Accident Analysis and Prevention 60 (2013) 148–157

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

The impacts of speed cameras on road accidents: An application of propensity score matching methods

Haojie Li, Daniel J. Graham*, Arnab Majumdar

Centre for Transport Studies, Dept of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

ARTICLE INFO

Article history:
Received 8 May 2013
Received in revised form 10 July 2013
Accepted 10 August 2013

Keywords:
Speed limit enforcement cameras
Causal analysis
Propensity score matching

ABSTRACT

This paper aims to evaluate the impacts of speed limit enforcement cameras on reducing road accidents in the UK by accounting for both confounding factors and the selection of proper reference groups. The propensity score matching (PSM) method is employed to do this. A naïve before and after approach and the empirical Bayes (EB) method are compared with the PSM method. A total of 771 sites and 4787 sites for the treatment and the potential reference groups respectively are observed for a period of 9 years in England. Both the PSM and the EB methods show similar results that there are significant reductions in the number of accidents of all severities at speed camera sites. It is suggested that the propensity score can be used as the criteria for selecting the reference group in before-after control studies. Speed cameras were found to be most effective in reducing accidents up to 200 meters from camera sites and no evidence of accident migration was found.

1. Introduction

The impacts of speed on accident severity are well known and speed limits in the UK define maximum desirable traffic speeds for the purposes of road safety in order to reduce the number of casualties. There are a number of policy measures that can be taken by governments in order to improve road safety by reducing traffic speed. An example of such a measure is that of speed limit enforcement cameras. These were first introduced in the UK in 1991 to persuade drivers to comply with these limits and their use has grown in recent years. Numerous studies have since shown that the introduction of speed cameras can reduce the vehicle speed as well as road accidents. The main challenge for their evaluation is the construction of the counterfactual outcomes, i.e. what would have happened to the "treated" units in the absence of any treatment. Since the counterfactual outcome cannot be observed, statistical methods are used for its estimation, in particular naïve Before-After and Empirical Bayes (EB) methods. Typically, a reference or control group is employed to estimate the counterfactual outcomes of the treatment group. However, due to the confounding factors, the characteristics of treated and untreated units may differ in the absence of any treatment. In other words, the characteristics of units that are treated differ in some systematic way from those that are not treated, and those characteristics also have a bearing on the incidence of selection bias (i.e. regression-to-mean (RTM)) and the severity of its impact. Only untreated units with similar characteristics to those treated can be used to approximate the counterfactual outcomes of the treatment group. A critical issue inadequately addressed in previous studies is the selection of this reference or control group.

This paper tackles this critical issue of the reference group by using the propensity score matching method (PSM) and subsequently uses this method to evaluate the effect of speed cameras on the reduction of road traffic accidents. The PSM method has become a popular approach to estimate causal treatment effects, but, to the best of the authors' knowledge, remains untried in road traffic safety research. What makes the PSM method attractive is that it gives a clear criterion by which to select the reference or control group.

And unlike the traditional matching methods, which implement matching on multiple dimensions, the PSM enables matching to be reduced to a single dimension, the propensity score. Therefore treated and untreated units with similar propensity scores can be compared to obtain the treatment effect. This method is then used to analyze a large dataset from speed camera sites in England to show its efficacy.

This paper is organized as follows. The literature review is presented in Section 2. The method and data used in this analysis are described in Sections 3 and 4. The results are presented and discussed in Sections 5 and 6. The conclusions are given in the final section.

2. Literature review

In the past decade, numerous studies have been conducted to investigate the impacts of speed enforcement cameras on

* Corresponding author. Tel.: +44 20 7594 6088; fax: +44 20 7594 6107.
E-mail address: d.j.graham@imperial.ac.uk (D.J. Graham).

0001-4575/$ – see front matter
http://dx.doi.org/10.1016/j.aap.2013.08.003

Table 1
A summary of studies on safety effects of speed cameras.

Authors	Data and units of analysis	Methods	Main findings
Goldenbeld and van Schagen (2005)	28 road sections; speed data of 5 years after and 1 year before the enforcement; accidents data of 5 years after and 8 years before the enforcement	Before and after study with comparison groups; long period data is used to control for RTM	A reduction of 21% in injury accidents and serious casualties; significant decrease in mean speed and speed limit violators
Hess and Polak (2003)	43 camera sites with 10 years accidents data in Cambridgeshire	Autoregressive integrated moving average (ARIMA)/seasonal ARIMA	A reduction of 31.26% in injury accidents
Newstead and Cameron (2003)	10 years crashes data for Queensland; segments of 2 km, 4 km and 6 km are treated as treatment groups separately	Quasi-experimental analysis with all areas outside camera sites as control groups	A reduction of 45% in fatal crashes and significant reduction in other kinds of crashes
Chen et al. (2002)	22 km corridor with 12 photo radar locations; data of 2 years before and 2 years after are used	Empirical Bayes with comparison groups	A 2.8 km/h reduction in mean speed; a decrease of 14% in collision at photo radar sites, 19% reduction in collision at non-photo radar sites and 16% total reduction along the corridor
Christie et al. (2003)	101 mobile speed camera sites in South Wales; 1996–2000 accidents data extracted from STATS 19	A circle zone based and a route based before and after analysis	The route based method is superior than the circle based one and suggested a 51% reduction in injury crashes.
ARRB Group Project Team (2005)	28 speed camera sites in New South Wales; speed and crashes data of 3 years before and 2 years after the operation of speed cameras	Quasi-experimental analysis	About 6 km/h fall in mean speed; reduction of 23% for casualty crashes, 20% for injury crashes and 90% for fatal crashes
Mountain et al. (2004)	62 fixed speed camera sites with 1 km upstream and downstream; crashes data of 3 years before and 3 years after the installation of cameras	Empirical Bayes with UK national accident totals as comparison groups	4.4 mph fall in mean speed; reduction of 26% and 34% for overall injury accidents and fatal and serious accidents separately
Cunningham et al. (2008)	14 corridors in Charlotte, USA; data of 4 years before and 1 year after the enforcement	Before and after study with comparison groups	Reduction of about 14% in collisions; around 7% decrease in mean speed
Mountain et al. (2005)	79 speed cameras with accidents of 3 years before and 3 years after the enforcement	Empirical Bayes is used to control for RTM	A reduction of 22% in personal injury accidents; engineering schemes incorporating vertical deflections offer the largest benefits
Shin et al. (2009)	A 6.5 mile urban freeway; data of 1.5 years	A before and after study with a comparison group; a before and after study with traffic flow correction; an empirical Bayes before and after study	A 9 mph fall in mean speed; a general reduction of 44–54% in total number of crashes
Keall et al. (2001)	Open roads with visible and hidden speed cameras; aggregated crash data for nearly 5 years	An interrupted time-series design with a comparison group	A reduction of 2.3 km/h in mean speed in speed camera areas; a net fall of 11% in the crash rate in the trail area
Gains et al. (2004, 2005)	24 areas with over 2300 speed camera sites in the UK; data of 3 years baseline and 3 years after program	A before and after study with a comparison group; an empirical Bayes before and after study	A fall of around 7% in mean speed at camera sites; a 33% reduction in personal injury collisions at speed camera sites
Jones et al. (2008)	29 camera locations in Norfolk, UK; monthly road traffic casualties data from 1999 to 2003	A before and after study with comparison groups; effects of RTM are estimated	A 1% decline in overall crashes and a 19% decline in crashes at speed camera sites

