffic safety can be realized by reductions in crash frequency, less severe crash outcomes, or both. Crash modification factors (CMFs) are commonly used to document the expected change (increase or decrease) in crash frequency, either after a safety countermeasure has been implemented, or when comparing a site with a treatment to similar sites without the treatment. A large collection of CMFs, based on scientifically rigorous evaluations, is included in the first edition of the American Association of State Highway and Transportation Officials' Highway Safety Manual (2010). HSM safety prediction algorithms predict the expected number of crashes on a road segment or an intersection based on safety performance functions (SPF) and CMFs. For at-grade intersections, an example HSM safety prediction algorithm is shown in the following equation.

$$N_{predicted\,\mathrm{int}} = N_{spf\,\mathrm{int}} \times C_i \times \prod_{i=1}^{n} CMF_i \qquad (1)$$

---

[*] Corresponding author. Tel.: +41 0764496995.
E-mail addresses: lekshmi.sasidharan@ivt.baug.ethz.ch, lechu18181@yahoo.com (L. Sasidharan), edonnell@engr.psu.edu (E.T. Donnell).

http://dx.doi.org/10.1016/j.aap.2014.05.017

where $N_{predicted\,int}$ is the predicted frequency of crashes per year at an intersection; $N_{spf\,int}$ is the predicted frequency of crashes per year at an intersection with base conditions, estimated using a SPF; C_i is a calibration factor for intersections for a specific geographical area; and CMF_i are crash modification factors for individual geometric design, traffic control features, or other safety treatments, that vary from the base conditions. The CMFs are indexed from $i = 1$ to n, where n is the number of CMFs that differ from the base conditions.

CMFs are useful in adjusting the base predictions and customizing the crash predictions for site-specific roadway and traffic conditions. The safety effect of a treatment or countermeasure can be determined using a CMF. The CMF for roadway intersection lighting included in the HSM is shown in the following equation.

$$CMF_{lighting} = 1 - 0.38 \times p_{ni} \tag{2}$$

where $CMF_{lighting}$ is the crash modification factor for the effect of lighting on total intersection crashes and p_{ni} is the proportion of total crashes for unlighted intersections that occur at night.

The HSM recommends that p_{ni} take the value of 0.260, 0.244, and 0.286 for three-leg stop-controlled, four-leg stop-controlled, and four-leg signalized intersections, respectively. The $CMF_{lighting}$ estimated in Eq. (2) can be used in Eq. (1) to estimate the expected number of total crashes per year at a lighted intersection. However, the role of intersection lighting (countermeasure) in reducing the probability of a fatal or severe injury crash outcome is as likely important as considering the frequency of total crashes when programming safety improvements. It is important to note that none of the studies included in the HSM considered the probability of occurrence of crashes at different severity levels, conditioned on crash occurrence. By estimating the probability of occurrence of crashes at different severity levels, the crash prediction algorithm that is currently used in the HSM could be modified to estimate the expected number of crashes at different severity levels as follows:

$$N_{predicted\,fatal\,\mathrm{int}} = N_{predicted\,\mathrm{int}} \times p_{fatal} \tag{3}$$

where $N_{predicted\,fatal\,int}$ is the number of predicted fatal intersection crashes per year; $N_{predicted\,int}$ is the number of predicted intersection crashes per year per Eq. (1); p_{fatal} = probability of a fatal crash based on the geometry, traffic, and other safety-influencing features present at an intersection.

Eq. (3) can be modified to include other crash severity outcomes, such as severe injury, minor injury, or property-damage only (PDO) crashes. The probability of occurrence of different crash severity levels (p_{fatal}, p_{injury}, p_{pdo}) when estimated for entities without a countermeasure(s) can be employed along with base predictions and entity-specific CMFs to predict the expected number of crashes at different crash severity levels. This can be compared to the crash frequency of different severity levels at entities with the same countermeasure to determine the effectiveness of the countermeasure in changing crash severity outcomes.

The objective of this paper is to explore the applicability of a propensity scores-potential outcomes framework in estimating the probability of occurrence of crashes at different severity levels. Rather than assuming or using fixed severity distributions to estimate the frequency of severe crashes when planning or implementing safety countermeasures, this method estimates the probability distribution for various severity outcomes and considers the uncertainty associated with the estimate. A dual modeling framework, based on regression estimation with propensity score-related variables, is used to estimate the probability of occurrence of different crash severity levels. The “propensity score” in this paper refers to the probability or chance of an intersection receiving lighting given the observed characteristics of the intersection, and is discussed in detail in the methodology section of this paper. The method is demonstrated by estimating the probability of occurrence of fatal and severe injury crashes using roadway intersection lighting data from Minnesota. The study also estimates the relative risk (risk ratio or RR) of unlighted intersections relative to lighted intersections based on the probability of occurrence of fatal and severe injury crashes. The proposed method involves identification of comparable lighted and unlighted intersections (mimicking randomization) via propensity score matching based on pre-defined calipers. A sensitivity analysis of RR, considering a range of caliper sizes in the matching process, was performed to assess how the results of the proposed propensity scores-potential outcomes framework vary. Because the caliper size used in matching produces different sample sizes in the estimation process, the effect of sample size was tested by conducting a model stability analysis by randomly dropping lighted and unlighted intersections from the analysis database. This study also includes a comparison of the results obtained using the RR estimated by the propensity score-potential outcomes framework, to results obtained using traditional binary logit models without propensity scores matching.

2. Background

The CMF Clearinghouse (FHWA, 2013) includes a large number of CMFs for different geometric design, traffic control and other safety treatments, many of which are estimated using different analytical methods, such as before-after observational studies (e.g., Hauer et al., 2002; Hauer and Persaud, 1983; Persaud et al., 2009, 2007a, 2007b, 2004; Harwood et al., 2002), cross-sectional studies (e.g., Lord and Mannering, 2010; Lord, 2006; Lord et al., 2010; Tarko and Kanodia, 2004; Donnell et al., 2010; Tsyganov et al., 2009), epidemiological case–control studies (e.g., Gross et al., 2009; Gross and Jovanis, 2007), and meta-analysis (Elvik, 1995; Bahar, 2010). The following section briefly describes each of these methods, including the advantages and limitations of the method.

2.1. Conventional methods to determine CMFs

The conventional methods that are used to determine CMFs for the installation of a treatment or safety countermeasure include observational before–after studies (e.g., empirical Bayes [EB] method, comparison group, yoked comparison), cross-sectional statistical modeling (e.g., Poisson regression, negative binomial regression), and epidemiological research methods (case–control or cohort studies). The EB method, which is accepted as the state-of-the-art observational before-after method in traffic safety research, estimates CMFs based on a SPF. The SPF is estimated from a reference group and is used to determine the expected number of crashes that would have occurred in the after period, had the treatment not been implemented. This estimate is then compared to the number of crashes that occurred in the after period at the treatment site(s) (Hauer, 1997).

The advantages of the EB method are that it properly accounts for regression-to-the-mean, and accounts for differences in traffic volume and crash trends between the before and after periods at the treatment sites. The limitations associated with the EB method are that it requires time to pass after a traffic safety countermeasure has been applied before an analysis can be completed; the traffic safety countermeasure of interest is often not the only change that has occurred at the treatment site(s) during the analysis time period; and, determination of treatment installation dates is often challenging in practice. Persaud and Lyon (2007) suggests that the reference group for EB method must be representative of the treated entities in terms of geometric design, traffic volumes, vehicle fleet and so on. However, no guidelines exist concerning how to most effectively select the group of reference sites for developing SPFs.