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
Available online 12 June 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.

© 2014 Elsevier Ltd. All rights reserved.

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.

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

* 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
0001-4575/© 2014 Elsevier Ltd. All rights reserved.

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.

$$CMF_{lighting} = 1 - 0.38 \times p_{ni} \quad (2)$$

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:

$$N_{predicted\,fatal\,\mathrm{int}} = N_{predicted\,\mathrm{int}} \times p_{fatal} \quad (3)$$

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.

Cross-sectional studies estimate CMFs by comparing the crash frequencies or rates between sites with and without the safety countermeasure of interest. The analysis of cross-sectional data is useful when there is more than one countermeasure applied to an entity; the before period data are not available; data exhibits under, over or equidispersion; or, the countermeasure installation date is unknown. Cross-sectional data from sites with and without the safety treatment of interest typically includes information about the crash, traffic volume, geometric features of the roadway, traffic control, posted speed limit, and other site-specific data. After a database of roadway characteristics and crash data are assembled, CMFs are estimated using statistical or econometric models of crash frequency. The CMF estimated using cross-sectional analysis methods may be biased if the traits of treated and untreated entities are not comparable, or they lack control for confounding variables. However, cross-sectional methods allow for inclusion of a large number of sites in the study sample and do not require a time sequence to estimate the effectiveness of a traffic safety countermeasure.

Case–control studies are suitable to analyze rare outcomes such as crashes when the number of confounding variables is small. Gross et al. (2009) and Gross and Jovanis (2007) are two examples of recent traffic safety research that have used case–control studies to estimate the CMF associated with different lane and shoulder widths. In a case–control study, the safety effect of a countermeasure is estimated from matched cases and controls using an odds ratio. The CMF is computed as the ratio of the odds of an event (crash) occurring in cases to the odds of it occurring in controls. An odds ratio of 1.0 indicates that the event (crash) is equally likely to occur in both groups. An odds ratio greater than 1.0 indicates that the event is more likely to occur in cases than in controls and vice versa. A drawback to the case–control methods relates to the number of confounding factors that can be considered in the matching process. When a large number of confounders are present, it is difficult to match cases and controls. Furthermore, matched case–control studies consider only binary outcomes (crash vs. non-crash). It is not clear if the method can be modified to consider multiple crash events when estimating CMFs (Gross and Donnell, 2011).

Many of the CMFs available in the CMF Clearinghouse estimated using the above methods consider only the effects of a safety countermeasure on crash frequency, while few consider the change in probability of severity outcomes. This paper aims at demonstrating the estimation of probability of occurrence of target crashes (fatal and severe injury crashes) for a countermeasure (roadway intersection lighting) using a propensity scores-potential outcomes approach.

2.2. Roadway intersection lighting

Nighttime driving can be particularly problematic as the risk of dying in a crash increases significantly at night. According to NHTSA (2007), the fatality rate per vehicle mile of travel is about three times higher at night than during the day. Roadway intersection lighting, a potential countermeasure to improve nighttime safety, was identified as a traffic safety countermeasure with the highest benefit cost ratio (21:1) among 27 low-cost safety treatments used in Minnesota (MnDOT, 2001). The treatment effect of intersection lighting reported in published studies is based on the frequency of nighttime, daytime or total crashes; nighttime crash rates; or night-to-day crash ratios. Few studies have estimated the effect of intersection lighting with regard to severity.

Based on a meta-analysis of 37 crash studies involving the effect of fixed roadway lighting systems, Elvik (1995) reported a 65 percent reduction in nighttime fatal crashes, a 30 percent reduction in nighttime injury crashes, and a 15 percent reduction in nighttime PDO crashes. Preston and Schoenecker (1999) reported a 50 percent reduction in injury and fatal crashes after installing roadway lighting at rural, at-grade intersections. Isebrands et al. (2004) reported a 41 percent reduction in nighttime injury/fatal crashes after the installation of lighting based on before-after studies conducted using data from Minnesota DOT. In 1957, Blythe suggested a 30 percent reduction in the total number of nighttime crashes and a 93 percent reduction in fatalities after lighting was installed at intersections in Indiana using a naïve before-after study (Schwab et al., 1982).

All of the above studies are based on the frequency of different crash severity levels. In this paper, an estimate of the probability of occurrence of fatal and severe injury crashes for lighted and unlighted intersections ($p_{fatal\ and\ severe}$), which later can be used per Eq. (3) to predict the number of fatal and severe injury crashes for lighted and unlighted intersections, is developed. An advantage of using this method is that, unlike the HSM, which uses a default distribution for different crash and severity types, we assess the probability distribution for different severity levels based on site-specific geometric and traffic conditions while considering the uncertainty associated with the estimate.

2.3. Propensity scores-potential outcomes approach

The propensity scores-potential outcomes framework is a type of causal inference method used to determine the effect of treatments or interventions based on observational, non-randomized data. These models are common in medical, economic, political and educational research (D'Agostino, 1998; Gelman and Meng, 2004; Dehejia and Wahba, 2002; Yanovitzky et al., 2005). There are different causal analysis methods in use, such as Rubin's causal model (RCM) based on potential outcomes (Rubin, 1973, 1978; Rubin and Thomas, 1996; Dehejia and Wahba, 1997; Little and Rubin, 2000; Schafer and Kang, 2008), graphical models based on causal diagrams (Pearl, 1995; Karwa et al., 2011), and sufficient-component cause models (Greenland and Brumback, 2002).

Propensity score modeling considers the conditional probability of an entity receiving treatment given the covariates (X) and outcomes (Y). When the treatment assignment mechanism is unconfounded, the propensity score $\hat{p}$ is represented as shown in the following equation:

$$\hat{p} = P(T = 1|X) \tag{4}$$

where T is the treatment status ($T=1$, treated [with countermeasure] and $T=0$, untreated [without countermeasure]); X is the covariates influencing treatment selection.

Randomized experiments, in which the entities are randomly assigned to treatments or controls, is the ideal way to estimate treatment effects. In this method, all entities are assumed to have equal probability of receiving the countermeasure. Furthermore, the random assignment ensures that the treated and untreated groups (with respect to probability) were the same before the treatment and that any difference in the outcome is due to the treatment effect, and not due to differences between two groups resulting from the assignment process. However, the probability of entities receiving the countermeasure ($\hat{p}$) is not equal in observational non-randomized studies. The $\hat{p}$ is unknown in observational studies and must be estimated from the sample based on important covariates influencing the treatment selection. The $\hat{p}$ is commonly estimated using logistic regression (Sasidharan and Donnell, 2013; Schafer and Kang, 2008; Hosmer and Lemeshow, 2003; Menard, 2002). In previous work by Sasidharan and Donnell (2013), the $\hat{p}$'s were used to identify lighted and unlighted intersections with the same propensity of receiving lighting, thereby mimicking randomization.

