The Heterogeneous Treatment Effects of Speed Cameras on Road Safety

Haojie Li*

School of Transportation

Southeast University

Daniel J. Graham

Centre for Transport Studies

Dept of Civil and Environmental Engineering

Imperial College London

*Corresponding author at: School of Transportation, Southeast University, Nanjing 210096, China.

Tel.: +86(0)25 83795642

E-mail address: h.li@seu.edu.cn

ABSTRACT

This paper analyses how the effects of fixed speed cameras on road casualties vary across sites with different characteristics and evaluates the criteria for selecting camera sites. A total of 771 camera sites and 4787 potential control sites are observed for a period of 9 years across England. Site characteristics such as road class, crash history and site length are combined into a single index, referred to as a propensity score. We first estimate the average effect at each camera site using propensity score matching. The effects are then estimated as a function of propensity scores using local polynomial regression. The results show that the reduction in personal injury collisions ranges from 10% to 40% whilst the average effect is 25.9%, indicating that the effects of speed cameras are not uniform across camera sites and are dependent on site characteristics, as measured by propensity scores. We further evaluate the criteria for selecting camera sites in the UK by comparing the effects at camera sites meeting and not meeting the criteria. The results show that camera sites which meet the criteria perform better in reducing casualties, implying the current site selection criteria are rational.

Key words: Speed Camera; Heterogeneous Treatment Effect; Propensity Score

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. Mountain et al.,2005; Gains et al., 2004; Gains et al., 2005; Li et al., 2013). 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 hypothesis is that 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 the site characteristics influence the effects of fixed speed cameras, and identify the locations which have benefited most from treatment.

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 site characteristics. 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 (Gains et al., 2005). 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 units most likely to benefit from the treatment, so as to improve the cost-effectiveness of intervention. In this paper we tackle this issue by applying and developing causal approaches for estimating heterogeneous treatment effects of speed cameras on road casualties.

The paper is organized as follows. The literature review is presented in Section 2. The method and data used in the 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

Several studies have been conducted to analyse the effects of speed enforcement cameras on safety (Christie et al., 2003; Mountain et al., 2004; Cunningham et al., 2008; Shin et al., 2009; Jones et al., 2008). 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 regards the selection of the reference or control group. Most studies to date have used before-and-after methods with control groups (Gains et al., 2004; Christie et al., 2003; Cunningham et al., 2008; 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 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. 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. However, treated and untreated units may differ in the absence of any treatment due to confounding characteristics, which affect both potential outcomes and treatment participation. In other words, confounding characteristics of units that are treated may differ in some systematic way from those that are not treated, and those characteristics also may have a bearing on the incidence of selection bias and the severity of its impact. This means that only untreated units with similar confounding characteristics to the 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 (Rosenbaum and Rubin, 1983) for selecting control groups and estimating causal effects. The PSM method has been widely used as a tool of evaluation in econometrics (Heckman et al., 1997; Hirano and Imbens, 2001; Dehejia, 2005; Dehejia and Wahba, 2002; Kurth et al., 2006; Lechner, 2001; Abadie and Imbens, 2004; Abadie and Imbens, 2009). Recently, this approach has been introduced and employed in evaluation studies of road safety measures (e.g. Li et al., 2013; Sasidharan and Donnell, 2013). We will discuss PSM in the next section.

The second issue arising from these studies is that only ATE is estimated, however, neglecting the fact that treatment effects can differ across the treated population. The ATE provides useful information, but policy makers also care about effects within specific subpopulations. Since road safety measures are usually costly, it is desirable that treatments are assigned to areas or units 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 previous studies 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 studies. For example, Xie et al. (2012) discuss a practical approach to studying HTE as a function of treatment propensity under the unconfoundedness assumption. Three methods, one parametric and two non-parametric, are described for analysing interactions between treatment effects 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 (2011). This study applies the approaches introduced by Xie et al. (2012) to estimate HTE of speed cameras on road casualties.

