The Heterogeneous Treatment Effects of Speed Cameras on Road Safety

Haojie Li*

Daniel J. Graham

Centre for Transport Studies

Dept of Civil and Environmental Engineering

Imperial College London

SW7 2AZ

Word Count: 6965

Number of tables: 2; Number of figures: 3

Number of references: 33

Submitted date: 30/07/2014

*Corresponding author at: Centre for Transport Studies, Imperial College London, London SW7 2AZ, UK.

Tel.: +44(0)20 75942706; fax: +44(0)20 75946102.

E-mail address: h.li09@imperial.ac.uk

ABSTRACT

This paper aims to analyse the heterogeneous treatment effects of fixed speed cameras and evaluate the criteria of camera sites selection. We employ a nonparametric method to estimate how treatment effects continuously change against propensity scores. We first estimate individual treatment effect for each treated site using propensity score matching. Then treatment effects are estimated as a function of propensity scores using local polynomial regression. The results show that the effects of speed cameras are not uniform across the treated sites and are dependent of sites characteristics, measured by propensity scores. We further evaluate the criteria for selecting camera sites in the UK by comparing the effects on treated sites meeting and not meeting the criteria. The results show that camera sites which meet the criteria perform better in reducing casualties, implying the site selection criteria are rational.

1. INTRODUCTION

Speed limit enforcement cameras were first introduced in the UK in 1991 and were extended widely in the last decade. Numerous studies have been conducted to investigate the effect of safety cameras, and results show that the implementation of safety cameras has reduced vehicle speed and casualty numbers near camera sites (e.g. *1-4*). Despite the wealth of empirical evidence it remains unclear how such effects may vary across sites, referred to as heterogeneity of treatment effect (HTE). The variation in treatment effects is related to the differences in site characteristics, specifically the extent to which site characteristics meet treatment assignment criteria. The main objective of this study is to analyse how effects of fixed speed cameras change against propensity of selection into treatment, and identify the locations most benefited.

Although the importance of HTE has been widely recognized in causal analysis, most previous studies on speed cameras usually report an average treatment effect (ATE), which neglects the fact that the effects of speed cameras may differ systematically by the propensity for treatment. This is due in part to the fact that causal approaches for exploring HTE, used routinely in other areas of science such as medicine and epidemiology, have not yet been adopted in road safety studies. Understanding HTE has important implications for policy making. Treatments or trials, such as speed cameras, are usually costly. For example, the annual cost of safety cameras is around £100 million for 2003/04 in the UK (*3*). It is desirable that the treatment is operated in a way that maximises effectiveness with limited resources. By revealing patterns of HTE, policy makers can assign treatments to individuals most likely to benefit from the treatment, so as to improve cost-effectiveness of treatment. This research tackles this issue by applying and developing causal approaches for estimating heterogeneous treatment effects of speed cameras on road safety conditional on unconfoundedness assumption.

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 Section 3 and Section 4. The results are presented and discussed in Section 5. The conclusions are given in the final section.

2. LITERATURE REVIEW

In the past decade, numerous studies have been conducted to investigate the impact of speed enforcement cameras on safety (*5-9*). In general, these studies show that the implementation of speed cameras has significantly reduced vehicle speeds and the number of casualties near camera sites. There are two outstanding issues, however, which have yet to be fully addressed in the previous evaluations of the effects of speed cameras on road casualties.

The first issue is regarding the selection of the reference or control group. Most studies to date have used before-and-after methods with control groups (*2, 5, 7, 9*). In these studies, a group of similar sites is usually selected as the control group in order to account for the general trend in casualties. However, this method is unable to control for effects of regression to mean (RTM), also known as 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 casualty rate will tend to be lower in subsequent years. The impact of the RTM is that it can make random variation appear as real change caused by treatments and therefore overestimate the effect of a safety treatment.

A reference or control group is usually required to estimate the counterfactual outcomes of the treatment group. Due to the confounding factors, however, 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 and the severity of its impact. This means that only untreated units with similar characteristics to those treated can be used to approximate the counterfactual outcomes of the treatment group. 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.

The propensity score matching (PSM) method is proposed by Rosenbaum and Rubin (*10*) for selecting control groups and estimating causal effects. The PSM method has been widely used as a tool of evaluation in econometrics (*11-18*). Recently, this approach has been introduced and employed in evaluation studies of road safety measures (*4, 19*). We will discuss PSM in the next section.

The second issue arising from these studies is that only ATE is estimated, neglecting the fact that treatment effects can differ across the treated population. ATE provides useful information but policy makers also care about effects within specific subpopulation. Since road safety measures are usually costly, it is desirable that treatments are assigned to areas or individuals which are most likely to benefit from the treatment. A good knowledge of the pattern of treatment effects can help policy makers to make optimal decisions with limited resources. Most of previous researches on the effect of speed cameras, however, focus on the average benefit, ignoring the fact that the impact may vary across sites with different characteristics.

Several approaches to estimating HTE based on the propensity scores have been proposed and applied in a few quantitative sociological researches. For example, Xie et al. (*20*) discuss a practical approach to studying HTE as a function of the treatment propensity under the unconfoundedness assumption. Three methods, one parametric and two non-parametric, are described for analysing interactions between treatment and the treatment propensity. They apply the three methods to estimate the effects of college attendance on women's fertility based on the work by Brand and Davis (*21*). This study applies the approaches introduced by Xie et al. (*20*) to estimate HTE of speed cameras on road casualties.