The propensity scores help to balance the covariates in the treated and untreated groups when the data are non-randomized and observational as in the case of traffic safety data. Balancing the covariates in the treated and untreated groups occurs by reducing the bias due to differences in observed covariates, thereby mimicking randomization between treated and untreated groups. According to Rosenbaum and Rubin (1983), the treatment assignment is strongly ignorable if unconfoundedness and common overlap hold, both of which are shown in Eqs. (5) and (6), respectively.

$$Y_i(0), Y_i(1) \amalg T_i|X_i \tag{5}$$

$$0 < \hat{p} < 1 \tag{6}$$

Potential outcomes are the outcomes for any entity "i" corresponding to two treatment conditions, treated ($T_i = 1$) and untreated ($T_i = 0$), observed simultaneously under the same period of time. For instance, if an intersection can be lighted or unlighted at the same time, then the number of crashes for that intersection observed during lighted (treated) and unlighted (untreated) conditions are the potential outcomes. However, this is not possible in practice (it is not possible to install and not install lighting for an intersection at the same time). An analogous concept occurs in randomized experiments, in which the treatments are assigned to homogenous entities by chance (all entities have equal probability of receiving and not receiving treatment). In this study, propensity scores are used to mimic randomization and subsequently estimate potential outcomes. These potential outcomes were used to determine the probability of occurrence of fatal and severe injury crashes for lighted and unlighted intersections.

Karwa et al. (2011) suggest that the "risk ratio (RR)," which is the ratio of the probability of occurrence of target crashes at untreated entities to the probability of occurrence of target crashes at treated entities, is a good measure to report the safety effectiveness of a treatment based on severity. The risk ratio, or relative risk, of unlighted intersections based on the probability of occurrence of fatal and severe injury crashes was also determined in this study using the potential outcomes. The gold standard of the unbiased and efficient estimate of RR, if all potential outcomes are observed, is shown in the following equation:

$$RR = \frac{\sum_{i=1}^{n} Y_i(1)/n}{\sum_{i=1}^{n} Y_i(0)/n} \tag{7}$$

where n is the total number of entities; $Y_i(1)$ is an outcome when all entities are treated and $Y_i(0)$ is an outcome when all entities are untreated.

Because of the missing potential outcomes or counterfactuals, the RR shown in Eq. (7) cannot be estimated in an observational study. The unbiased estimate of RR for a completely randomized study is shown in Eq. (8):

$$\hat{RR} = \frac{\sum_i T_i Y_i(1)/\sum_i T_i}{\sum_i (1-T_i) Y_i(0)/\sum_i (1-T_i)} \tag{8}$$

where $\sum_i T_i$ is the total number of treated entities and $\sum_i (1-T_i)$ is the total number of untreated entities.

In observational studies such as those used in traffic safety research, countermeasures may be installed based on outcomes (crashes). This dependence can be considered by making use of important covariates that affect the outcome and countermeasure selection process. These covariates assist in recovering information about the counterfactuals and do not redefine the causal effect that is estimated (Schafer and Kang, 2008). An unbiased estimate of RR can be determined from observational data, provided it satisfies the following assumptions:

1. *Stable unit treatment value assumption (SUTVA)* (Rubin, 1990): This assumption states that the treatment applied to one entity does not affect the outcome of any other (i.e., no interference among the entities).
2. *Positivity*: This assumption requires that there be a non-zero probability of receiving every level of treatment for the combination of values of exposure and covariates that occur among entities in the population (Rubin, 1978; Hernan and Robins, 2006). The positivity assumption can be made when each homogenous entity can be assigned to the treatment or non-treatment group.
3. *Unconfoundedness:* The treatment assignment mechanism is said to be unconfounded if the treatment status (T_i) is conditionally independent of the potential outcomes, given a set of covariates X_i. This is represented as shown in the following equation:(9)$T_i \amalg Y_i(0), Y_i(1)|X_i$

The treatment status T_i is unconditionally independent of the potential outcomes [$Y_i(1)$ and $Y_i(0)$] by design in the case of a randomized experiment. For non-randomized observational data, however, independence is achieved by balancing observed covariates using methods such as propensity score matching, stratification, inverse propensity score weighting, weighted residual bias correction, etc. (Sasidharan and Donnell, 2013; Schafer and Kang, 2008).

The remainder of the paper is organized as follows. The next section discusses the analysis methodology and is followed by a description of the data. The data analysis and results sections follow, and the paper concludes with a discussion and conclusions resulting from the research.

3. Analysis methodology

In this paper, the use of a propensity scores-potential outcomes framework to estimate the probability of occurrence of fatal and severe injury crashes for lighted and unlighted intersections is proposed. In addition, the caliper size used in the propensity scores matching is assessed based on a sensitivity analysis. The RR is the measure used to determine the estimated probabilities of fatal and severity injury crash occurrence in the sensitivity analysis.

This section describes the methodology used in the study to determine the probability of occurrence of different crash severity levels based on the severity of crashes. The methodology includes the following two steps:

(1) mimicking randomization by identifying comparable treated and untreated entities, and
(2) estimating severity probabilities using regression estimation with propensity score-related variables based on the treated and untreated entities identified in step 1 and estimating RR as the ratio of the severity probabilities for untreated and treated entities.

A detailed description on the estimation of propensity scores for treated and untreated entities in the traffic safety context can be found in Sasidharan (2011), and Sasidharan and Donnell (2013), while a more general discussion of the various propensity score estimation procedures can be found in Schafer and Kang (2008). These studies also provide a discussion on different methods used to identify comparable treated and untreated entities based on propensity scores, such as matching, stratification, inverse propensity score weighting, and weighted residual bias corrections.

3.1. Identification of comparable treated and untreated entities

In this paper, we have used a common sampling scheme in the potential outcomes literature to identify comparable treated and untreated entities, which is referred to as pair-wise matching based on propensity scores (Sasidharan and Donnell, 2013; Schafer and Kang, 2008; Park and Saccomanno, 2007; Dehejia and Wahba, 1997; Rubin, 1973; Rosenbaum and Rubin, 1985; Rubin and Thomas, 1992). The basic idea is to pair each treated entity to an untreated entity with the most comparable propensity score. Matching on propensity scores mimics the results of a randomized block experiment in which entities having the same propensity scores are randomly assigned to treated or untreated groups (Schafer and Kang, 2008). Matching based on propensity scores provides advantages over matching based on individual covariates as the latter method is cumbersome and will lead to a large loss of data when there are many confounding variables. Matching based on the propensity score, estimated using all important covariates influencing the treatment selection, ensures that the distribution of observed covariates for the treated and untreated groups is approximately the same for matched groups.