3. METHODS

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

3.1 Propensity Score Matching

The treatment indicator is defined as T_i=1 if unit *i* receives the treatment and T_i =0 otherwise. $Y_i(T)$ denotes the potential outcome for unit *i*, where *i*=1,…, *N* and *N* denote the total population. For instance, *E[Y(0)|T=1]* is the expected value of the outcome *Y* of treated units when not exposed to the treatment. 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 (Holland, 1986). In practice, control groups are usually selected from untreated units to construct counterfactual outcomes for treated units. However, since the treatment assignment is usually not random and affected by pre-treatment variables, there can be systematic differences between treated and untreated units, and they can affect the potential outcomes, *Y*.

The basic idea behind matching is to match each treated unit to untreated units with the same values on observed characteristics, such as a vector of control variables ***X***. The matching approach becomes more difficult to implement as the number of observed control variables 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 confounding variables. Conditional on the propensity score, differences in observed outcomes between the two groups can be solely attributed to the intervention impacts.

The validity of this approach rests on two assumptions, conditional independence assumption (CIA) and overlap assumption, which can be described as:

$(Y(1), Y(0)) \perp T | P(\mathbf{X}), \ \forall \mathbf{X}$ (Conditional independence assumption)

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

For a full discussion of these assumptions please see Abadie and Imbens (2009). It is important to check the validity of the assumptions before estimating the treatment effects. There are several methods for checking these two assumptions and assessing the matching quality. We will discuss this in detail later.

Because linear probability models produce predictions outside the [0, 1] bounds of probability, logit and probit models are usually used to estimate propensity scores. For binary treatment, logit and probit models usually yield similar results, hence the choice between them is not critical. Please refer to the paper by Smith (1997) for further discussion of this point. In this paper, a logit model is used:

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

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

3.2 Inferences on Treatment Effects

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

3.2.1 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 control unit j that is matched with the treated unit i. In general, 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. (2010).

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

δ_{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 (2003) and can be installed in Stata.

3.2.2 Heterogeneous Treatment Effects

We now discuss the approach for estimating HTE using a smoothing method (Xie et al., 2012). The procedures can be illustrated as following steps.

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

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

(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 change against propensity scores. Specifically, individual treatment effects estimated in step *(2)* can be described as a function of the propensity score:

δ_{HTE}=f(P(**X**))

Where *P(X)* is the propensity score given observed control variables ***X***, *f(P(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 Control Variables Included in PSM

The validity of PSM largely relies on the CIA, $(Y(1), Y(0)) \perp T|\boldsymbol{X}$, where ***X*** is a vector of confounders. Only variables that affect both treatment participation and potential outcomes should be included in the propensity score model. Discussions are available regarding variables choice when using propensity score methods (Heckman et al., 1997; Rubin and Thomas, 1996; Brookhart et al., 2006; Bryson et al., 2002; Augurzky and Schmidt, 2000). The inclusion of control variables would be less complicated if criteria for treatment assignment were available. Where such criteria are not available, it is still possible to choose control variables based on previous empirical findings. In this study two sets of variables are chosen for inclusion in PSM.

4.1.1 Variables suggested in the handbook for camera site selection

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 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) 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.

A site is defined as a stretch of road that has a fixed speed camera or has had one in the past. A site length is the distance between two points within which collisions, casualties and speeds are measured and camera enforcement takes place. For fixed cameras the choice of a monitoring length for collisions was difficult as there has, until recently, been very little information available concerning the likely area of influence of cameras and there is no standard monitoring length. Different authorities use different site lengths although 500m either side of the camera has probably been most common (Gains et al., 2004). In this study, camera sites data is collected from eight English administrative districts, where different site lengths are selected by local authorities. Although site lengths vary slightly, they can be categorized into three bands: 200m, 500m and 1km. The segments are centered on the camera for most sites. If a segment meets a major junction or another segment, the segment lengths are not even on both sides, e.g. 200m upstream and 800m downstream.