3. METHODS

As discussed earlier, the conventional propensity score methods can be applied for identifying potential comparison or control groups with similar characteristics to the treated ones. In addition, matching approaches can be developed by combining with linear and non-linear regressions to explore and estimate HTE.

In this section, we first introduce the propensity scores and the conditions under which it can be used to evaluate the effect of interventions. Then two approaches based on propensity scores are discussed for ATE and HTE estimation.

3.1 Motivations

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,…, N and N denote the total population. The treatment effect for unit i can be described as:

$\delta_i=Y_i(1)-Y_i(0)$ (Individual Treatment Effect)

The fundamental problem of causal inference is that it is impossible to observe the outcomes of the same unit i in both treatment conditions at the same time (*22*). In practice, control groups are usually selected from untreated units to construct counterfactual outcomes for treated units. In simple control studies, the average treatment effect on the treated (ATE) is estimated by taking comparisons of the average outcomes between treated and control units, which can be defined as:

$$\delta_{ATE}=E[Y(1)|T=1]-E[Y(0)|T=0]$$
$$=E[Y(1)-Y(0)|T=1]+\{E[Y(0)|T=1]-E[Y(0)|T=0]\}$$

In the above equation, the term in curly brackets is zero for randomized experiments, where the probability of assignment to treatment does not depend on potential outcomes. That is,

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

Then $E[Y(0)|T=1]= E[Y(0)|T=0]$ and therefore

$$\delta_{ATE}=E[Y(1)|T=1]-E[Y(0)|T=0]$$
$$=E[Y(1)-Y(0)|T=1]+\{E[Y(0)|T=1]-E[Y(0)|T=0]\}$$
$$=E[Y(1)-Y(0)|T=1] \quad \text{(ATE with randomized assignment)}$$

which is an unbiased estimator of the ATE. However, in most cases it is not always feasible to implement a randomized experiment due to high costs and ethical issues. In other words, since the treatment assignment is not random and affected by pre-treatment variables, there can be systematic differences between treated and untreated units, and they can affect the outcome, *Y*. That is the selection of treated units is affected by a vector of covariates ***X***, which also have impacts on potential outcomes.

To relax the strict requirement for randomized experiments, unconfounded assignment is proposed by Rosenbaum and Rubin (*10*). In the context of unconfounded assignment, it still requires that the probability of receiving the treatment is dependent on the potential outcomes, however, conditional on covariates ***X***:

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

With unconfounded assignment it is possible to account and adjust for differences in pre-treatment covariates and outcomes between treatment and control groups in order to properly estimate the effect of treatment:

$$\delta_{ATE}=E[Y(1)|T=1, X]-E[Y(0)|T=0, X]$$
$$=E[Y(1)-Y(0)|T=1, X]+\{E[Y(0)|T=1, X]-E[Y(0)|T=0, X]\}$$
$$=E[Y(1)-Y(0)|T=1, X] \quad \text{(ATE with unconfounded assignment)}$$

3.2 Propensity Score Matching

As discussed above, in most cases treatment assignments are not random and the inferences could be biased due to confounding factors, ***X***. Conditional on unconfounded assignment, however, the treatment effects can be properly estimated.

Matching is one such approach based on unconfounded assignment. The basic idea behind matching is to match each treated unit to an untreated unit with the same values on observed characteristics, such as a vector of covariates ***X***. The matching approach becomes more difficult to implement as the number of observed covariates used increases, however. This obstacle can be overcome by matching on a single index instead of multiple dimensions. The most well-known index is the propensity score, which is the probability that a unit is selected into the treatment group conditional on observed covariates. 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. But the validity of this approach rests on two assumptions which we will discuss in the next section.

3.3 Assumptions

The first assumption for an unconfounded assignment is known as the Conditional Independence Assumption (CIA), which assumes all observed differences in characteristics between the treated and untreated units are controlled for, and the outcomes that would result in the absence of treatment are the same for both groups.

Rosenbaum and Rubin (*10*) proposed the idea of balancing scores, suggesting that if potential outcomes are independent of treatment conditional on covariates ***X***, they are also independent of treatment conditional on a balancing score, such as the propensity score. The CIA based on the propensity score can thus be described as:

$(Y(1), Y(0)) \perp T|P(\mathbf{X}), \ \forall \mathbf{X}$ (Unconfoundedness given the propensity score)

The second problem is regarding the chance of finding a match for each individual with the same propensity score. It is likely that there is no match in the control group with a similar propensity to that of any treated individual. So it requires that individuals with the same ***X*** values have a positive probability of being in both treated and untreated groups. In other words, the proportion of treated and untreated individuals must be greater than zero for every possible value of ***X***. This condition is usually referred as the common support condition or overlap condition, which ensures sufficient overlap in the characteristics between the treated and untreated units to find adequate matches. The overlap condition can be described as:

$0<P(D=1|X)<1$ (Overlap Condition)

There are several methods for checking these two assumptions and assessing the matching quality. We will discuss this in detail later.

3.4 Estimation of Propensity Score

Because linear probability models produce predictions outside the [0, 1] bounds of probability, logit and probit models are usually used estimating propensity score. For binary treatment, logit and probit models usually yield similar results, hence the choice between them is not critical (see further discussion of this point in *23*). In this paper, a logit model is used:

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

Where α is the intercept and β' is the vector of regression coefficients. The selection of covariates included in PSM will be discussed in section 4.

3.5 Inferences on Treatment Effects

Here we discuss propensity score matching and regression method for estimating ATE and HTE under the unconfoundedness assumption.