In this paper, matching was done using 1: 1 nearest neighbor (NN) matching with predefined calipers. Details of 1:1 NN matching can be found in Sasidharan and Donnell (2013). If $\hat{p}_i$ is the estimated propensity score, then for each entity in the treated group, a pool of potential matches in the untreated group is identified whose logit propensities are in the interval $\hat{p}_i \pm c$, where c is the caliper distance specified for matching. The treated entity was then matched to the untreated entity in the pool with the closest propensity score. NN matching was done without replacement of untreated entities after matching. The caliper "c" value needs to be specified to ensure precise matching. However, there is no specific c value that should be used, so this study also examined the change in treatment effect as a result of varying the caliper size.

After matching the treated and untreated entities, the balance of covariates for the matched sample was checked. To eliminate the dependence on sample size, the balance of covariates was assessed using the standardized bias or standardized difference in means (Schafer and Kang, 2008; Sasidharan and Donnell, 2013). The standardized bias is estimated as the difference of the sample means in the treated and untreated sub-samples (full or matched) as a percentage of the square root of the average of the sample variances in the treated and untreated groups (Rosenbaum and Rubin, 1985) as shown in the following equation.

$$\text{Standardized bias} = \frac{100 \times (\bar{x}_T - \bar{x}_{UT})}{\sqrt{(S_T^2 + S_{UT}^2)/2}} \tag{10}$$

where $\bar{x}_T$ is the sample mean in the treated group; $\bar{x}_{UT}$ is the sample mean in the untreated group; S_T^2 is the sample variance in the treated group and S_{UT}^2 is the sample variance in the untreated group.

A properly balanced group of treated and untreated entities is shown mathematically in the following equation:

$$P(X_i|p_i = a, T_i = 1) = P(X_i|p_i = a, T_i = 0) \tag{11}$$

where p_i is the propensity score ($0 < a < 1$); X_i is the set of covariates and T_i is the treatment status (1 = with treatment, 0 = without treatment).

After identifying the comparable treated and untreated entities based on propensity scores, the probabilities of occurrence of fatal and severe injury crashes at treated and untreated entities were estimated using regression estimation with the propensity scores method.

3.2. Regression estimation with propensity scores

Fatal and severe injury crashes were considered target crashes in this paper and were coded as a binary variable (taking value 1 if the entity experienced at least one target crash during the analysis period and 0 otherwise). Probability distributions for the occurrence of target crashes were estimated separately using binary logistic regression for the matched treated and untreated groups. The matched dataset for treated and untreated groups were divided into five groups of approximately equal size based on the 20th, 40th, 60th, and 80th percentiles of the estimated propensity scores. Cochran (1968) suggests that dividing data into five groups will remove about 90 percent of the selection bias from an estimate of a population mean. The indicator variables corresponding to these subclasses are also included as predictors in the binary logistic regression models, along with other covariates, because the inclusion of these subclass categorical variables are expected to remove 90 percent of the bias in the regression estimate (Cochran, 1968; Little and An., 2004; Schafer and Kang, 2008; Karwa et al., 2011). The probability of occurrence of target crashes (θ) using a binary logit model is estimated as shown in the following equation.

$$\theta = \frac{\exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n\right)}{\left(1 + \exp\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n\right)\right)} \tag{12}$$

where $\beta_0,\ \beta_1, \ldots, \beta_n$ is the coefficients estimated by maximum likelihood method; $x_1,\ x_2, \ldots, x_n$ is the covariates and n is the number of covariates.

The expected probabilities of occurrence of target incidents were estimated for both treated and untreated groups. The RR was then estimated as shown in the following equation:

$$RR = \frac{E[\theta_{iUT}(1)]}{E[\theta_{iT}(1)]} \tag{13}$$

where RR is the risk ratio; $E[\theta_{iT}(1)]$ is the expected probability of occurrence of target incidents for the treated group; $E[\theta_{iUT}(1)]$ is an expected probability of occurrence of target incidents for the untreated group.

We assume that $E[\theta_{iT}(1)] > 0$. The estimated RR is the relative risk of target incident occurrence for the untreated group. The confidence interval for the risk ratio was estimated using the bootstrap method. Bootstrapping is a resampling method in which a number of resamples of the observed dataset is generated by random sampling with replacement from the original dataset. These resamples are used to estimate the models, and a histogram of the RR is developed, which is known as the bootstrap distribution of RR from which the confidence interval for RR is obtained by trimming off $100(\alpha/2)$ percent from the lower and upper ends of the distribution (Poi, 2004). A sensitivity analysis was conducted to examine the effect of different caliper values on the risk ratio. Model stability was tested by estimating risk ratios for different sample sizes, generated by randomly dropping observations for treated and untreated entities.

3.3. Traditional analysis using binary logit models

As a means of comparison, binary logit models for sites with and without lighting were estimated using the entire dataset (all entities), irrespective of its chances for being in the treated or untreated group. The methodology employed is consistent with Eq. (12), which considers the probability of occurrence of target crashes (θ). Corresponding risk ratios were estimated in a manner consistent with Eq. (13). The risk ratios estimated in this analysis were compared to those that were estimated using the risk ratio method described in Section 3.2.

4. Example modeling application

As described previously in this paper, the countermeasure selected in this study was fixed roadway intersection illumination in Minnesota. The analysis database (same database as in Sasidharan and Donnell, 2013) contained information from the

Table 1
Crashes/intersection/year by different levels of severity.

Crash severity levels	Unlighted intersections		lighted intersections	
	Crashes/int/year	Proportion	Crashes/int/year	Proportion
Fatal crashes	0.011	0.014	0.012	0.005
Incapacitating injury crashes	0.026	0.034	0.048	0.019
Non-incapacitating Injury	0.115	0.152	0.287	0.116
Possible injury crashes	0.166	0.219	0.552	0.222
Property damage only crashes	0.101	0.134	0.389	0.157
Unknown	0.339	0.447	1.19	0.480

Minnesota Highway Safety Information System (HSIS) roadway inventory and crash data files. Four years (1999–2002) of crash and corresponding roadway inventory data were used in the analysis. A total of 25,832 observations (from 6464 intersections) were available for analysis. More than 42 percent of the intersections contained some form of roadway lighting. The crash data were coded as fatal, incapacitating injury, non-incapacitating injury, possible injury, unknown injury, and property damage only crashes (KABCO scale). For this study, the interest was on fatal and severe injury crashes. Details concerning the number of crashes per intersection per year, and the proportion of crashes in each severity level for lighted and unlighted intersections are shown in Table 1.