safety (Goldenbeld and van Schagen, 2005; Hess and Polak, 2003; Newstead and Cameron, 2003; Chen et al., 2002; Christie et al., 2003; ARRB Group Project Team, 2005; Mountain et al., 2004; Cunningham et al., 2008; Mountain et al., 2005; Shin et al., 2009; Keall et al., 2001; Gains et al., 2004, 2005; Jones et al., 2008). Table 1 summarizes these studies, which show that the implementation of speed cameras has significantly reduced the vehicle speed and the number of accidents near camera sites. In these studies, it is also suggested that the impact of speed cameras relates to various factors, such as types of road, the speed limit and site length, all of which will be adjusted for in this study. Furthermore we will tackle several outstanding issues which have yet to be fully addressed in the previous evaluations of the effects of speed cameras on road accidents.

Most studies to date have used naïve before-and-after methods with control groups (Goldenbeld and van Schagen, 2005; Christie et al., 2003; Cunningham et al., 2008; Gains et al., 2004; Jones et al., 2008). In these studies, a group of similar sites is usually selected as the control group in order to account for the general trend in accidents. However, this method is unable to control for effects of the RTM, sometimes called selection bias, which is a type of bias due to a flaw in the sample selection process. In the context of road safety, the RTM occurs when evaluating the effect of treatments that aim to make dangerous sites safer. Black spot sites with high recent crash record are often chosen and their accidents rate will tend to be lower in subsequent years. The impact of selection bias is that it can make random variation appear to as real change caused by treatments and therefore overestimate the effect of a safety treatment.

The empirical Bayes (EB) method has been suggested by Hauer et al. (2002) as effective in controlling for the RTM effect. In recent years, the EB approach has been applied in several evaluations of speed camera schemes (Elvik, 1997; Chen et al., 2002; Shin et al., 2009; Mountain et al., 2005). The EB method combines the observed crash number and the expected crash number estimated by prediction models and safety performance functions (SPF). Reference or comparison groups are usually employed to calibrate the SPF to account for the trend in accidents, as well as the effects of changes in flow, between the before and after periods. Ideally reference groups should have the same or similar traffic flow and road characteristics, i.e. the reference group must be representative of the treated sites. However, in previous research, not only is there insufficient justification of the selection of control groups, how the treatment and control groups are matched is also unclear. In fact, this issue of similarity is also critical when selecting the control group in the before and after method.

Rosenbaum and Rubin (1983) propose the propensity score matching method as a solution to this problem of similarity. The propensity score is the probability that a unit is selected into the treatment group conditional on observed covariates. “Similar” groups can then be defined clearly as those with similar propensity scores, and by this avoid selection bias to ensure that the difference between the treatment and control groups can be attributed to the treatment. The propensity score matching method has been widely studied and used in many evaluations of social, economic and medical programs (Heckman et al., 1997; Rudner and Peyton, 2006; Hirano and Imbens, 2001; Dehejia, 2005; Dehejia and Wahba, 2002; Kurth et al., 2006; Lechner, 2001; Abadie and Imbens, 2004, 2009).

The phenomenon of accident migration may occur and manifest itself in various ways. One way is seen on roads nearby to the speed camera sites if drivers should choose alternative routes to avoid these sites. In turn this may lead to an increase in accident numbers on alternative routes that avoid speed cameras whilst decreasing numbers at such sites. Mountain et al. (2005) show that speed enforcement cameras can affect route choice and have a significant effect on accidents at camera sites. The authors also suggest that the traffic flows before and after the installation of speed cameras should be monitored and accounted for in the model. Both Mountain et al.’s (2005) study and that of Christie et al. (2003) estimate the variation of speed camera effects at different distances from the cameras. There are two reasons for doing this. First, it is possible that drivers may decelerate and accelerate abruptly before and after the camera sites. This is known as the “kangaroo” effect and is another manifestation of accident migration. When this happens, it is necessary to know if there is unexpected increase in accidents upstream or downstream of the camera sites. Second, although Department for Transport (2004) provides site selection guidelines on length of camera sites, there is little knowledge of the most effective area of speed cameras. This paper will use a similar definition of section length to that of Mountain et al. (2005) and Christie et al. (2003).

3. Method

3.1. Propensity score matching

In the case of a randomized experiment, the treatment status T_i is unconditionally independent of the potential outcomes Y_i. For non-randomized observational data, such independence cannot be achieved due to the confounding factors $\boldsymbol{X}$, that is, covariates that affect both the probability of treatment exposure and potential outcomes. Consequently, simple comparison of mean outcomes between treated and untreated groups will not in general reveal the causal effect. However, conditional independence of potential outcomes and treatment status can be ensured by adjusting for the vector of covariates $\boldsymbol{X}$, then consistent causal estimates of treatment effects can be obtained.

A widely used method called the propensity score matching (PSM) is applied to evaluate the effect of fixed speed cameras on road accidents. The idea behind this method is to construct a control group that is similar to the treatment group in all relevant pre-treatment covariates $\boldsymbol{X}$. Instead of matching directly on all the covariates $\boldsymbol{X}$, PSM has the advantage of reducing the multiple dimension of matching to a single dimension, the propensity score, which is the probability of receiving a treatment. Conditional on the propensity score, differences in observed outcomes between the two groups can be solely attributed to the intervention impacts. In other words, adjusting for the propensity score is enough to eliminate the bias created by all confounding factors.

3.1.1. Notation

The treatment indicator is defined as T_i, where $T_i = 1$ if unit i receives the treatment and $T_i = 0$ otherwise. Let $Y_i(T)$ denote the potential outcome for unit i, where $i = 1, \ldots, N$ and N denote the total population. The treatment effect for unit i can be described as:

$$\delta_i = Y_i(1) - Y_i(0).$$

In practice, the parameter of interest is usually the average treatment effect on the treated (ATT), which can be defined as:

$$\delta_{\text{ATT}} = E(\delta|T = 1) = E(y(1)|T = 1) - E(Y(0)|T = 1)$$

3.1.2. PSM assumptions

Two crucial assumptions underlying the PSM method are introduced by Rosenbaum and Rubin (1983).