3.5.1 Average Treatment Effects

Once the propensity score is estimated, the most straightforward approach for estimating treatment effects is matching. In general, the treatment effect can be estimated as $Y_i(1)-Y_{j(i)}(0)$, where $Y_{j(i)}$ is the outcome for the comparison unit j that is matched with the treated unit i. In the early stage, treated units were paired with those in the comparison on a one-to-one basis. For each treated unit, only the unit in the comparison group with the most similar propensity score is matched with that treated unit. Usually, such pairwise matching is performed without replacement, which means each comparison group member can be used as a matched unit only once. The problem is that the matching performance could be poor when the sample size of the control group is small or when there is little common support for two groups (*14*). More studies use matching with replacement and other matching algorithms other than pairwise matching.

In contrast to matching one treated unit with only one comparison unit, using all comparison units that are sufficiently close to a given treated unit is a more stable approach. To account for the sampling error, it is important to include only those comparison units that are close, to within a certain tolerance, to a given treated unit. There are four mostly used matching algorithms: nearest neighbour 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. (*24*).

Then the effects can be calculated by averaging the differences in outcomes between treated units and matched comparison units.

$$\delta_{ATE} = E[Y(1)-Y(0)|T=1] = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_{j(i)}(0))$$

A number of statistical software programs are available to perform matching and evaluate average effects. A frequently used program, **psmatch2**, has been developed by Leuven and Sianesi (*25*) and can be installed in Stata. All matching algorithms can be implemented in this program. Functions, such as common-support graphing (**psgraph**) and covariate balance tests (**pstest**) are also included in **psmatch2**.

It is also important to estimate the standard errors to indicate the sampling error. Bootstrap methods are widely used to obtain standard errors in PSM and can be easily implemented in **psmatch2** or the Becker and Ichino (*26*) PSM estimation program.

Another widely used approach in before-after traffic safety countermeasures evaluations is empirical Bayes (EB) method. Applications of the EB method in analysis of speed camera effects include studies by Mountain et al. (*1*), Shin et al. (*8*), and Gains et al. (*3*). Recently, Li et al. (*4*) compare the EB with PSM for ATE estimation and find similar results from these two methods. Their results also suggest that the propensity score can be used as the criteria for the reference group selection. To avoid confusing the focus, we exclude the EB

method in this study.

3.5.2 Heterogeneous Treatment Effects

We discuss the approach for estimating HTE using smoothing method (*20*). The procedures can be illustrated as following steps.

(1) The first step is to estimate propensity scores for all individuals in treated and comparison groups as discussed in ATE estimation.

(2) In this step, each treated individual is matched to a comparison individual or multiple individuals with the similar propensity score. The treatment effect for each treated individual can be obtained by taking the differences in outcomes between treated individuals and matched comparison individuals.

(3) Given sufficient observations it is possible to fit curves and surfaces to data by smoothing. One widely used approach to smoothing data is local polynomial regression. The curve depicts how treatment effects continuously change against propensity scores. Specifically, individual treatment effects estimated in step (2) can be described as a function of the propensity score:

$$\delta_{HTE}=f(P(\mathbf{X}))$$

Where $P(\boldsymbol{X})$ is the propensity score given observed covariates $\boldsymbol{X}$, $f(P(\boldsymbol{X}))$ is assumed as a polynomial function, which can be expressed as:

$$f(P(\mathbf{X})) = \mu_0 + \mu_1 P(\mathbf{X}) + \mu_2 P(\mathbf{X})^2 + \mu_3 P(\mathbf{X})^3 \ldots. + \mu_m P(\mathbf{X})^m$$

The best fitting power, *m*, is selected by maximizing the likelihood of this equation.

4. DATA

4.1 Confounders

It is explicit that the validity of PSM largely relies on unconfoundedness assumption, $(Y(1), Y(0)) \perp T \mid \boldsymbol{X}$, where $\boldsymbol{X}$ is a vector of confounders. Only covariates that affect both treatment participation and potential outcomes should be included in PSM. In practice, however, selection of such covariates can be complex due to the lack of precise knowledge of treatment assignment mechanism. Discussions are available regarding covariates choice in PSM (*11, 27-30*).

On the one hand, although all the observed covariates could be included, this could generate problems with the common support (*29*). Another reason for avoiding over-parameterized models is that although the inclusion of non-significant covariates will not affect the unbiasedness and consistency of the estimates, it can increase their variance, especially with small samples (*30*).

On the other hand, it is suggested that omitting important covariates can cause serious bias in estimation (*11*). Rubin and Thomas (*27*) recommend that a covariate should only be excluded if there is consensus that the covariate is unrelated to either the outcome or participation. If there are doubts about this, it is advised to include the relevant covariates.

A simulation study by Brookhart et al. (*28*) illustrates how the choice of variables included in PSM can affect the bias, variance, and mean squared error of estimated treatment effects. Their results suggest that the optimal practice, in terms of bias and precision, is to include all covariates that affect the outcome regardless of whether they have impacts on treatment assignment. In contrast, however, adding a covariate unrelated to the outcome but related to treatment assignment will increase the variance without decreasing bias.

4.2 Covariates Included in PSM

The covariates inclusion would be less complicated if criteria for treatment participation were available. Where such criteria are not available, it is still possible to choose covariates based on previous empirical findings. In this study two sets of covariates are considered to be included in PSM.

4.2.1 Covariates suggested in the handbook for camera site selection

Currently, in the UK, formal site selection guidelines for fixed speed camera sites exist (*2*), as shown below.

(1) Site length: Between 400-1500 metres.

(2) Number of fatal and serious collisions (FSCs): at least 4 FSCs per km in the last three calendar years.