Table 1 shows that, even though the number of crashes per intersection per year was higher at lighted intersections when compared to unlighted intersections (mean traffic volumes at lighted intersections were 2.5 times higher than at unlighted intersections), the proportion of target crashes (fatal and incapacitating injury crashes) were lower at lighted intersections when compared to unlighted intersections.

The following describes how the modeling assumptions were met in the present study:

a. *SUTVA*: Crashes were coded as intersection-related if they occurred within 250 feet of an intersection. The distance between intersections was not measured and was unavailable in the database; however, it was assumed that the intersections were far enough apart so that lighting installed at one intersection did not influence the outcome (crashes) at another intersection, thereby satisfying the SUTVA assumption.
b. *Positivity*: In the present study, none of the intersections had an estimated propensity score of zero or one, indicating that there was a positive chance that every intersection may be lighted or unlighted.
c. *Unconfoundedness*: The warrants for intersection lighting in Minnesota (MnDOT, 2003) indicate that fixed roadway illumination systems may be installed based on traffic volumes; crash frequency; the presence of a traffic signal, flashing beacons, or school crossings; the presence of intersection channelization on high-speed (40 mph or higher) approaches; or, ambient light that adversely affects driver visibility. The covariates that influence the installation of lighting (based on lighting warrants) were included in the propensity score model. Pre-treatment outcome (crashes) information was not available because the data were cross-sectional and, therefore, covariates that influence the occurrence of crashes were also included in the propensity score model to better predict the treatment assignment mechanism.

Table 2 shows the definitions and descriptive statistics for continuous variables and proportions for categorical variables. 888 of the 6464 intersections were signalized while the remaining 5576 operated under stop-controlled conditions. There were three intersection forms coded in the database: cross, tee, and skew. Approximately 49 percent of the intersections were four-legged, “cross” intersections. Nearly 40 percent were three-legged, “tee” intersections. The remaining 11 percent of the intersections were skew. There were 38,437 reported crashes in the analysis database. All crashes that occurred within 250 feet of where intersection roadway alignments cross were included in the database, except those involving crashes on icy roads, which totaled 827 crashes. These crash types, while a small proportion of the sample, were excluded because lighting is unlikely to affect these events.

5. Results

This section includes the propensity score model estimated using important covariates influencing the installation of intersection lighting. It also includes the results of the NN matching process, which was used to identify comparable lighted and unlighted intersections. Finally, the estimation of the probability of occurrence of fatal and severe injury crashes using the regression estimation method is described in this section.

5.1. Propensity score model

The propensity score model developed for the treatment assignment (presence of intersection lighting) is shown in Table 3. The pre-treatment outcome (crash) data were not available and, therefore, variables that influence the outcome (crashes) were included to form a rich propensity score model that specified the treatment assignment mechanism. The propensity score model for the present study is the same as the model estimated in Sasidharan and Donnell (2013). Propensity scores were estimated using binary logistic regression as only two treatment levels (lighting vs. no lighting) were considered. Covariates that have signs in the expected direction, and interaction terms that can influence the treatment selection, were also included in the propensity score model, irrespective of statistical significance.

As shown in Table 3, the following variables are associated with a lower probability of lighting presence at an intersection: minor stop control indicator, high speed indicator, depressed median indicator, two lane major road indicator, skew intersection type indicator, divided major road indicator, and the urban–suburban and no left shoulder interaction. The following variables are associated with a higher probability of lighting presence at intersections: urban–suburban indicator, signal control indicator, no access control indicator, no left shoulder indicator, no right shoulder indicator, log (major AADT), log (minor AADT), skew-minor stop interaction, urban–suburban and two lane interaction, high speed and two lane interaction, high speed and depressed median interaction, and high speed and no right shoulder interaction. Of particular note is that the majority of the covariates that are associated with an increase in the propensity of an intersection receiving lighting (coefficient > 0) are warrants for lighting in Minnesota (e.g., urban–suburban area type, signal control, traffic volumes). This suggests that the estimated propensity score model is in accordance with the actual treatment assignment for lighting.

Table 2
Variable definitions and proportions of Minnesota data.

Continuous variables

Variable	Min.	Max.	Mean	Std. dev.
Night crash frequency (per year)	0	28	0.366	0.969
Day crash frequency (per year)	0	55	1.121	2.457
Major road average daily traffic	40	77,430	8284	9381
Percentage heavy vehicles on major road	0	61.11	8.888	5.1092
Minor road average daily traffic	1	77,430	3164	5179

Categorical variables

Variable	Proportion in sample
Area type indicator	1: 45.6%
(1 = urban/suburban; 0 = rural)	0: 55.4%
Traffic control indicator	0: 85.5%
(1 = signal; 2 = minor stop-controlled; 0 = all way stop-controlled)	1: 13.7%
	2: 0.80%
Lighting indicator	1: 42.1%
(1 = present; 0 = not present)	0: 57.9%
Intersection type indicator	1: 10.1%
(1 = skew; 0 = cross or tee)	0: 89.9%
[a]High speed indicator	1: 67.3%
(1 = 50 mph or greater; 0 otherwise)	0: 32.7%
[a]No access control indicator	1: 93.7%
(1 = no access; 0 = partial access control)	0: 6.3%
[a]Depressed median indicator	1: 11.5%
(1 = depressed median; 0 = barrier or no median)	0: 88.5%
[a]Two lane	1: 76.5%
(1 = Two lane; 0 = otherwise)	0: 23.5%
[a]Paved left-shoulder indicator	1: 53.9%
(1 = paved shoulder; 0 = unpaved or no shoulder)	0: 46.1%
[a]No left shoulder	1: 29.0%
(1 = no left shoulder; 0 = otherwise)	0: 71.0%
[a]Paved right-shoulder indicator	1: 49.1%
(1 = paved shoulder; 0 = unpaved or no shoulder)	0: 50.9%
[a]No right shoulder	1: 24.2%
(1 = no right shoulder; 0 = otherwise)	0: 75.8%
[a]Divided	1: 20.4%
(1 = divided; 0 = undivided)	0: 79.6%

[a] Indicates that data were available for the major intersecting roadway only.