Assumption 1 *(Conditional Independence Assumption (CIA):).*

$$(Y(1),\ Y(0)) \perp T|X$$

This assumption is also known as the unconfoundedness condition and states that the potential outcomes are independent of the treatment status after controlling for covariates $\boldsymbol{X}$. In other words, the treatment assignment can be considered as a random assignment conditional on $\boldsymbol{X}$. The CIA ensures that differences between treated and untreated units can be accounted for and the untreated units can be used to estimate a counterfactual outcome for the treatment group. With a high dimensional vector $\boldsymbol{X}$, the matching can be complicated. To deal with this problem, Rosenbaum and Rubin (1983) propose the propensity score, $P(T = 1|\boldsymbol{X}) = P(\boldsymbol{X})$, which is the probability of being in a treatment conditional on covariates $\boldsymbol{X}$. They show that potential outcomes are also independent of treatment conditional on the propensity score. The CIA based on the propensity score can be described as:

$$(Y(1),\ Y(0)) \perp T|P(X).$$

Assumption 2 *(Common Support Condition (CSC):).*

$$0 < P(T = 1|X) < 1$$

This assumption ensures that units with the same $\boldsymbol{X}$ values have a positive probability of being both treated and untreated. The CSC is also known as the overlap condition, because there is sufficient overlap in the $\boldsymbol{X}$ of the treated and untreated units to find adequate matches.

3.1.3. Implementing PSM

The procedure for estimation of treatment effects using PSM can be illustrated in three steps:

(1) *Estimating the propensity score*. The first step when implementing PSM is to estimate the propensity score. For a binary treatment variable, logit and probit models are usually preferred to a linear probability model, which may generate predictions outside the [0,1] bounds of probabilities. Because logit and probit models usually produce similar results, the choice is not critical. In this paper, a logit model is used described as follows:

$$P(T = 1|X) = \frac{EXP(\alpha + \beta' X)}{1 + EXP(\alpha + \beta' X)}$$

where α is the intercept and β' is the vector of regression coefficients.

One key issue when specifying the propensity score model is the inclusion of covariates. To satisfy the unconfoundedness condition, variables that influence both the selection of treatment groups and potential outcomes should be included in the model. If there are explicit criteria used in the treatment group selection, such criteria must be included in the propensity score model. Other factors can be decided based on a sound knowledge of previous studies. It is also suggested that over-parameterized models should be avoided (Bryson et al., 2002; Augurzky and Schmidt, 2000). There are two reasons for this. First, including extraneous covariates may violate the overlap condition. Second, the inclusion of non-significant covariates can increase the variance.

In conclusion, following covariates should be included in the model based on these criteria:

(a) Covariates that strongly influence the selection into the treatment and outcomes;
(b) Covariates which are statistically significant in the regression model;
(c) Covariates that have been suggested as important factors affecting outcomes in previous research.

(2) Choosing matching algorithm. After estimating the propensity score, a matching algorithm is selected to construct the control group from non-treated units. There are four mostly used matching algorithms: nearest neighbor matching, caliper and radius matching, stratification and interval matching, kernel and local linear matching. For detailed discussion of these matching algorithms, please refer to the work by Heinrich et al. (2010).

There is no theoretical guidance on how to select the most appropriate matching algorithm. Given a large sample, the result from all algorithms should be similar and therefore the choice is not critical. However, the matching approach can have a considerable impact on the results when only a small sample is available. In this case, the decision is made based on the distribution of the estimated propensity score for treated and untreated groups. For example, if some treated units have many close neighbors while others do not, the kernel and local linear matching would be more appropriate. In general, the reasonable way is to try every matching approach. If there is a large difference in the results, further judgment and consideration is required.

(3) Estimating Treatment Effects. Once treated units have found matches from the untreated group, the treatment effect can be evaluated by taking differences in outcomes between treated units and their matches. A number of programs are available for STATA and other statistical software programs. Bootstrapping methods are usually applied to calculate standard errors. The program used in this study is **psmatch2** in STATA, which is developed by Becker and Ichino (2002).

(4) Difference-In-Difference (DID) Matching. The conditional independence assumption is too strong and may not hold when unobserved factors that may influence outcomes are not included in the model. However, the CIA can be relaxed by using the DID matching estimator (Heckman et al., 1997). Given data from the pre-treatment period, any time-invariant confounder can be controlled for. In the DID matching approach, the dependent variable is the difference between pre-intervention and post-intervention periods. The treatment effect can be simply calculated by applying the procedures described above.

3.1.4. Issues in PSM estimation

Two important issues in PSM estimation are discussed in this section. First, the validity of the overlap condition needs to be checked. It is assumed that there is a positive probability of receiving the treatment for both the treated and untreated units conditional on covariates $\boldsymbol{X}$. The most straightforward way to evaluate this condition is by the visual inspection of the density distribution of the propensity score for both the treatment and control groups. The histograms of the distribution of propensity scores for both groups, together with a comparison of the minima and maxima values in each group, can help to provide clear knowledge of the extent to which there is overlap in the propensity score between the two groups.

The second issue regards the assessment of the matching quality. It should be noted that the real purpose of matching is to balance the characteristics between the treatment and control groups. Rosenbaum and Rubin (1983) propose a theorem stating that after conditioning on $P(T=1|\boldsymbol{X})$, covariates $\boldsymbol{X}$ should be independent of the treatment decision.

$$X \perp T | P(T=1|X)$$

A balancing test is proposed by Rosenbaum and Rubin (1983) and applied by Dehejia and Wahba (2002).The units are first divided into blocks with similar propensity scores. Within each block, the t-test is used to test whether the distribution of covariates $\boldsymbol{X}$ is the same in both groups. If a block has unbalanced covariates, then that blocked is divided into smaller blocks and the evaluation is repeated. If differences still remain, then the specification for propensity score is revisited and higher order or interaction terms are included. Then the process is repeated again from the beginning. Such a stratification test can be done using the program **psmatch2** in STATA.

It is worth noting that the violation of the assumptions underpinning the PSM can bias the estimates. In some cases, when we have little theoretical or empirical evidence on the nature of selection into a treatment, it is difficult to know which factors influence participation and outcomes, and thus what set of covariates to be included in the propensity score model. Since estimates of treatment effects can be sensitive to the covariates used, this may bias estimates. This may happen with innovative treatments which we have little knowledge about. In this instance, pre-intervention research would be helpful to identify the covariates involved. It is also possible that the selection into treatments is driven by factors that are not observable, and the matching estimator may be seriously biased. As discussed earlier, with additional pre-intervention data, a modified version, the DID matching method can be applied to correct for some of this bias, as long as the effects of unobserved factors is fixed over time.

In summary, the procedures for using the PSM to evaluate safety effects of speed cameras can be illustrated as following steps.

(1) The data for both camera sites and comparison sites, such as accidents record and site information, is collected and constructed in a single data set.
(2) Covariates are selected to be included in the logit model to estimate the propensity score, which is the probability of being treated as camera sites for both groups.
(3) The distributions of propensity scores are compared between camera sites and comparison sites to check the overlap condition. If the condition is not satisfied, then covariates included in the logit model need to be re-selected.
(4) It is recommended that multiple matching algorithms are applied to increase the credibility of the PSM.
(5) A balancing test is conducted to test whether the camera sites and comparison sites are statistically similar after matching. If significant differences are found, the logit model is re-specified and the process is repeated from the beginning.
(6) The safety effects of speed cameras can be evaluated by taking differences in outcomes between matched camera sites and

comparison sites. In the original PSM, outcomes are observed accidents number in the post-intervention period, while the outcome variables are the difference between pre-intervention and post-intervention in the DID matching.