The first three guidelines can be thought of as primary criteria and the latter two as secondary criteria (Gains et al., 2005). 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, because data concerning speeds are not routinely collected for road sections before they have been selected for further investigation and possible remedial treatment. Whether the exclusion of the speed information will lead to bias in estimates depends on the extent to which the speed variables play a part in site selection. The models used in this study were based on data for 771 cameras sites throughout the UK. These roads can be reasonably assumed to be representative of the typical speed distribution throughout the UK. National data suggests that, for typical 30 mph roads, speed distributions and percentages of speeding are very similar to those at the sites with speed cameras (Gains et al., 2004). It is probably because that speeding is endemic on 30mph roads and the speed criteria for site selection are not particularly restrictive. For example, in 1998, an average of 70% of cars on 30mph roads in GB exceeds the speed limit with a mean speed of 33 mph (Mountain et al., 2005). Since the speed criteria would be met on most 30mph roads, there is no reason to suppose that the speed related criteria played much part in selecting 30mph camera sites. This is also broadly the case for 40mph zones. 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 collision history, both of which are also important predictors in road safety analysis. Pre-treatment casualty records are valuable control variables 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 safety analysis. In addition, it is suggested that drivers may choose alternative routes to avoid speed cameras sites (Allsop, 2010; Mountain et al., 2004). Collision 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 effect due to traffic flow is controlled for in this study by including the annual average daily traffic (AADT) in the propensity score model.

4.1.2 Variables suggested as important factors affecting road casualties

Notwithstanding the guidelines discussed above there are sites not meeting the criteria which may still be treated as enforcement sites for one or more of the other reasons, such as community concern, collision frequency and engineering factors (DfT, 2005). 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 (1996) and Brookhart et al. (2006), unless there is consensus that the variable is unrelated to treatment assignment, variables that affect potential outcomes should be included, because they decrease the variance of the estimated treatment effect without increasing bias. Hence variables suggested as important factors when analysing road casualties at camera sites are also considered. Thus variables further included in the propensity score model are: speed limit, road class (e.g. Motorway, A road, B road, Minor road) 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 (Mountain et al., 2005; Gains et al., 2005; Christie et al., 2003). The data set used is Ordnance Survey (OS) Meridian ™ 2 for the period from 1999-2007, so the variance in road characteristics over time is controlled for.

4.2 Sample Size

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. All speed cameras are designed to measure the speed of approaching vehicles, or departing vehicles, or both, depending upon the type of camera. 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 collision migration (Mountain et al, 2005; Christie et al., 2003). To control for this effect, in this study the effects of speed camera are investigated in both directions no matter which type of the camera is. Control sites, which have never had a camera in the past and are at least 1.5km from any camera site, are randomly chosen in these districts to ensure that they are not affected by speed cameras. A total of 4787 sites without cameras covering all site length bands are used as potential control sites for matching. In the following section, a series of tests are used to check whether the control group can provide adequate matches.

The established dates for camera sites range from 2002 to 2004: 215 sites in 2002, 352 sites in 2003 and 204 sites in 2004. To ensure that three years data before and after are available for all camera sites, nine years data from 1999 to 2007 are used in this study. 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 Estimation of Propensity Scores

The first step is to estimate propensity scores for all treated and untreated units. Table 1 shows that all control variables except "minor road" are significant in the estimation of the propensity scores. This is probably because there are only 19 observations for speed cameras installed on minor roads in the study sample. "Motorway" is excluded due to collinearity. 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 described in the handbook (DfT, 2005). The result, in general, confirms that the control variables included in the propensity score model are important in predicting the possibility of being selected as camera sites. The estimation model shows a low Pseudo R^2 value. As Westreich et a*l. (*2011*)* emphasized, however, the primary purpose of the propensity score model is not to predict treatment assignment, but to balance control variables in order to control for confounding.

TABLE 1 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
AADT 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
Pseudo R Square: 0.28	AIC: 2145.4					
Observations: 5558						

5.2 Tests of Matching Quality

We first check the validity of the propensity score matching 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 control groups. The region of common support is defined as the area where the support of the propensity scores overlaps for the treated and control groups. Units within the region of common support is called "on support", and vice versa. Figure 1 shows the distribution of propensity scores for both groups. We observe 771 sites and 4787 sites for the treatment and the potential control groups respectively, with only seven treated sites are off support and discarded[1]. Table 2 is also provided as a supplement to Figure 1. It can be seen that camera and control sites have common support with scores below 0.65. For scores between 0.65-0.7, only control sites are observed, while camera sites with a score greater than 0.7 are off support. Therefore there is sufficient overlapping of the distributions between camera and potential control sites.

[1] Due to the small percentage of "off support" treated units (0.9%), the green columns are not clear. However, they can be seen if Figure 1 is sufficiently enlarged.