(3) Number of personal injury collisions (PICs): at least 8 PICs per km in the last three calendar years.

(4) 85^{th} percentile speed at collision hot spots: 85^{th} 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. Secondary criteria such as the 85^{th} percentile speed and percentages of vehicles over the speed limit are not normally publically available for all sites on UK roads, however. For untreated sites with a speed limit of 30 mph or 40 mph, the national average mean speed and percentages of speeding are similar to the data for the camera sites (*1-2*). The focus groups for this study are sites with a 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.

Selection of speed camera sites, therefore, is primarily based on site length and accident history, both of which are also important predictors in road safety analysis. Pre-treatment casualty records are valuable covariates when estimating with PSM because they are important predictors of treatment entry and subsequent outcomes in post-treatment period. The road length is also an important exposure variable to obtain risk estimates in road accident analysis. In addition, the handbook for camera site selection suggests that "The decision as to what type of camera (fixed or mobile) should be used at a site should take into account the volume, severity and distribution of collisions" (*31*). This indicates that traffic volume also influences site selection and needs to be included in PSM.

4.2.2 Covariates suggested as important factors affecting road casualties

Notwithstanding the guidelines discussed above there are sites not meeting the criteria which may still be selected as enforcement sites for one or more of the other reasons, such as community concern, collision frequency and engineering factors (*31*). In other words, there are unknown factors that affect treatment assignment but are not explicitly described in the handbook for camera site selection. As suggested by Rubin and Thomas (*27*) and Brookhart et al. (*28*), unless there is consensus that the covariate is unrelated to treatment participation, covariates that affect the outcome should be included in PSM, because they decrease the variance of the estimated treatment effect without increasing bias. Hence covariates suggested as important factors when estimating road casualties at camera sites are also considered to be included in PSM. Covariates further included in PSM are speed limit, road types and the number of minor junctions within the site length, which have been suggested as important factors when estimating the safety impact of speed cameras (*1, 3, 5*).

4.3 Sample Size

The PSM method is known as a "data-hungry method" in terms of the number of treated and untreated units. Matching can only be 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 scores, these treated units will be discarded and the estimation of ATE will be biased. Thus, a large untreated pool is required to ensure adequate matches. The literature is not explicit, however, 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 (*12, 14-15, 32-33*). The ratio chosen in this study was around 7:1, which was assumed to be

sufficient to ensure the matching quality. Due to data restrictions, 771 camera sites from the following eight English administrative districts were 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. The accident data for the three years before and after the camera installation were acquired for every site and the research period covered nine years from 1999 to 2007. Whilst concerns have been raised about the completeness and reliability of accident data in STATS19, in the case of casualties at speed camera sites, given the nature of such sites, it is likely that all casualties were captured and that the data is therefore reliable and complete.

5. HTE OF SPEED CAMERA ON ROAD CASUALTIES

5.1 The Estimation of Propensity Scores

The first step is to estimate the propensity score. Table 1 shows that all covariates except minor road are significant in the estimation of the propensity score. This is probably because there are only 19 observations for speed cameras installed on minor road in the study sample. It is also worth noting that only the number of FSCs is positively correlated to propensity scores, while the number of PICs is not. This indicates that local authorities put more values on the number of FSCs than PICs in practice, which is also consistent with the rules for proposed cameras sites described in the handbook (*31*). The result, in general, confirms that the covariates included in the propensity score model are important in predicting the possibility of being selected as camera sites.

TABLE 1The Propensity Score Model Check

	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 30mph	1.017	(0.101)	10.11	0.000	0.820	1.214
Speed Limit 40mph	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					

5.2 Tests of Matching Quality

We first check the validity of the PSM method before estimating the effects of speed cameras. There are two routine tests, one of which 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 (*29*). However, if the proportion is too large, the true treatment effect can be misestimated. Figure 1 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.