Fig. 1 shows the diagram of applying the PSM to the estimation of safety effects of speed cameras.

3.1.5. A numerical example

To illustrate the basic mechanics of the PSM, a simple example is shown in Table 2. The matching algorithm used is nearest neighbor matching, where the treatment sites are matched with the comparison sites which have the most similar propensity score.

The average effect on treated sites is estimated by averaging the differences over all the treated sites. In this case, the estimated treatment effect is (2 + 4.5 + 2 + 2 + 3)/5 = 2.7.

This example is very simple and only nearest neighbor matching is applied. The problem could be complex with a large sample size and high dimensionality of covariates. Then other algorithms should be used.The Bayesian Approaches Using Propensity Scores

Empirical Bayes (EB) methods have been introduced and widely used in before-and-after traffic safety countermeasures evaluation (Hauer, 1997; Hauer et al., 2002; Persaud et al., 2009; Persaud and Lyon, 2007; Sayed et al., 2004; Hirst et al., 2004).Recent applications in analysis of speed camera effects include studies by Mountain et al. (2005), Shin et al. (2009), and Gains et al. (2005). In the EB approach, the predicted number of crashes without treatment is derived by combining the observed crash counts in before period and expected number of crashes from Safety Performance Functions (SPFs).

The validity of the EB approach heavily relies on the availability of a proper reference group and an inappropriate reference group can bias the estimation of SPFs. The propensity score, however, can be used to find untreated sites that are similar to treated sites and thereby enable the reference group to be constructed. In this study, two reference groups are applied in the EB approach. One reference group is selected based on the propensity score, while the other one contains all the potential reference sites. Comparisons are conducted between results from the EB and PSM models.

The SPF used in this study is based on the model proposed by Mountain et al. (1997), which has been used in other studies (Gains et al., 2005; Mountain et al., 2004, 2005).

The number of observed crashes y can be modeled as:

$$y \sim Poisson(\mu\varepsilon)$$

$$\log\mu = \alpha + \log L + \beta_1 \log V + \beta_2 \left(\frac{J}{L}\right) + \varepsilon$$

where L is the site length, V is the AADF at each section and J is the number of minor junctions within site length. ε is a Gamma distributed random error term. The parameters of this model depend on the accident type, speed limit, the carriageway type, and road class.

The EB estimate of total number of crashes $\widehat{\mu_B}$ in a before period of t_B years is

$$\widehat{\mu_B} = t_B\hat{\mu}$$

Then the predicted number of crashes in a before period, $\widehat{M_B}$ can be obtained by$\widehat{M_B} = \rho\widehat{\mu_B} + (1-\rho)X_B$where X_B is the observed number of crashes in the before period and$\rho = \left(1 + \frac{\widehat{\mu_B}}{\phi}\right)^{-1}$ where φ is the shape parameter for the NB distribution.

To account for the trend in accidents between the before and after periods, the expected accidents in the after periods are calibrated using a reference group. The estimate of accidents number in the after period had the treatment not occur, $\widehat{M_A}$, can be calculated after adjusting the time trend effect:

$$\widehat{M_A} = \left(\frac{N_{A\_POP}}{N_{B\_POP}}\right)\widehat{M_B}$$

where N_{A_POP} and N_{B_POP} are the numbers of crashes for total population in the before and after periods.

To control for the effect of any flow changes due to the treatment, the expected flow in the after period had the treatment not occurred, V'_A can be estimated as$V'_A = \left(\frac{V_{A_POP}}{V_{B_POP)}}\right)V_B$where $V_{A\text{-}POP}$ and $V_{B\text{-}POP}$ are the traffic flow for whole population in the before and after periods, V_B is the observed traffic flow in the before period.

The estimate of crashes number in the after period can be refined as

$$\widehat{M'_A} = \left(\frac{V_A}{V'_A}\right)^{\beta} 1\widehat{M_A}$$

where V_A is the observed traffic flow.

The treatment effect can be obtained as$\delta_{AAT} = \frac{\frac{X_A}{t_A} - \frac{\widehat{M'_A}}{t_A}}{X_B/T_B}$where X_A is the observed number of crashes in the after period.

4. Data

4.1. Covariates

Clear knowledge of the criteria for treatment assignment can help decide which covariates to be included in the propensity score model and hence improve the validity of the PSM method. Currently in the UK, formal site selection guidelines for fixed speed camera sites exist (Gains et al., 2004) as shown below.

(1) Site length: between 400–1500 m.
(2) Number of fatal and serious collisions (FSCs): at least 4 FSCs per km in last three calendar years.
(3) Number of personal injury collisions (PICs): at least 8 PICs per km in last three calendar years.
(4) 85th percentile speed at collision hot spots: 85th percentile speed at least 10% above speed limit.
(5) Percentage over the speed limit: at least 20% of drivers are exceeding the speed limit.

The first three guidelines can be thought of as primary criteria and the latter two as secondary criteria. There are though sites unable to meet the above criteria that may be still selected as enforcement sites for reasons such as community concern and engineering factors.

Selection of the speed camera sites was primarily based on accident history. Accident data can be obtained from the STATS 19 database and located on the map using MapInfo. However, secondary criteria such as the 85th percentile speed and percentages of vehicles over the speed limit are normally unavailable for all sites on UK roads. If speed distributions differ between the treated and untreated groups, then the failure to include the speed data could bias the estimation, an issue discussed in previous research (Mountain et al., 2005; Gains et al., 2004). For untreated sites with the speed limit of 30 mph and 40 mph, the national average mean speed and percentages of speeding are similar to the data for the camera sites. The focus groups for this study are sites with the speed limit of 30 mph and 40 mph throughout the UK. It is reasonable to assume that there is no significant difference in the speed distribution between the treated and untreated groups and hence exclusion of the speed data will not affect the accuracy of the propensity score model.