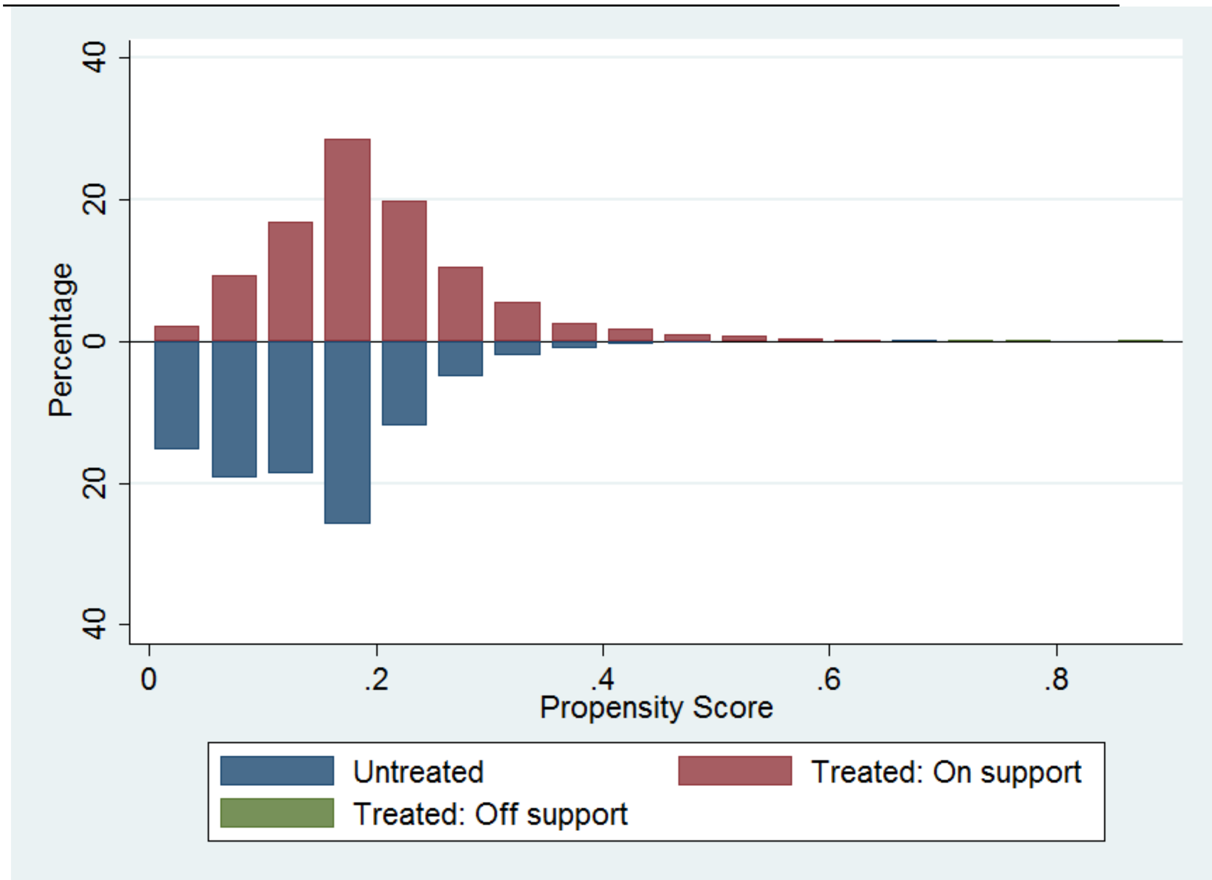

FIGURE 1 Propensity score distribution

TABLE 2 Summaries of Propensity Scores for Treated and Untreated Groups

Propensity Score	Treated Obs.	Mean	Min	Max	Untreated Obs.	Mean	Min	Max
0-0.05	20	0.036	0.014	0.049	718	0.024	0.007	0.049
0.05-0.1	71	0.078	0.052	0.099	930	0.074	0.050	0.099
0.1-0.15	123	0.130	0.101	0.149	914	0.128	0.100	0.149
0.15-0.2	216	0.173	0.150	0.199	1267	0.173	0.150	0.199
0.2-0.25	147	0.223	0.200	0.249	546	0.221	0.200	0.249
0.25-0.3	77	0.272	0.251	0.299	238	0.272	0.250	0.299
0.3-0.35	40	0.325	0.301	0.348	97	0.322	0.301	0.349
0.35-0.4	21	0.369	0.350	0.394	42	0.369	0.351	0.399
0.4-0.45	18	0.422	0.402	0.443	11	0.423	0.400	0.448
0.45-0.5	12	0.478	0.461	0.490	5	0.473	0.455	0.482
0.5-0.55	10	0.520	0.500	0.539	4	0.524	0.504	0.542
0.55-0.6	7	0.572	0.555	0.597	4	0.562	0.553	0.591
0.6--0.65	2	0.635	0.628	0.642	2	0.624	0.614	0.633
0.65-0.7	0	-------			9	0.682	0.672	0.691
0.7-0.75	2	0.741	0.736	0.746	0	-------		
0.75-0.8	2	0.769	0.759	0.778	0	-------		
0.8-0.85	1	0.815	0.815	0.815	0	-------		
0.85-0.9	2	0.880	0.862	0.897	0	-------		

A balancing test is performed to assess the matching quality as this test can verify that the treatment is independent of the control variables after matching. Table 3 shows the t-test of differences in variable means before