5.2. Identification of comparable lighted and unlighted intersections

The comparable lighted and unlighted intersections were selected for further analysis based on propensity scores. This was done by plotting the histograms for the distribution of logit propensity scores for matched and unmatched lighted and unlighted intersections, and based on the absolute standardized difference in bias for the predictors for lighted and unlighted intersections. For details of the histogram, please refer to Sasidharan and Donnell (2013). The graph showing the absolute standardized bias or absolute standardized difference in means for the covariates considered in the study before and after 1:1 NN matching is shown in Fig. 1.

Fig. 1 shows that the percent bias, or absolute standardized difference in means, is large for the unmatched lighted and unlighted intersection sample when compared to the 1:1 matched intersections. Fig. 1 shows that the percent bias for the matched sample is less than five percent for all covariates except the major road AADT variable, for which the bias is 10 percent.

5.3. Regression estimation using propensity scores

The binary logistic regression models for the matched lighted and unlighted intersections are shown in Table 4. The probability of occurrence of fatal and severe injury crashes for lighted and unlighted intersections are the potential outcomes in this study. The expected probability of occurrence of target crashes for lighted and unlighted intersections are estimated using binary logistic regression.

Propensity 1, Propensity 2, Propensity 3 and Propensity 4 are the categorical variables created based on the 20th, 40th, 60th and 80th percentiles, respectively, of propensity scores. The expected probability of occurrence of target crashes for matched lighted intersections was 0.029 (i.e., one in every 35 crashes at a lighted intersection is a fatal or severe injury outcome). The expected probability of occurrence of target crashes estimated for matched unlighted intersections is 0.033, which indicates that one in every

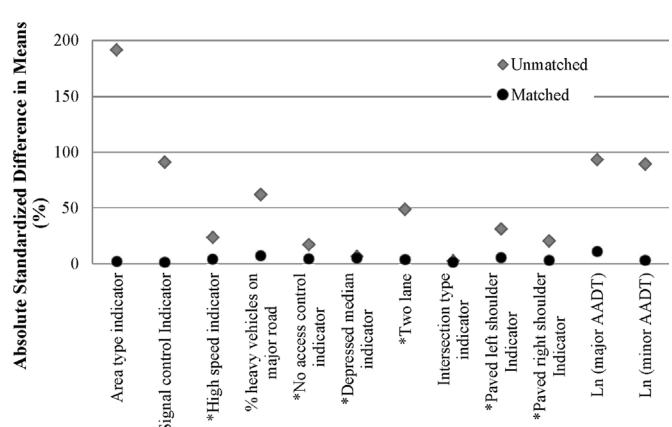

Fig. 1. Absolute standardized difference in means before and after 1:1 NN Matching.
* Data were available for the major intersecting roadway only.

Table 3
Propensity score model.

Variable	Coefficient	Std. err.	Z	$P>z$	95% Conf. interval	
Natural logarithm of major road AADT	0.538	0.028	18.91	<0.001	0.482	0.594
Natural logarithm of minor road AADT	0.176	0.014	12.9	<0.001	0.149	0.202
Percentage heavy vehicles on major road	−0.005	0.005	−1.02	0.307	−0.015	0.005
Area type indicator (1 = urban/suburban; 0 = rural)	2.177	0.095	22.98	<0.001	1.991	2.363
Minor stop control indicator (1 = minor stop; 0 = otherwise)	−0.543	0.249	−2.18	0.029	−1.032	−0.055
Signal control indicator (1 = signalized; 0 = otherwise)	2.071	0.269	7.7	<0.001	1.544	2.598
High speed indicator (1 = 50 mph or greater; 0 otherwise)	−0.621	0.158	−3.93	<0.001	−0.93	−0.312
[a]No access control indicator (1 = no access; 0 = partial access control)	0.148	0.097	1.53	0.127	−0.042	0.338
[a]Depressed median indicator (1 = depressed median; 0 = barrier or no median)	−0.516	0.171	−3.02	0.003	−0.85	−0.181
[a]No left shoulder indicator (1 = no left shoulder; 0 = otherwise)	1.318	0.212	6.21	<0.001	0.902	1.733
[a]No right shoulder indicator (1 = no right shoulder; 0 = otherwise)	0.452	0.199	2.27	0.023	0.062	0.842
Skew intersection type indicator (1 = skew; 0 = cross or tee)	−0.122	0.244	−0.5	0.617	−0.601	0.357
[a]Two lane (1 = two lane; 0 = otherwise)	−0.615	0.146	−4.22	<0.001	−0.900	−0.329
[a]Divided road indicator (1 = divided; 0 = otherwise)	−0.345	0.131	−2.64	0.008	−0.601	−0.089
Skew-minor stop interaction	0.499	0.254	1.97	0.049	0.002	0.996
Urban–suburban and no left shoulder interaction	−0.507	0.136	−3.72	<0.001	−0.774	−0.24
Urban–suburban and two lane interaction	0.433	0.101	4.27	<0.001	0.234	0.632
High speed and two lane interaction	0.264	0.152	1.74	0.082	−0.034	0.561
High speed and depressed median interaction	0.658	0.19	3.47	0.001	0.287	1.03
High speed and No left shoulder interaction	−0.894	0.214	−4.18	<0.001	−1.313	−0.475
High speed and No right shoulder interaction	1.387	0.224	6.18	<0.001	0.947	1.826
Constant	−6.625	0.411	−16.11	<0.001	−7.431	−5.819
Number of observations = 22,092	LR $\chi^2(22)$ = 15,582.57					
Log likelihood = −7444.67	Psuedo R^2 = 0.5114					

[a] Indicates that data were available for the major intersecting roadway only.

Table 4
Binary logistic regression for matched lighted and unlighted intersections.