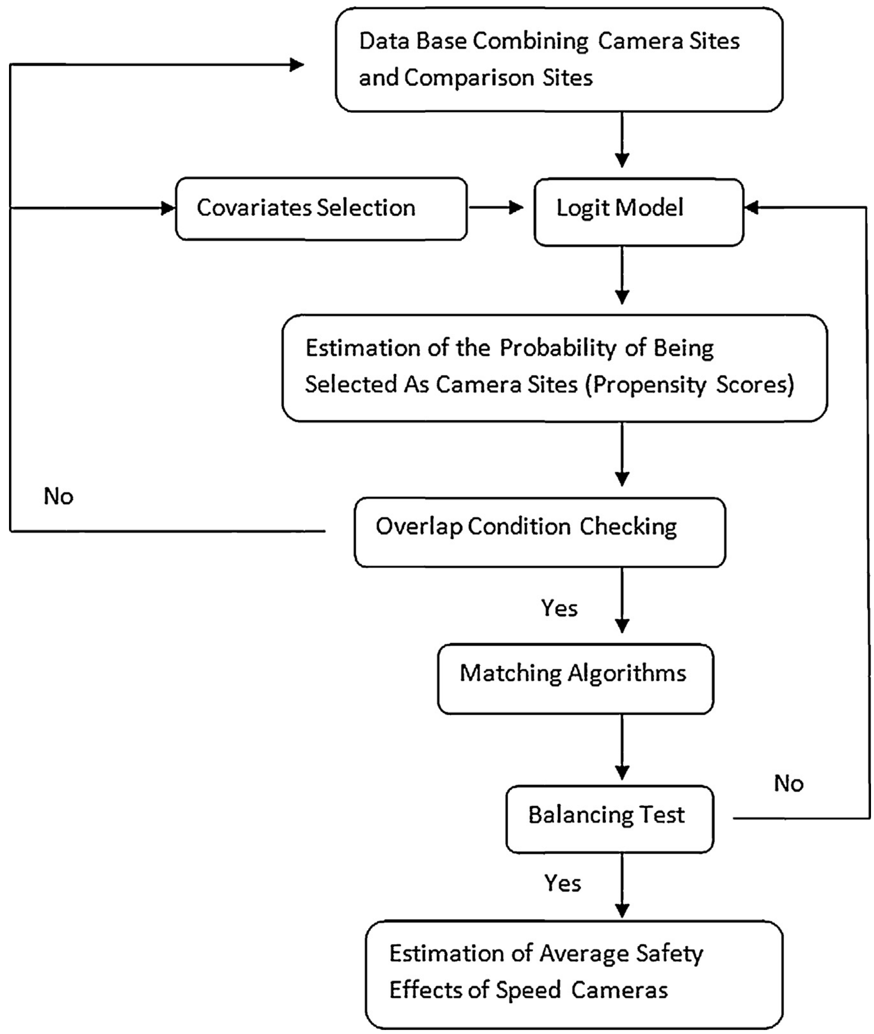

Fig. 1. The diagram of the application of the PSM to the evaluation of safety effects of speed cameras.

In order to account for the "kangaroo" effect (Elvik, 1997; Thomas et al., 2008), in this paper, the effective length of camera sites is determined by investigating a different range of distance to camera sites, including 200 m, 500 m and 1 km on both sides of the camera. However, if a route meets a major junction or a route for another camera site, the route is terminated (Christie et al., 2003). Therefore some sites cannot be evaluated for all distance bands. The percentages of such sites are 0% for 200 m sites, 3% for 500 m sites and 9% for 1 km sites respectively.

It is also possible that drivers may choose alternative routes to avoid speed cameras sites. Accident reduction at camera sites may include the effect induced by a reduced traffic flow. The benefits of speed cameras will therefore be overestimated without controlling for the change in traffic flow. The annual average daily flow (AADF) is available for both treated and untreated roads and the effect due to traffic flow is controlled for in this study by including the AADF in the propensity score model.

In addition to the criteria that strongly influence the treatment assignment, factors that affect the outcomes should also be taken into account when the propensity score model is specified. We further include road characteristics such as: road types, speed limit, and the number of minor junctions within site length, which are

Table 2
An example of applying the PSM using nearest neighbor matching.

I	T	Propensity score	Accident number in the post-intervention period	Matching comparison sites	$Y(1)$	$Y(0)$	$Y(1) - Y(0)$
1	0	0.2	0	–	–	–	–
2	0	0.3	2	–	–	–	–
3	0	0.4	5	–	–	–	–
4	0	0.6	6	–	–	–	–
5	0	0.9	11	–	–	–	–
6	1	0.6	8	[4]	8	6	2
7	1	0.5	10	[3,4]	10	5.5	4.5
8	1	0.3	4	[2]	4	2	2
9	1	0.1	2	[1]	2	0	2
10	1	0.2	3	[1]	3	0	3

suggested as important factors when estimating the safety impact of speed cameras (Gains et al., 2005; Christie et al., 2003; Mountain et al., 1997).

The final propensity score model can be described as below:

$$P(T=1|X)=\frac{EXP(\alpha+\beta_1 C_{FSC}+\beta_2 C_{PIC}+\beta_3 L+\beta_4 V+\beta_5 A+\beta_6 B+\beta_7 M+\beta_8 S_{30}+\beta_9 S_{40}+\beta_{10} J)}{1+EXP(\alpha+\beta_1 C_{FSC}+\beta_2 C_{PIC}+\beta_3 L+\beta_4 V+\beta_5 A+\beta_6 B+\beta_7 M+\beta_8 S_{30}+\beta_9 S_{40}+\beta_{10} J)}$$

where C_{FSC} and C_{PIC} are the FSCs and PICs in the last three years before the camera installation. The road type is defined by binary indicators, A (A road), B (B road) and M (Minor road). S_{30} (S_{40}) equals 1 if the speed limit is 30mph (40mph) and 0 otherwise. Other covariates are as defined above.

4.2. Sample size

The PSM method is known as a "data-hungry method" in terms of the number of treated and untreated units. Matching can be only implemented when there is sufficient overlap between both treatment and control groups for every propensity score block. If no match can be found for treated units at some propensity score, these treated units will be discarded and the estimation of the ATT will be biased. So a large untreated pool is required to ensure adequate matches. However, the literature is not explicit on how large the untreated group should be. According to previous research, the ratio of the number of control group candidates to the number of treatment group members ranges from 1.5:1 to over 30:1 (Hirano and Imbens, 2001; Dehejia and Wahba, 2002; Kurth et al., 2006; Smith and Todd, 2005; Peikes et al., 2008). The ratio chosen in this study is around 7:1, which is assumed sufficient to ensure the matching quality. Due to the data restriction, 771 camera sites from eight following English administrative districts are included in the treatment group: Cheshire, Dorset, Greater Manchester, Lancashire, Leicester, Merseyside, Sussex and West Midlands. A total of 4787 potential control sites were selected randomly within these districts. Because the established dates of speed cameras vary from 2002 to 2004, the research period is chosen from 1999 to 2007 to ensure the availability of accident data for 3 years before and after the camera installation for every camera site. For untreated sites, the before and after periods are defined the same as those of the most proximate camera sites. Whilst concerns have been raised about the completeness and reliability of accident data in STATS19, in the case of accidents at speed camera sites, given their nature of such sites, it is likely that all accidents at were captured and the data is reliable and complete.

It is worth noting that there is a difference in the data requirement for the EB and PSM methods. Although both methods rely on a large sample of untreated units, the EB method further requires that the untreated group must be representative of the treated sites in order to estimate the SPFs, which constrains the application of the EB method.

5. Results

5.1. The estimation of propensity scores

The first step in the propensity score matching method is to estimate the probability of being selected in the treatment group. The logit model is regressed on the covariates and the covariates that influence the participation and the outcome should be included in the model. Table 3 shows that all covariates except minor road are significant in the estimation of the propensity score. This is probably because there are very few speed cameras installed on minor road in the study sample. The result confirms that the covariates included in the propensity score model are important in predicting the possibility of being selected as camera sites.