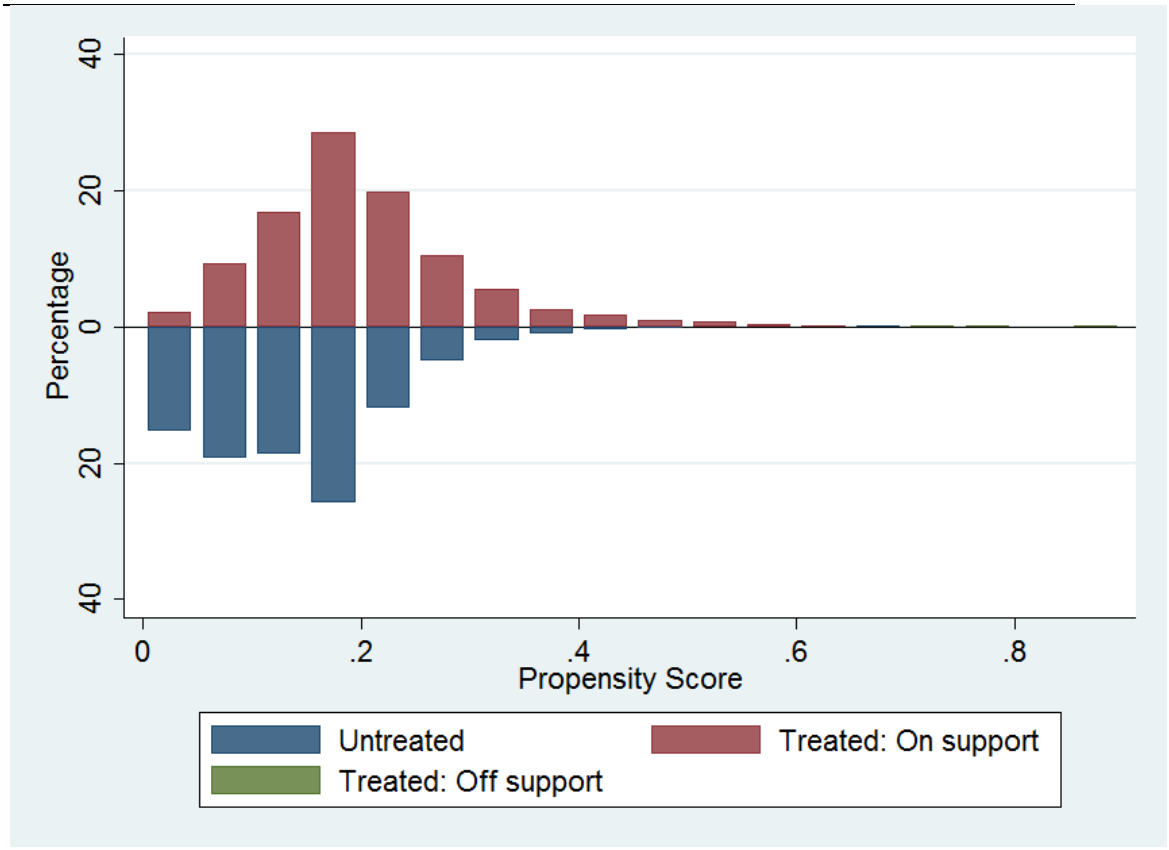

FIGURE 1 Propensity score distribution.

A balancing test is performed to assess the matching quality as this test 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 2 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 2 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.

TABLE 2 Checking the Covariates Balance between Groups Before and After Using Nearest Neighbours (k=5) Matching

Variable	Sample	Mean Treated	Control	%bias	%reduced |bias|	t-test 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	19039	18020	10.1	88.3	2.52	0.012
	Matched	19049	19168	-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 30mph	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 40mph	Unmatched	0.1219	0.1828	-17.0	97.9	-4.14	0.000
	Matched	0.1224	0.1237	-0.4		-0.08	0.938

5.3 HTE of Speed Cameras on Road Casualties

The hypothesis of this study is that the average effect of treatment on the population bears little resemblance to the real effect estimated for individual treated sites. We estimate HTE of speed cameras on road casualties using smoothing method in this section. We first estimate propensity scores for all individuals in treated and comparison groups. The treated and comparison individuals are then matched via kernel matching (caliper=0.05). The treatment effect for each treated individual is estimated by taking the differences in outcomes between matched pairs. Smooth curves through the data points of individual treatment effects are plotted using local polynomial regressions. Local linear and quadratic are employed instead of higher-degree polynomials which may tend to overfit the data. The first two graphs in Figure 2 show "U" shape curves of HTE for PICs and FSCs. It is worth noting, however, the 95% quantile of propensity score is 0.32, above which there is only 5% of the population. To get a clearer pattern of HTE, we exclude the sites with propensity scores higher than 0.32 and re-examine the pattern of HTE as shown in the two graphs at the bottom of Figure 2.

We then compare ATE and HTE of speed cameras on PICs and FSCs. Li et al. (*4*) estimate ATE of speed cameras using the PSM and EB methods. Their results show that the average reduction in annual PICs is around 1.1 per km. The estimation of HTE in this study, however, suggests that the reduction in annual PICs ranges from 0.5 to 3 per km. Similar case applies to the effect on FSCs. The annual reduction in FSCs is estimated to be 0.13 per km on average, while the approximate number varies from -0.1 to 0.4 per km in HTE estimation. It is explicit that the treatment effect is not uniform and is highly dependent on propensity scores. It can be speculated that a treatment decision based on the average effect for the entire population would underestimate the effect in the subpopulation with high propensity scores and vice versa.

Despite that the propensity score is a useful index by simplifying matching, it is not available to decision makers. Furthermore, treatment decision is made based on the criteria (a set of observables) rather than propensity scores. Thus it is important to investigate how the distribution of treatment effects is related to treatment assignment criteria.