Target crashes	Lighted intersections			Unlighted intersections		
	Coef.	SE	Z	Coef.	SE	z
Intersection type indicator (1 = skew; 0 = cross or tee)	0.401	0.316	1.27	1.559	0.281	5.56
Area type indicator (1 = urban/suburban; 0 = rural)	−0.7	0.446	−1.57	−0.423	0.281	−1.5
[a]High speed Indicator (1 = 50 mph or greater; 0 otherwise)	−0.01	0.23	−0.04	−0.265	0.218	−1.22
Ln(AADT-major-night)	0.585	0.169	3.46	1.195	0.228	5.24
Ln(AADT-minor-night)	0.193	0.072	2.68	0.494	0.09	5.46
[a]Percentage Heavy vehicles on major road	0.006	0.027	0.23	−0.015	0.028	−0.52
[a]No access control indicator (1 = no access; 0 = partial access control)	−0.046	0.337	−0.14	1.035	0.311	3.33
[a]Depressed median indicator (1 = depressed median; 0 = barrier or no median)	−0.147	0.499	−0.29	0.989	0.593	1.67
[a]No left shoulder indicator (1 = left shoulder absent; 0 = otherwise)	−0.04	0.548	−0.07	0.657	0.569	1.15
[a]No right shoulder indicator (1 = right shoulder absent; 0 = otherwise)	−0.924	0.673	−1.37	−1.254	0.637	−1.97
[a]Two lane indicator (1 = two lane; 0 = otherwise)	0.019	0.379	0.05	0.104	0.395	0.26
Propensity 1	0.026	0.475	0.05	0.856	0.578	1.48
Propensity 2	−0.828	0.456	−1.81	−0.269	0.326	−0.82
Propensity 3	0.094	0.652	0.14	1.156	0.679	1.7
Propensity 4	−0.14	0.751	−0.19	−0.652	0.568	−1.15
Constant	−9.038	2.142	−4.22	−19.215	2.565	−7.49
	No of observation = 2433			No of observation = 2433		
	LR chi^2(16) = 80.28			LR chi^2(16) = 227.88		
	Log likelihood = −390.78			Log likelihood = −373.54		

[a] Indicates that data were available for the major intersecting roadway only.

Table 5
Binary logistic regression for all lighted and unlighted intersections (no propensity scores matching).

Target crashes	Lighted intersections			Unlighted intersections		
	Coef.	SE	Z	Coef.	SE	z
Intersection type indicator (1 = skew; 0 = cross or tee)	0.31	0.14	2.15	0.61	0.17	3.53
Area type indicator (1 = urban/suburban; 0 = rural)	−0.27	0.14	−1.97	−0.61	0.18	−3.32
[a]High speed indicator (1 = 50 mph or greater; 0 otherwise)	−0.06	0.10	−0.63	−0.16	0.12	−1.35
Ln(AADT-major-night)	0.51	0.09	5.58	0.42	0.08	5.16
Ln(AADT-minor-night)	0.32	0.05	7.07	0.39	0.04	8.91
[a]Percentage heavy vehicles on major road	0.00	0.01	−0.17	−0.02	0.01	−1.29
[a]No access control Indicator (1 = no access; 0 = partial access control)	−0.06	0.14	−0.43	0.23	0.20	1.12
[a]Depressed median indicator (1 = depressed median; 0 = barrier or no median)	0.48	0.19	2.55	0.44	0.43	1.00
[a]No left shoulder indicator (1 = left shoulder absent; 0 = otherwise)	0.34	0.20	1.68	0.45	0.45	1.00
[a]No right shoulder indicator (1 = right shoulder absent; 0 = otherwise)	−0.47	0.18	−2.59	−1.19	0.56	−2.14
[a]Two lane indicator (1 = two lane; 0 = otherwise)	−0.22	0.15	−1.41	−0.23	0.29	−0.78
Constant	−10.18	0.94	−10.80	−9.74	0.93	−10.43

[a] Indicates that data were available for the major intersecting roadway only.

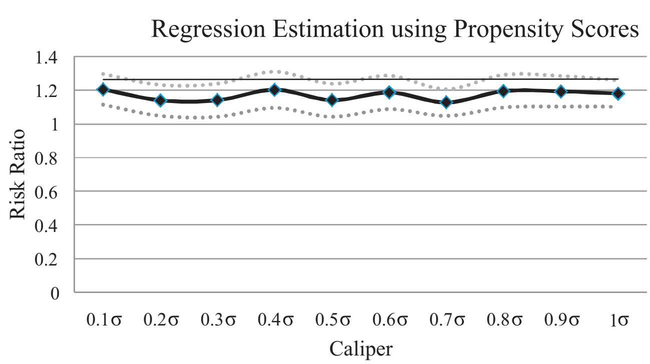

Fig. 2. Risk ratios for different caliper values.

30 crashes is a target crash. The RR is 0.033/0.028 = 1.14. This indicates that the probability of a target crash occurring at unlighted intersections is 1.14 times higher than at lighted intersections. The bootstrap confidence interval estimated for the RR estimate ranges from 1.04 to 1.23.

Sensitivity analysis was conducted for different caliper values ranging from 0.1σ to 1σ, where σ is the within group standard deviation of the propensity scores. Fig. 2 shows the risk ratios for different caliper values and the 95 percent bootstrap confidence intervals.

Fig. 2 shows that the estimates of RR for the occurrence of target crashes at matched intersections are comparable for different caliper values. The bootstrap confidence intervals for all caliper values overlap indicating that the results do not vary much among calipers ranging from 0.1σ to 1σ. To test the model stability, 25 percent, 50 percent and 75 percent of the lighted and unlighted intersections were dropped randomly from the database. The risk ratios were estimated for a randomly selected caliper value of 0.2σ.

Table 6 shows the risk ratios and bootstrap confidence intervals for three different samples after dropping 25 percent, 50 percent and 75 percent of the lighted and unlighted intersections.

Table 6 shows that the risk ratios estimated for a caliper value of 0.2σ for three different sets of samples are comparable to the estimated risk ratio for the entire sample when using the same caliper value. The standard errors increase as more observations are dropped, which was expected as the standard errors highly depend on the sample size. Because of the higher standard errors, the bootstrap confidence intervals are wide when a large number of observations are dropped.

5.4. Traditional analysis using binary logit models

The binary logistic regression models for all lighted and unlighted intersections, irrespective of propensity scores, were estimated for comparison purposes. Table 5 shows the results of the traditional binary logit analysis. The expected probability of occurrence of target crashes for lighted intersections was estimated as 0.046 (i.e., one in every 21 crashes at lighted intersection is a fatal or severe injury crash). The expected probability of occurrence of target crashes estimated for unlighted intersections was 0.028 (i.e., one in every 36 crashes is a target crash) and the RR is 0.028/0.046 = 0.608. This indicates that the presence of intersection lighting increases the probability of occurrence of a target crash at lighted intersections by approximately 40 percent compared to unlighted intersections. This result contradicts with the findings of most published literature on lighting. This result might be due to the fact that the lighted intersections considered in this study have much higher traffic volumes, have complex intersection geometry, or are located on high-speed roads when compared to unlighted intersections. This result suggests that strategies are needed to identify comparable treated and untreated entities

Table 6
Risk ratios for different sample sizes.

Sample	Sample size		Coef.	S.E.	Z	P > z	95% CI	
	Unmatched	Matched						
25% dropped	16,539	3508	1.157	0.066	16.920	<0.001	1.028	1.286
50% dropped	12,928	2424	1.170	0.071	17.400	<0.001	1.030	1.309
75% dropped	6465	1092	1.165	0.129	9.040	<0.001	0.912	1.417

before performing statistical analysis that lead to the determination of a treatment effect for a safety countermeasure.