5.2. Tests of matching quality

Before estimating the effects of speed cameras, the validity of the PSM method must be checked. One approach is through a visual inspection of the propensity score distribution for both the treatment and comparison groups. From the histograms of propensity scores for both groups, the extent to which there is overlap in the scores between the treatment and comparison groups is apparent. Observations that fall outside the region of common support must be discarded and cannot be estimated. The estimation will be unaffected if the proportion of discarded observations is small (Bryson et al., 2002). However, if the proportion is too large, the true treatment effect can be misestimated. Fig. 2 shows the distribution of propensity scores for both groups. We observe 771 sites and 4787 sites for the treatment and the potential comparison groups respectively, with only seven treated sites are outside the region of common support and discarded. Therefore there is sufficient overlapping of the distributions.

The next step is to perform the balancing tests to assess the matching quality as these tests can verify that treatment is independent of the covariates after matching. The PSM method aims to balance characteristics between the treatment and comparison groups, i.e., there should be no significant differences between covariate means of the treatment and comparison groups after matching. Table 4 shows the t-test of differences in covariate means before and after the matching. It can be seen that there are significant differences in all covariates except site length when using all sites as the comparison group. It is clear that the characteristics between groups are imbalanced and the estimation of the treatment effect can be biased. The PSM method is subsequently used to construct matched comparison groups. Table 4 shows that all covariates are balanced between the treatment and matched comparison groups. Consequently the bias due to the differences in observable characteristics is reduced.

5.3. Effects of speed cameras on road accidents

The effects of speed cameras on road accidents are estimated using three different methods: a naïve before and after approach, the PSM method and the EB method. Since different algorithms can be chosen when employing the PSM method, the robustness of the results must be checked to ensure that the

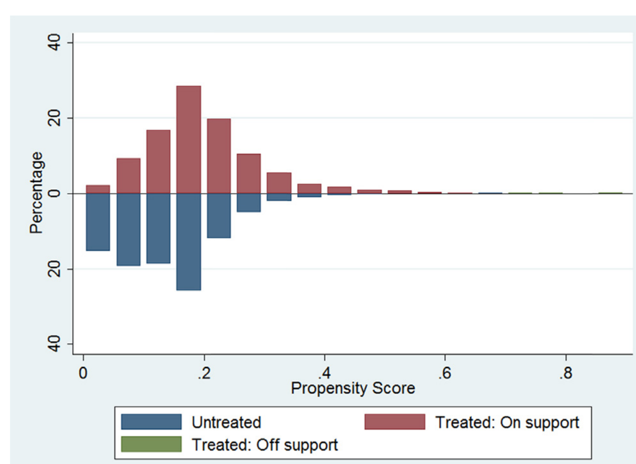

Fig. 2. Propensity score distribution.

Table 3
The propensity score model.

	Coef. (Std. Err.)		z	P > z	[95% Conf. Interval]	
Number of minor junctions	0.023	(0.007)	3.33	0.001	0.009	0.036
AADF in baseline years	1.30E−05	(2.36E−06)	5.52	0.000	8.40E−06	1.76E−05
PICs in baseline years	−0.013	(0.003)	−4.13	0.000	−0.019	−0.007
FSCs in baseline years	0.159	(0.018)	8.67	0.000	0.123	0.194
Site length	−0.141	(0.064)	−2.20	0.028	−0.267	−0.015
A road	−0.377	(0.128)	−2.95	0.003	−0.627	−0.126
B road	−0.307	(0.135)	−2.27	0.023	−0.572	−0.042
Minor road	−0.078	(0.193)	−0.40	0.686	−0.457	0.301
Speed limit 30 mph	1.017	(0.101)	10.11	0.000	0.820	1.214
Speed limit 40 mph	0.594	(0.106)	5.61	0.000	0.387	0.802
Constant	−1.876	(0.168)	−11.14	0.000	−2.206	−1.546
Observations	5558					

estimation does not depend upon the chosen algorithm. In this study, results from five algorithms, two of which are of one type (K-nearest neighbors), are compared to increase our confidence in the PSM method. The matching algorithms used are: K-nearest neighbors matching ($K = 1$), K-nearest neighbors matching ($K = 5$), radius matching (caliper = 0.05), stratification matching and kernel matching (caliper = 0.05).The EB method is used and compared with the PSM method with two reference groups used in this method. One reference group is selected based on the propensity score, while the other one contains all the potential reference sites. For FSCs, only the effect on absolute accident numbers is estimated. Because the effect on annual FSCs per km in percentage is estimated as the ratio of changes in absolute accident number to the accident number in the pre-treatment period, and the sample for FSCs in the pre-treatment period is zero-inflated, large numbers of data will be discarded, thereby making the result unreliable.

Table 5 presents the estimations of effects of speed cameras on annual PICs and FSCs per km. The observed reduction in annual PICs per km is 1.441 in absolute numbers and 30.7% in percentage. When applying the PSM method, the results are very similar for all five algorithms, where the average reduction in PICs is around 1.068 (25.9% in percentage). Such similar results indicate that the estimations are independent of the algorithms used and increase our confidence in the PSM method. The results from the EB using all sites as the reference group show a reduction of 0.854 in absolute numbers and 23.3% in percentage, slightly lower than the results from the PSM method. Then matched sites are used as the reference group and the estimated reduction in PICs is 1.026 (25.7% in percentage). In Table 5 the effect on FSCs are also analyzed using three methods. Unsurprisingly, the result from the naïve before-and-after approach shows the largest fall of 0.342 in absolute number. The PSM methods give a consistent estimation for all five algorithms with the average reduction of 0.132. A similar result is obtained from the EB method using matched sites as the reference group, where the reduction is 0.135.

Table 6 summarizes the effects of speed cameras on PICs and FSCs given different distances from camera sites. These results suggest that speed cameras are most effective up to 200 meters, where the reduction in annual PICs per km is approximately 1.350 (27.5% in percentage).The effectiveness decreases as the distance from the camera site increases, with the estimations approximately 1.135 (26.4% in percentage) for up to 500 meters and 0.570 (18.5% in percentage) for up to 1 km respectively. In terms of effects on FSCs at different distance from cameras, certain figures in Table 6 are insignificant. This is probably due to too few FSCs being observed to give conclusive estimates. Nevertheless, it is obvious that the reduction of FSCs within 200 meters from the camera, 0.188, is the largest. Up to 500 meters, this reduction is 0.164, whilst for up to

Table 4
Checking the covariates balance between groups before and after using nearest neighbors ($K = 5$) matching.