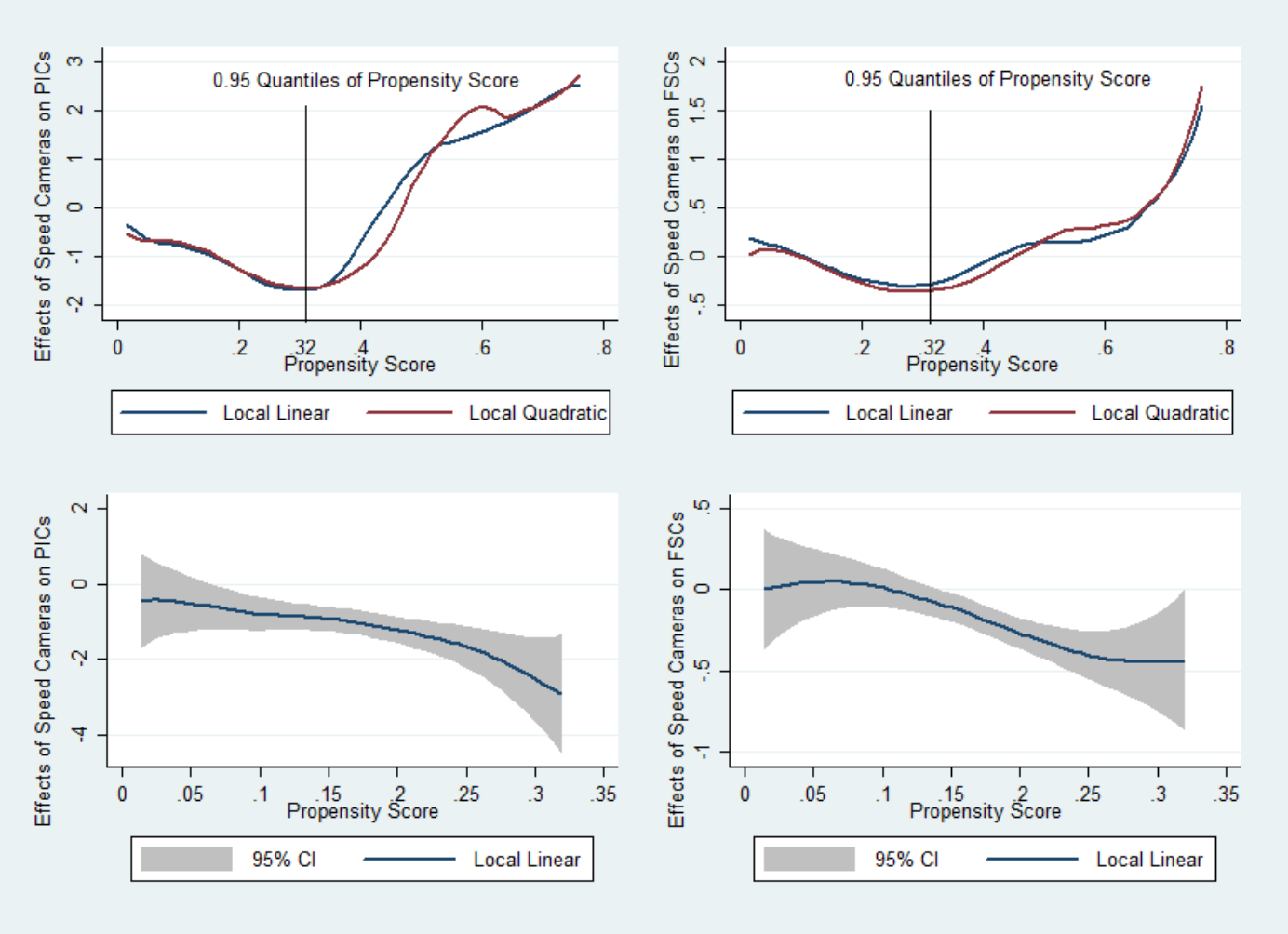

FIGURE 2 HTE of speed cameras.

5.4 Criteria for Proposed Camera Sites

In this section we evaluate the criteria for selecting camera sites in the UK. We assume that the main criterion affecting camera sites selection is historical casualties (*2, 31*). It is worth noting that the rules for proposed fixed speed camera sites are slightly different in the handbook published by DfT (*31*). Most rules are consistent with the criteria described in the four-year report (*2*), however, a criterion termed "total value required" is introduced in this handbook. That is, "new camera sites will be selected using an assessment that includes the level of fatal, serious and slight collisions. The combined level of collisions will be expressed as numerical scale" (*31*). For example,

Fatal and serious injury collision = 5 (i.e. 2 serious collision = 10)

Slight injury collision = 1 (i.e. 5 slight collisions = 5).

The total value required is 22 per km for proposing fixed effects camera on a road with a speed limit of 40 mph or less. This criterion is termed as risk values for clarity in this study. The risk values are excluded in the propensity score models to avoid the perfect multicollinearity with the number of PICs and FSCs. The risk values, however, are treated as the primary criterion to be evaluated in this study to avoid the complexity due to multiple criteria.

It is very likely that the rules are not strictly complied with in practice. And sites not meeting the criteria may still be selected as exceptional sites for other reasons, such as community concern and engineering factors. Therefore treated sites meeting and not meeting the criteria are both observed in the sample. The idea is to compare the treatment effects between these two types of treated sites. If the treatment is more effective with the sites not meeting the criteria, then we may conclude that the criteria for selecting camera sites are not optimised. There are 414 observations out of 771 camera sites with risk values higher than 22 per km.

The patterns of treatment effects across the risk values are described in Figure 3, where the required minimum risk value for installing speed cameras is also marked. Smooth curves are plotted using local linear and quadratic polynomial regressions. The speed cameras show no effects or negative effects on reducing PICs and FSCs where risk values are lower than 22 per km. In contrast, significant reductions on both PICs and FSCs are observed at sites with risk values higher than 22 per km.

It is suspected that treatment evaluation at black spot sites with high casualty record may suffer the RTM effect. That is, the reduction in casualties can be caused by random variation rather than the effects of speed cameras. In this study, by controlling for the RTM effect using PSM, Figure 3 shows that treatment effects increase as risk values increase, suggesting that speed cameras are more effective at sites with higher historical casualties records. However an opposite trend is observed for sites with risk values higher than 90 per km. This is probably due to the small sample size, which are 31 for the control sites compared to 48 for the treated sites.

In addition, we examine the distribution of treatment effects for treated sites in Figure 3. Camera sites with risk values higher than 22 per km are found to have higher treatment effects, indicating that sites meeting the selection criterion perform better in reducing casualties. According to the above results, we can conclude that it is reasonable to use risk values of 22 per km as the main criterion for selecting speed camera sites.