6. Conclusions

This paper demonstrated the application of the propensity scores-potential outcomes framework to estimate the probability of occurrence of fatal and severe injury crashes using fixed roadway intersection lighting data from Minnesota. The probabilities were estimated based on regression estimation with the propensity score-related variables method. The method also yielded RR for the presence of intersection lighting under non-experimental settings in which the treated group differed from the pool of untreated groups. Comparable lighted and unlighted intersections (with observed covariates) were identified using propensity score based 1:1 NN matching with calipers. The standardized biases for observed covariates were much lower for matched intersections when compared to the unmatched ones, indicating that the intersections were comparable after matching. For these pseudo-randomized intersections, the probabilities of target crash occurrence at unlighted and lighted intersections were estimated. The expected probability of occurrence of a fatal or incapacitating injury crash at an unlighted intersection was 1 in every 30 crashes compared to 1 in every 35 crashes for lighted intersections. The RR, estimated as the ratio of these probabilities, indicated that the probability of occurrence of fatal and severe injury crashes at unlighted intersections was 1.14 times higher than at lighted intersections. The results of the model stability analysis for different samples indicated that the estimated risk ratios were comparable for different sample sizes even though the standard errors increase as the sample size decreases. The comparison of the results of propensity score-potential outcomes analysis to the RR estimated using traditional binary logit models show the importance of using propensity scores in the identification of comparable treated and untreated entities in safety analysis. The traditional logit models, without propensity scores matching, indicated that the probability of target crashes at lighted intersections was approximately 40 percent higher than the probability of target crashes at unlighted intersections, which is not consistent with published literature on the safety effects of intersection lighting.

Few published studies have determined the role of a countermeasure in preventing the occurrence of fatal or severe injury crash outcomes. The propensity score-potential outcomes framework discussed in this paper is a method that is widely accepted in other fields such as statistics, politics, medicine, and economics, to determine the effectiveness of interventions from non-randomized observational data. However, this method has not yet been thoroughly evaluated in traffic safety countermeasure evaluations. The regression estimation method using propensity-based covariates appears to be viable method to determine safety effectiveness of countermeasures based on severity from readily available observational data. Another advantage of using this method is that the treatment effect estimate is doubly robust (i.e., it is not biased even if one of the two models [propensity score or safety estimates] is incorrect) as a result of using the dual modeling approach (separate models for treatment assignment mechanism and safety estimates) (Elliott and Little, 2000; Little and An., 2004; Kang and Schafer, 2007; Schafer and Kang, 2008).

Unlike randomized experiments in which both observed and unobserved variables are randomly distributed across groups, the propensity score-potential outcomes framework only mimics randomization based on observed covariates and, therefore, hidden bias may exist. Addressing this limitation is problematic. Using a hypothetical hidden bias of varying magnitudes, Rosenbaum (2002) tried to examine the influence of bias on the conclusions of the study. However, even in this analysis, there was no way to know the existence and magnitude of a hidden bias. Therefore, future research on this topic in the propensity scores-potential outcomes framework is suggested within the context of traffic safety countermeasure evaluations.

In summary, the doubly robust potential outcomes-propensity scores framework was successfully used to estimate the probability of occurrence of fatal and severe injury crashes at lighted and unlighted intersections. This work contributes to the existing published literature by providing a method to integrate severity probability outcomes in the Highway Safety Manual crash prediction algorithm, in place of the fixed severity probability distributions used in the current framework. This framework could be used effectively to estimate the probability of occurrence of different crash severity levels for any traffic safety countermeasure, which in turn offers a mechanism to consider both crash frequency and severity effects of safety countermeasures in the Highway Safety Manual crash prediction algorithm framework.

References

American Association of State Highway Transportation Officials, 2010. Highway Safety Manual, Washington, DC.

Bahar, G., 2010. Methodology for the development and inclusion of crash modification factors in the first edition of the highway safety manual. In: Transportation Research Circular E-C 142. Transportation Research Board, Washington, DC.

Cochran, W.G., 1968. The effectiveness of adjustment by subclassification in removing bias in observational studies. Biometrics 24, 205–213.

D'Agostino Jr., R.B., 1998. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat. Med. 17, 2265–2281.

Dehejia, R., Wahba, S., (Ph.D. dissertation) 1997. Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. In: Econometric Methods for Program Evaluation. Harvard University (Chapter 1).

Dehejia, R., Wahba, S., 2002. Propensity score matching methods for non-experimental causal studies. Rev. Econ. Stat. 84 (1), 151–161.

Donnell, E.T., Porter, R.J., Shankar, V.N., 2010. A framework for estimating the safety effects of roadway lighting at intersections. Saf. Sci. 48 (10), 1436–1445.

Elliott, M.R., Little, R.J.A., 2000. Model-based alternatives to trimming survey weights. J. Off. Stat. 16, 191–209.

Elvik, R., 1995. Meta-analysis of evaluations of public lighting as accident countermeasure. In: Transportation Research Record 1485. Transportation Research Board, National Research Council, Washington, DC, pp. 112–123.

Federal Highway Administration, 2013. Crash Modification Factor Clearinghouse, Available at: http://www.cmfclearinghouse.org/ (accessed on 28.11.13).

Gelman, A., Meng, X.L., 2004. Applied Bayesian Modeling and Causal Inference From Incomplete-data Perspectives. Wiley, New York.

Greenland, S., Brumback, B., 2002. Theory and methods an overview of relations among causal modeling methods. Int. J. Epidemiol. 31, 1030–1037.

Gross, F., Donnell, E.T., 2011. Case–control and cross-sectional methods for estimating crash modification factors: comparisons from roadway lighting and lane and shoulder width safety effect studies. J. Saf. Res. 42 (2), 117–130.

Gross, F., Jovanis, P.P., 2007. Estimation of the safety effectiveness of lane and shoulder width: case–control approach. J. Transport. Eng. 133 (6), 362–369.

Gross, F., Jovanis, P.P., Eccles, K.A., Ko-Yu, C., 2009. Safety evaluation of lane and shoulder width combinations on rural, two-lane, undivided roads. In: Final Report, FHWA-HRT 09-031.

Harwood, D.W., Bauer, E., Potts, I.B., Torbic, D.J., Richard, K.R., Kohlman Rabbani, E.R., Hauer, E., Elefteriadou, L., 2002. Safety Effectiveness of Intersection Left- and Right-turn Lanes. FHWA-RD-02-089. Federal Highway Administration, McLean, Va.