Variable	Sample	Mean			%reduced	t-Test	
		Treated	Control	%bias	|bias|	t	p > |t|
Number of minor junctions	Unmatched	5.4578	3.5233	35.6	97.6	10.84	0.000
	Matched	5.3307	5.2852	0.8		0.16	0.873
AADF in baseline years	Unmatched	19,039	18,020	10.1	88.3	2.52	0.012
	Matched	19,049	19,168	−1.2		−0.22	0.823
PICs in baseline years	Unmatched	12.722	8.3347	34.7	99.2	9.60	0.000
	Matched	12.510	12.474	0.3		0.05	0.959
FSCs in baseline years	Unmatched	1.8431	1.0391	41.7	97.6	12.63	0.000
	Matched	1.7969	1.7773	1.0		0.18	0.861
Site length	Unmatched	0.7118	0.7009	2.3	−137.2	0.59	0.554
	Matched	0.7094	0.7353	−5.4		−1.05	0.294
A road	Unmatched	0.7276	0.7984	−16.7	74.2	−4.47	0.000
	Matched	0.7279	0.7096	4.3		0.79	0.427
B road	Unmatched	0.2101	0.1613	12.6	89.3	3.37	0.001
	Matched	0.2096	0.2148	−1.3		−0.25	0.803
Minor road	Unmatched	0.0376	0.0230	8.5	82.2	2.42	0.016
	Matched	0.0378	0.0404	−1.5		−0.26	0.792
Speed limit 30 mph	Unmatched	0.7575	0.5118	52.7	97.9	12.90	0.000
	Matched	0.7565	0.7513	1.1		0.24	0.813
Speed limit 40 mph	Unmatched	0.1219	0.1828	−17.0	97.9	−4.14	0.000
	Matched	0.1224	0.1237	−0.4		−0.08	0.938

Table 5
Effects of speed cameras on annual PICs/FSCs per km.

	Changes in annual PICs per km in absolute numbers					Changes in annual PICs per km in percentage					Changes in annual FSCs per km in absolute numbers				
	Treatment group	Control group	Changes	S.E.	*T*-Stat	Treatment group	Control group	Percentage changes	S.E.	*T*-Stat	Treatment group	Control group	Changes	S.E.	*T*-Stat
Unmatched	771	4787	−1.441	0.131	−11.02	726	4077	−30.70%	0.041	−7.5	771	4787	−0.342	0.037	−9.25
DID propensity score matching															
K-nearest neighbors matching ($K=1$)	764	663	−1.035	0.21	−4.92	726	600	−29.70%	0.051	−5.83	771	663	−0.141	0.06	−2.34
K-nearest neighbors matching ($K=5$)	764	2923	−1.068	0.168	−6.33	726	2625	−24.60%	0.034	−7.21	771	1676	−0.124	0.049	−2.5
Radius matching (caliper = 0.05)	769	4626	−1.081	0.155	−6.97	724	3921	−25.20%	0.031	−7.99	769	4626	−0.131	0.046	−2.82
Stratification matching	769	4628	−1.042	0.15	−6.96	725	4078	−24.70%	0.029	−8.48	769	4628	−0.135	0.044	−3.05
Kernel matching (bandwidth = 0.05)	771	4626	−1.117	0.147	−7.61	726	4077	−25.10%	0.032	−7.89	771	4626	−0.131	0.046	−3.01
Average effect			−1.068					−25.90%					−0.132		
Empirical Bayes using all sites as reference group	771	4787	−0.854	0.102	−8.34	726	4077	−23.30%	0.037	−7.01	771	4787	−0.197	0.044	−4.42
Empirical Bayes using matched sites as reference group	764	2923	−1.026	0.127	−8.04	726	2625	−25.70%	0.036	−6.36	771	1676	−0.135	0.069	−1.98

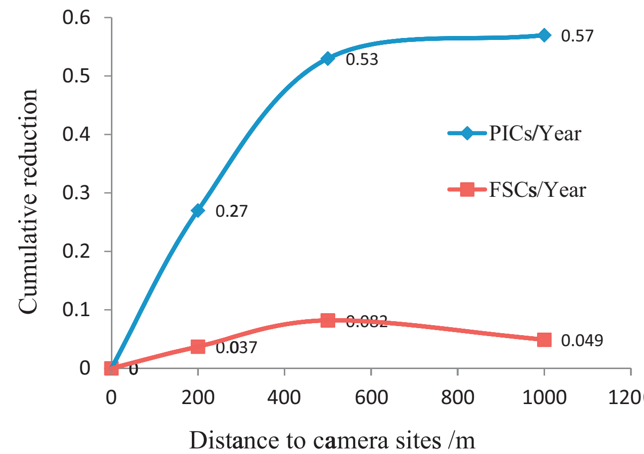

Fig. 3. Cumulative reductions in annual tables PICs/FSCs.

1 km the average reduction is 0.049, although the estimations using all algorithms are insignificant.

6. Discussion and conclusion

The issues of selecting control groups to account for confounding factors and how the treatment and control groups are matched are critical in assessing the impacts of road safety measures. This can be seen particularly when assessing the effect on road traffic accidents due to the introduction of speed cameras in the UK. This paper introduced the PSM method to account for these two issues and then applied it to data from the UK to assess the impact of speed cameras. Similar estimation results indicate that the PSM method and the EB method are comparable. The EB method requests a reference group that should be similar to the treatment group, though, it is not explicit how this reference group is selected in previous studies. This paper also shows that the characteristics of the treatment and comparison groups are well balanced after matching. The results confirm that the EB method using matched sites as the reference group is superior to the one using all sites. Therefore, the authors suggest that propensity scores can be applied as the criterion when constructing the reference group. Indeed the construction of such a reference group can be used in any road traffic safety analysis where a safety measure has been implemented, not simply for assessing the impacts of speed cameras.

This paper also has two major findings on the impacts of speed cameras on accidents. The first relates to the distance at which speed cameras have their greatest impact. For both PICs and FSCs, there is a reduction in accidents with a decrease in the distance from cameras and speed cameras are found to be most effective up to 200 m from camera sites. The reduction in accidents for up to 500 m is also prominent. Fig. 3 presents the cumulative reductions in annual PICs and FSCs. The cumulative reduction increases dramatically from 0 to 500 m and this tendency reduces from 500 m to 1 km. It is obvious that the reduction in accidents due to the effect of speed cameras is negatively correlated to the distance from the camera sites. However, it is unclear whether this relationship holds over larger distances (i.e. over 1000 m) because data restrictions prevented reliable estimation. The suggested site length by DfT (2004) is between 400 m to 1.5 km, which tallies with the effective length estimated in this paper.

The second finding relates to accident migration. Having controlled for accident migration due to the choice of alternative routes to avoid speed cameras by including the covariate AADF, this paper finds there is no evidence of the “kangaroo” effect, i.e. no increase in accidents upstream and downstream camera sites. This is an

Table 6
Effects of speed cameras on annual PICs/FSCs per km at different distance from cameras.

	Changes in annual PICS per km			Percentage changes in annual PICS per km			Changes in annual FSCS per km		
	0–200 m	0–500 m	0–1 km	0–200 m	0–500 m	0–1 km	0–200 m	0–500 m	0–1 km
K-nearest neighbors matching (K = 5)	−1.372**	−1.103*	−0.598**	−27.3%**	−25.2%*	−16.6% *	−0.115	−0.149***	−0.047
Radius matching (Caliper = 0.05)	−1.387*	−1.148*	−0.720*	−29.2%*	−26.3%*	−18.7% *	−0.209***	−0.169**	−0.049
Stratification Matching	−1.324*	−1.149*	−0.467**	−25.7%*	−26.2%*	−19.1% *	−0.167***	−0.180*	−0.053
Kernel matching (bandwidth = 0.05)	−1.318*	−1.141*	−0.496**	−27.7%*	−27.9% *	−19.4% *	−0.180	−0.161**	−0.045
Average effect	−1.35	−1.135	−0.421	−27.50%	−26.40%	−18.50%	−0.188	−0.164	−0.048

Notes: Figures are significant at: *99%, **95%, ***90.

important finding in that it shows the drivers do not alter their behavior to deliberately decelerate and accelerate abruptly before and after the camera sites. Rather speed cameras have a constant effect on driver behavior in reducing their speed.