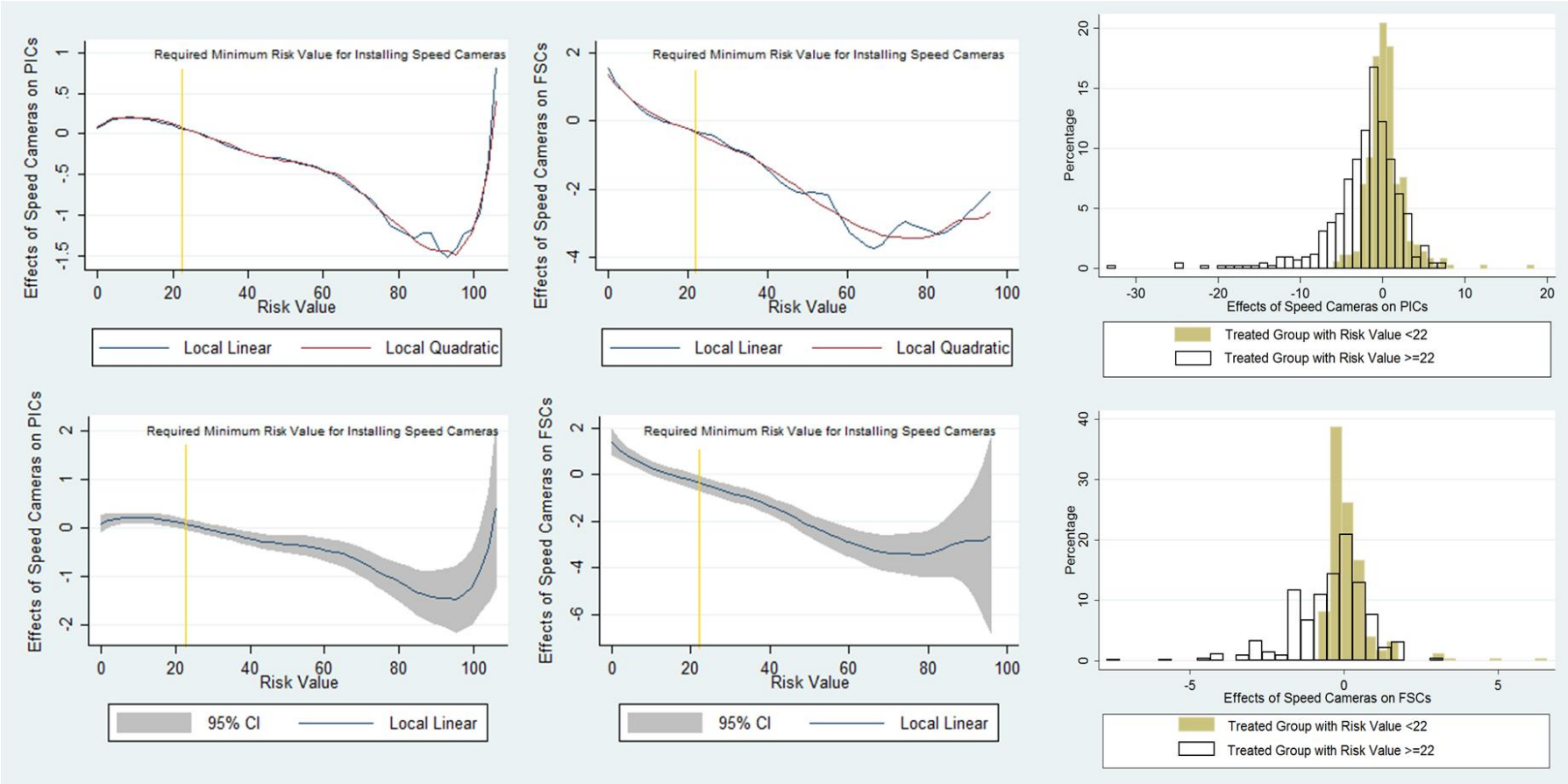

FIGURE 3 HTE on PICs and FSCs by risk values

6. DISCUSSIONS AND CONCLUSIONS

Most previous studies on speed cameras only estimate average treatment effects and neglect the fact that the effects may vary across sites. This is probably due to the fact that causal approaches for exploring the heterogeneity of treatment effect have not yet been widely used in road safety studies. This study contributes to the literature by applying causal models to estimate heterogeneous treatment effects of speed cameras on road safety conditional on unconfoundness assumption. We hypothesize that the responses to the treatment may vary due to differences in site characteristics measured by propensity scores. Local polynomial regressions are employed to plot smooth curves of individual treatment effects, providing us the pattern of treatment effects as a continuous function of propensity scores. The HTE estimation is then compared with ATE of speed cameras on PICs and FSCs. The results suggest that the treatment decision based on ATE can be misled because the average effect for the entire population would underestimate the effect in the subpopulation with high propensity scores and vice versa.

The camera sites are selected based on certain criteria, such as the number of KSIs and PICs in the baseline years. It remains unclear, however, whether such criteria are optimized for effectiveness. This study evaluates the criteria for selecting camera sites in the UK by comparing the effects on treated sites meeting and not meeting the criteria. To avoid the complexity due to multiple criteria, risk values are selected as the main criterion for evaluation. In general, it is found that camera sites which meet the criterion perform better in reducing casualties. Only 57 percent of the treated sites, however, meet the requirement for risk values. Despite that there are exceptional reasons for selecting sites not meeting the criteria, the results suggest that installing speed cameras at sites with risk values lower than 22 per km can be ineffective.