Hauer, E., Persaud, B., 1983. A common bias in before and after comparison and its elimination. In: Transportation Research Record 905. Transportation Research Board. National Research Council, Washington, DC, pp. 164–174.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M., 2002. Estimating safety by the empirical Bayes method: a tutorial. Transport. Res. Rec. 1784, 126–131.

Hauer, E., 1997. Observational before-after studies in traffic safety. Pergamon, New York.

Hernan, M.A., Robins, J.M., 2006. Estimating causal effects from epidemiological data. J. Epidemiol. Commun. Health 60, 578–586.

Hosmer, D.W., Lemeshow, S., 2003. Applied Logistic Regression, 2nd ed. Wiley, New York.

Isebrands, H., Hallmark, S., Hans, Z., McDonald, T., Preston, H., Storm, R., 2004. Safety Impacts of Street Lighting at Isolated Rural Intersections: Part II, Year 1 Report. Center for Transportation Research and Education. Iowa State University, Ames, IA.

Kang, J.D.Y., Schafer, J.L., 2007. Demystifying double robustness: a comparison of alternative strategies for estimating population means from incomplete data. Stat. Sci. 26, p.523–539.

Karwa, V., Slavkovic, A.B., Donnell, E.T., 2011. Causal inference in transportation safety studies: comparison of the potential outcomes and causal Bayesian networks. Ann. Appl. Stat. 5 (2B), 1428–1455.

Little, R.J., Rubin, R.B., 2000. Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu. Rev. Publ. Health 21, 121–145.

Little, R.J.A., An., H., 2004. Robust likelihood-based analysis of multivariate data with missing values. Statistica Sinica 14, 949–968.

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accid. Anal. Prevent. 38 (4), 751–766.

Lord, D., Geedipally, S.R., Guikema, S., 2010. Extension of the Application of Conway-Maxwell-Poisson Models: Analyzing Traffic Crash Data Exhibiting Under-Dispersion. Risk Anal. 30 (8), 1268–1276.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transport. Res. A, http://dx.doi.org/10.1016/j.tra.2010.02.001.

Menard, S., 2002. Applied Logistic Regression Analysis, 2nd ed. Sage, Thousand Oaks, CA.

Minnesota Department of Transportation, 2001. Traffic Safety Fundamentals Handbook. Office of Traffic Safety, St. Paul, MN.

Minnesota Department of Transportation, 2003. Roadway Lighting Design Manual. Office of Traffic Safety, St. Paul, MN.

National Highway Traffic Safety Administration, 2007. Traffic Safety Facts: Passenger Vehicle Occupant Fatalities by Day and Night – A Contrast.

Park, Y.P., Saccomanno, F.F., 2007. Reducing treatment selection bias for estimating treatment effects using propensity score method. J. Transport. Eng. Am. Soc. Civil Eng. 133 (2), 112–118.

Pearl, J., 1995. Causal diagrams for experimental research. Biometrika 82, 669–710.

Persaud, B.N., Lyon, C., 2007. Empirical Bayes before-after safety studies: lessons learned from two decades of experience and future directions. Accid. Anal. Prev. 39 (3), 546–555.

Persaud, B.N., Retting, R.A., Lyon, C.A., 2004. Crash reduction following installation of centerline rumble strips on rural two-lane roads. Accid. Anal. Prev. 36, 1073–1079.

Persaud, B.N., Lyon, C., Nguyen, T., 2007a. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. In: Transportation Research Record 1665, Paper No. 99-0534.

Persaud, B.N., Retting, R.A., Garder, P.E., Lord, D., 2007b. Safety effect of roundabout conversions in the United States empirical Bayes observational before–after study. In: Transportation Research Record 1665, Paper No. 99-0534.

Persaud, B.N., Lan, C., Lyon, R., Bhim, 2009. Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. Accid. Anal. Prevent..

Poi, B.P., 2004. From the help desk: some bootstrapping techniques. Stata J. 3, 312–328.

Preston, H., Schoenecker, T., 1999. Safety Impacts of Street Lighting at Rural Intersections. Report No. 1999-17. Minnesota Department of Transportation, St. Paul, MN.

Rosenbaum, P.R., 2002. Observational Studies, 2nd ed. Springer, New York, NY.

Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70, 41–55.

Rosenbaum, P.R., Rubin, D.B., 1985. Reducing bias in observational studies using the subclassification on the propensity score. J. Am. Stat. Assoc. 79, 516–524.

Rubin, D.B., 1973. The use of matched sampling and regression adjustment to remove bias in observational studies. Biometrics 29, 185–203.

Rubin, D.B., 1978. Bayesian inference for causal effects: the role of randomization. Ann. Stat. 6, 34–58.

Rubin, D.B., 1990. Neyman (1923) Causal inference in experiments and observational studies. Stat. Sci. 5, 472–480.

Rubin, D.B., Thomas, N., 1992. Characterizing the effect of matching using linear propensity score methods with normal distributions. Biometrika 79, 797–809.

Rubin, D.B., Thomas, N., 1996. Matching using estimated propensity scores: relating theory to practice. Biometrics 52, 249–264.

Sasidharan, L., 2011. Causal Modeling Approach to Determine the Effectiveness of Traffic Safety Countermeasures. The Pennsylvania State Univeristy, Univeristy Park USA (P.hD dissertation).

Sasidharan, L., Donnell, E.T., 2013. Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: exploratory analysis using intersection lighting data. Accid. Anal. Prev. 50, 539–553.

Schafer, J.L., Kang, J., 2008. Average causal effects from nonrandomized studies: a practical guide and simulated example. Psychol. Methods 13 (4), 279–313.

Schwab, R.N., Walton, N.E., Mounce, J.M., Rosenbaum, M.J., 1982. Synthesis of Safety Research Related to Traffic Control and Roadway Elements-Volume 2. Chapter 12: Highway Lighting. Report No. FHWA-TS-82-233. Federal Highway Administration.

Tarko, A.P., Kanodia, M., 2004. Effective and Fair Identification of Hazardous Locations, Transportation Research Record, Journal of the Transportation Research Board. No. 1897, 64–70.

Tsyganov, A.R., Warrenchuk, N.M., Machemehl, R.B., 2009. Driver performance and safety effects of edge lines on rural two-lane highways, 88th Annual Meeting of Transportation Research Board, Washington, D.C., USA.

Yanovitzky, I., Zanutto, E., Hornik, R., 2005. Estimating causal effects of public health education campaigns using propensity score methodology. Eval. Program Plann. 28, 209–220.