and after the matching. It can be seen that all control variables are balanced between the treatment and matched control groups. Consequently the bias due to the differences in observable characteristics is reduced. In addition, the balance test can be also used to check the distribution of control variables between treatment and control groups. For example, Figure 2 shows the density plots of “PICs in baseline years” and “number of minor junctions” for treatment and control groups before and after matching, which indicate that the propensity score matching can balance not only the means but also the distribution of control variables.

TABLE 3 Checking the Variables Balance between Groups Before and After Using Nearest Neighbours (k=5) Matching

Variable	Sample	Mean Treated	Mean Control	%bias	%reduced |bias|	t-test t	t-test 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
AADT 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

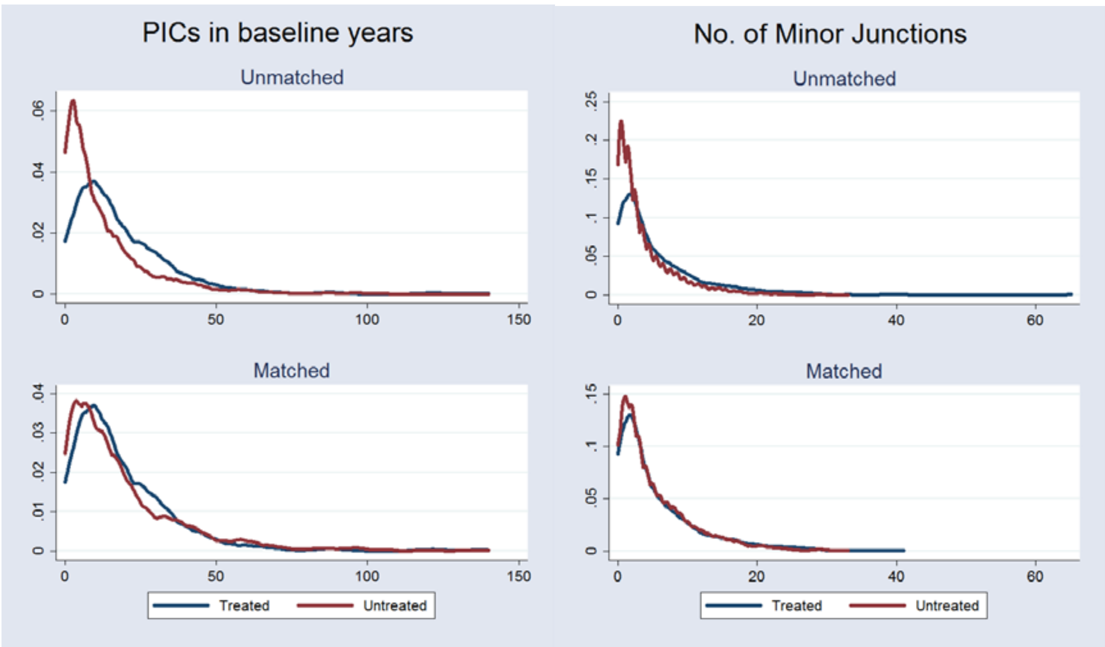

Figure 2 Density plots of "PICs in baseline years" and "number of minor junctions"

5.3 HTE of Speed Cameras on Road Casualties

The aim of this study is to evaluate how site characteristics influence the effects of fixed speed cameras. We estimate HTE of speed cameras on road casualties using smoothing methods in this section.