References

1. Mountain, L. J., W. M. Hirst, and M. J. Maher. 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, 2005, pp. 742-754.
2. Gains, A., B. Heydecker, J. Shrewsbury, and S. Robertson. The national safety camera programme 3-year evaluation report. PA Consulting Group and UCL for Department for Transport, London, 2004.
3. Gains, A., B. Heydecker, J. Shrewsbury, and S. Robertson. The national safety camera programme 4-year evaluation report. PA Consulting Group and UCL for Department for Transport, London, 2005.
4. Li, H., D. J. Graham, and A. Majumdar. The impacts of speed cameras on road accidents: An application of propensity score matching methods. Accident Analysis & Prevention, 60, 2013, pp. 148-157.
5. Christie, S. M., R. A. Lyons, F. D. Dunstan, and S. J. Jones. Are mobile speed cameras effective? A controlled before and after study. Injury Prevention, 9, 2003, pp. 302-306.
6. Mountain, L. J., W. M. Hirst, and M. J. Mahar. Costing lives or saving lives: a detailed evaluation of the impact of speed cameras. Traffic Engineering & Control, 45(8), 2004, pp. 280-287.
7. Cunningham, C. M., J. E. Hummer, and J. Moon. Analysis of Automated Speed Enforcement Cameras in Charlotte, North Carolina. Transportation Research Record, 2078, 2008, pp. 127-134.
8. Shin, K., S. P. Washington, and I. van Schalkwyk. Evaluation of the Scottsdale Loop 101 automated speed enforcement demonstration program. Accident Analysis and Prevention, 41, 2009, pp. 393-403.
9. Jones, A. P., V. Sauerzapf, and R. Haynes. The effects of mobile speed camera introduction on road traffic crashes and casualties in a rural county of England. Journal of Safety Research, 39, 2008, pp. 101-110.
10. Rosenbaum, P. R., and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. Biometrika, 70, 1983, pp. 41-55.

11. Heckman, J. J., H. Ichimura, and P. E. Todd. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. The Review of Economic Studies, 64, 1997, pp. 605-654.
12. Hirano, K., and G. W. Imbens. Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization. Health Services & Outcomes Research Methodology, 2, 2001, pp. 259-278.
13. Dehejia, R.. Practical propensity score matching: a reply to Smith and Todd. Journal of Econometrics, 125, 2005, pp. 355-364.
14. Dehejia, R. H., and S. Wahba. Propensity score-matching methods for nonexperimental causal studies. The Review of Economics and Statistics, 84(1), 2002, pp. 151-161.
15. Kurth, T., A. M. Walker, R. J. Glynn, K. A. Chan, J. M. Gaziano, K. Berger, and J. M. Robins. Results of Multivariable Logistic Regression, Propensity Matching, Propensity Adjustment, and Propensity-based Weighting under Conditions of Nonuniform Effect. American Journal of Epidemiology, 163, 2006, pp. 262-270.
16. Lechner, M..Programme heterogeneity and propensity score matching: an application to the evaluation of active labour market policies. Review of Economics and Statistics, 84, 2001, pp. 205-220.
17. Abadie, A., and G. Imbens. Large Sample Properties of Matching Estimators for Average Treatment Effects. Working Paper, Harvard University, 2004.
18. Abadie, A., and G. Imbens. Matching on the Estimated Propensity Score. Working Paper, Harvard University, 2009.
19. Sasidharan, L., and E. T. Donnell. Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: Exploratory analysis using intersection lighting data. Accident Analysis & Prevention, 50, 2013, pp. 539-553.
20. Xie, Y., J. E. Brand, and B. Jann. Estimating Heterogeneous Treatment Effects with Observational Data. Sociological Methodology, 42, 2012, pp. 314-347.
21. Brand, J. E., and D. Davis.The Impact of College Education on Fertility: Evidence for Heterogeneous Effects. Demography, 48, 2011, pp. 863-887.
22. Holland, P. Statistics and Causal Inference (with discussion). Journal of the American Statistical Association, Vol. 81, 1986, pp. 945-970.
23. Smith, H. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies," Sociological Methodology, 27, 1997, pp. 325-353.
24. Heinrich, C., A. Maffioli, and G. Vázquez. A Primer for Applying Propensity-Score Matching. Inter-American Development Bank, Technical Notes, No. IDB-TN-161, 2010.
25. Leuven E., and B. Sianesi. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components S432001, Boston College Department of Economics, 2003 (revised 12 Feb 2014).
26. Becker, S. O., and A. Ichino.Estimation of Average Treatment Effects Based on Propensity Scores. The Stata Journal, 2(4), 2002, pp. 358-377.
27. Rubin, D. B., and N. Thomas. Matching Using Estimated Propensity Scores: Relating Theory to Practice. Biometrics, 52, 1996, pp. 249-264.
28. Brookhart, M. A., S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn, and T. Sturmer. Variable selection for propensity score models. American Journal of Epidemiology, 163, 2006, pp. 1149-1156.
29. Bryson, A., R. Dorsett, and S. Purdon. The Use of Propensity Score Matching in the Evaluation of Labour Market Policies. Working Paper No. 4, Department for Work and Pensions, 2002.
30. Augurzky, B. and C.M. Schmidt. The Propensity Score – A Means to an End. IZA Discussion paper, No. 271, Bonn, 2000.
31. Department for Transport. Handbook of Rules and Guidance for the National Safety Camera Programme for England and Wales for 2006/07. 2005.
32. Smith, J. A., and P. E. Todd. Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of Econometrics, 125, 2005, pp. 305-353.

33. Peikes, D. N., L. Moreno, and S. M. Orzol. Propensity Score Matching: A Note of Caution for Evaluators of Social Programs. The American Statistician, 62, 2008, pp. 222-231.