References

Abadie, A., Imbens, G., 2004. Large Sample Properties of Matching Estimators for Average Treatment Effects. Working Paper. Harvard University.

Abadie, A., Imbens, G., 2009. Matching on the Estimated Propensity Score. Working Paper. Harvard University.

ARRB Group Project Team, 2005. Evaluation of the Fixed Digital Speed Camera Program in NSW.

Becker, S.O., Ichino, A., 2002. Estimation of Average Treatment Effects Based on Propensity Scores. The Stata Journal 2 (4), 358–377.

Bryson, A., Dorsett, R., Purdon, S., 2002. The Use of Propensity Score Matching in the Evaluation of Labour Market Policies. Working Paper No. 4. Department for Work and Pensions.

Chen, G., Meckle, W., Wilson, J., 2002. Speed and safety effect of photo radar enforcement on a highway corridor in British Columbia. Accident Analysis and Prevention 34, 129–138.

Christie, S.M., Lyons, R.A., Dunstan, F.D., Jones, S.J., 2003. Are mobile speed cameras effective? A controlled before and after study. Injury Prevention 9, 302–306.

Cunningham, C.M., Hummer, J.E., Moon, J., 2008. Analysis of automated speed enforcement cameras in Charlotte, North Carolina. Transportation Research Record 2078, 127–134.

Dehejia, R.H., Wahba, S., 2002. Propensity score-matching methods for nonexperimental causal studies. The Review of Economics and Statistics 84 (1), 151–161.

Dehejia, R., 2005. Practical propensity score matching: a reply to Smith and Todd. Journal of Econometrics 125, 355–364.

Department for Transport, 2004. Handbook of Rules and Guidance for the National Safety Camera Programme for England and Wales for 2005/06.

Elvik, R., 1997. Effects on accidents of automatic speed enforcement in Norway. Transportation Research Record 1595, 14–19.

Gains, A., Heydecker, B., Shrewsbury, J., Robertson, S., 2004. The National Safety Camera Programme 3-Year Evaluation Report. PA Consulting Group and UCL for Department for Transport, London.

Gains, A., Heydecker, B., Shrewsbury, J., Robertson, S., 2005. The National Safety Camera Programme 4-Year Evaluation Report. PA Consulting Group and UCL for Department for Transport, London.

Goldenbeld, C., van Schagen, I., 2005. The effects of speed enforcement with mobile radar on speed and accidents. An evaluation study on rural roads in the Dutch province Friesland. Accident Analysis and Prevention 37, 1135–1144.

Hauer, E., 1997. Observational Before–After Studies in Road Safety. Pergamon Press, Oxford, UK.

Hauer, E., Harwood, D.W., Council, F.M., Griffith, M.S., 2002. Estimating safety by the empirical Bayes method: a tutorial. Transportation Research Record 1784, 126–131.

Heckman, J.J., Ichimura, H., Todd, P.E., 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training programme. The Review of Economic Studies 64, 605–654.

Heinrich, C., Maffioli, A., Vázquez, G., 2010. A Primer for Applying Propensity-Score Matching. Inter-American Development Bank, Technical Notes No. IDB-TN-TN161.

Hess, S., Polak, J., 2003. Effects of speed limit enforcement cameras on accident rates. Transportation Research Record 1830, 25–34.

Hirano, K., Imbens, G.W., 2001. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. Health Services & Outcomes Research Methodology 2, 259–278.

Hirst, W.M., Mountain, L.J., Maher, M.J., 2004. Sources of error in road safety scheme evaluation: a quantified comparison of current methods. Accident Analysis and Prevention 36, 705–715.

Jones, A.P., Sauerzapf, V., Haynes, R., 2008. The effects of mobile speed camera introduction on road traffic crashes and casualties in a rural county of England. Journal of Safety Research 39, 101–110.

Keall, M.D., Povey, L.J., Frith, W.J., 2001. The relative effectiveness of a hidden versus a visible speed camera programme. Accident Analysis and Prevention 33, 277–284.

Kurth, T., Walker, A.M., Glynn, R.J., Chan, K.A., Gaziano, J.M., Berger, K., Robins, J.M., 2006. Results of multivariable logistic regression, propensity matching propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. American Journal of Epidemiology 163, 262–270.

Lechner, M., 2001. Programme heterogeneity and propensity score matching: an application to the evaluation of active labour market policies. Review of Economics and Statistics 84, 205–220.

Mountain, L.J., Hirst, W.M., Mahar, M.J., 2004. Costing lives or saving lives: a detailed evaluation of the impact of speed cameras. Traffic Engineering & Control 45 (8), 280–287.

Mountain, L., Mahar, M., Fawaz, B., 1997. The effects of trend over time on accident model predictions. In: Proceedings of the PTRC 25th European Transport Forum, P419, pp. 145–158.

Mountain, L.J., Hirst, W.M., Maher, M.J., 2005. Are speed enforcement cameras more effective than other speed management measures? The impact of speed management schemes on 30 mph roads. Accident Analysis and Prevention 37, 742–754.

Newstead, S., Cameron, M., 2003. Evaluation of the crash effects of the Queensland speed camera program. Monash University Accident Research Centre, Report No. 204.

Peikes, D.N., Moreno, L., Orzol, S.M., 2008. Propensity score matching: a note of caution for evaluators of social programs. The American Statistician 62, 222–231.

Persaud, B., Lan, B., Lyon, C., Bhim, R., 2009. Comparison of empirical Bayes and full Bayes approaches for before–after road safety evaluations. In: Presented at the 89th Annual Meeting of the Transportation Research Board, Washington, DC.

Persaud, B., Lyon, C., 2007. Empirical Bayes before–after safety studies: lessons learned from two decades of experience and future directions. Accident Analysis and Prevention 39 (3), 546–555.

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

Rudner, L.M., Peyton, J., 2006. Consider propensity scores to compare treatments. Practical Assessment, Research & Evaluation 11 (9).

Sayed, T., deLeur, P., Sawalha, Z., 2004. Evaluating the insurance corporation of British Columbia road safety improvement program. Transportation Research Record 1865, 57–63.

Shin, K., Washington, S.P., van Schalkwyk, I., 2009. Evaluation of the Scottsdale Loop 101 automated speed enforcement demonstration program. Accident Analysis and Prevention 41, 393–403.

Smith, J.A., Todd, P.E., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics 125, 305–353.

Thomas, L.J., Srinivasan, R., Decina, L.E., Staplin, L., 2008. Safety effects of automated speed enforcement programs: critical review of international literature. Transportation Research Record 2078, 117–126.