We first estimate average treatment effects using five different matching methods. In this study, the matching algorithms used are: K-nearest neighbours matching (K=1), K-nearest neighbours matching (K=5), radius matching (caliper=0.05), stratification matching and kernel matching (caliper=0.05). The effects of speed cameras are estimated as absolute numbers and percentages.

The next step is to estimate HTE. The treated and control units are first matched via kernel matching (caliper=0.05). The treatment effect for each treated unit is estimated by taking the difference 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 results from local linear and quadratic regressions are very similar. The first three graphs in Figure 3 show approximately U-shape curves of HTE on PICs and FSCs. It is worth noting, however, for camera sites with higher propensity scores, the pool of potential control sites is not as large as those with lower scores. This may influence the matching quality. In addition, the sample size of such camera sites is also small, only accounting for 5% of the population. This may also have an impact on the estimates. To get a clearer pattern of HTE, we thus exclude the sites with propensity scores higher than 0.32 (0.95 quantile) and re-examine the pattern of HTE as shown in the three graphs at the bottom of Figure 3.

We then compare ATE and HTE of speed cameras on PICs and FSCs. As shown in Table 4, 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. The reduction in PICs as percentages ranges from 10% to 40%, whilst the average treatment effect on PICs as percentage is 25.9%. In terms of 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. The average reduction in FSCs as percentages is about 10% using naive before-after method. However, this effect becomes insignificant after using matching methods.

Although the propensity score is a useful index in the sense that it simplifies matching, it is not available to decision makers. Furthermore, treatment decisions are 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.

Table 4 Average Effects of Speed Cameras on Annual PICs/FSCs per km

Matching Methods	Effects on annual PICs per km in absolute numbers					Effects on annual PICs per km in percentage				
	Changes (PICs/year/km)	S. E.	T-Stat	No. of Camera Sites	No. of Control Sites	Percentage Changes	S. E.	T-Stat	No. of Camera Sites	No. of Control Sites
Unmatched	-1.441	0.131	-11.02	771	4787	-30.70%	0.041	-7.5	726	4077
K-nearest Neighbours Matching (K=1)	-1.035	0.21	-4.92	764	663	-29.70%	0.051	-5.83	726	600
K-nearest Neighbours Matching (K=5)	-1.068	0.168	-6.33	764	2923	-24.60%	0.034	-7.21	726	2625
Radius Matching (Caliper=0.05)	-1.081	0.155	-6.97	769	4626	-25.20%	0.031	-7.99	724	3921
Stratification Matching	-1.042	0.15	-6.96	769	4628	-24.70%	0.029	-8.48	725	4078
Kernel Matching (Bandwidth=0.05)	-1.117	0.147	-7.61	771	4626	-25.10%	0.032	-7.89	726	4077
Average Effect	-1.068					-25.90%				

Matching Methods	Effects on annual FSCs per km in absolute numbers					Effects on annual FSCs per km in percentage				
	Changes (FSCs/year/km)	S. E.	T-Stat	No. of Camera Sites	No. of Control Sites	Percentage Changes	S. E.	T-Stat	No. of Camera Sites	No. of Control Sites
Unmatched	-0.342	0.037	-9.25	771	4787	-9.90%	0.043	-2.3	512	2435
K-nearest Neighbours Matching (K=1)	-0.141	0.06	-2.34	771	663	-5.20%	0.052	-0.99	512	409
K-nearest Neighbours Matching (K=5)	-0.124	0.049	-2.5	771	1676	-6.20%	0.047	-1.31	512	2435
Radius Matching (Caliper=0.05)	-0.131	0.046	-2.82	769	4626	-3.85%	0.043	-0.89	509	2435
Stratification Matching	-0.135	0.044	-3.05	769	4628	-3.76%	0.043	-0.87	509	2437
Kernel Matching (Bandwidth=0.05)	-0.131	0.046	-3.01	771	4626	-3.54%	0.046	-0.82	509	2435
Average Effect	-0.132					-4.51%				

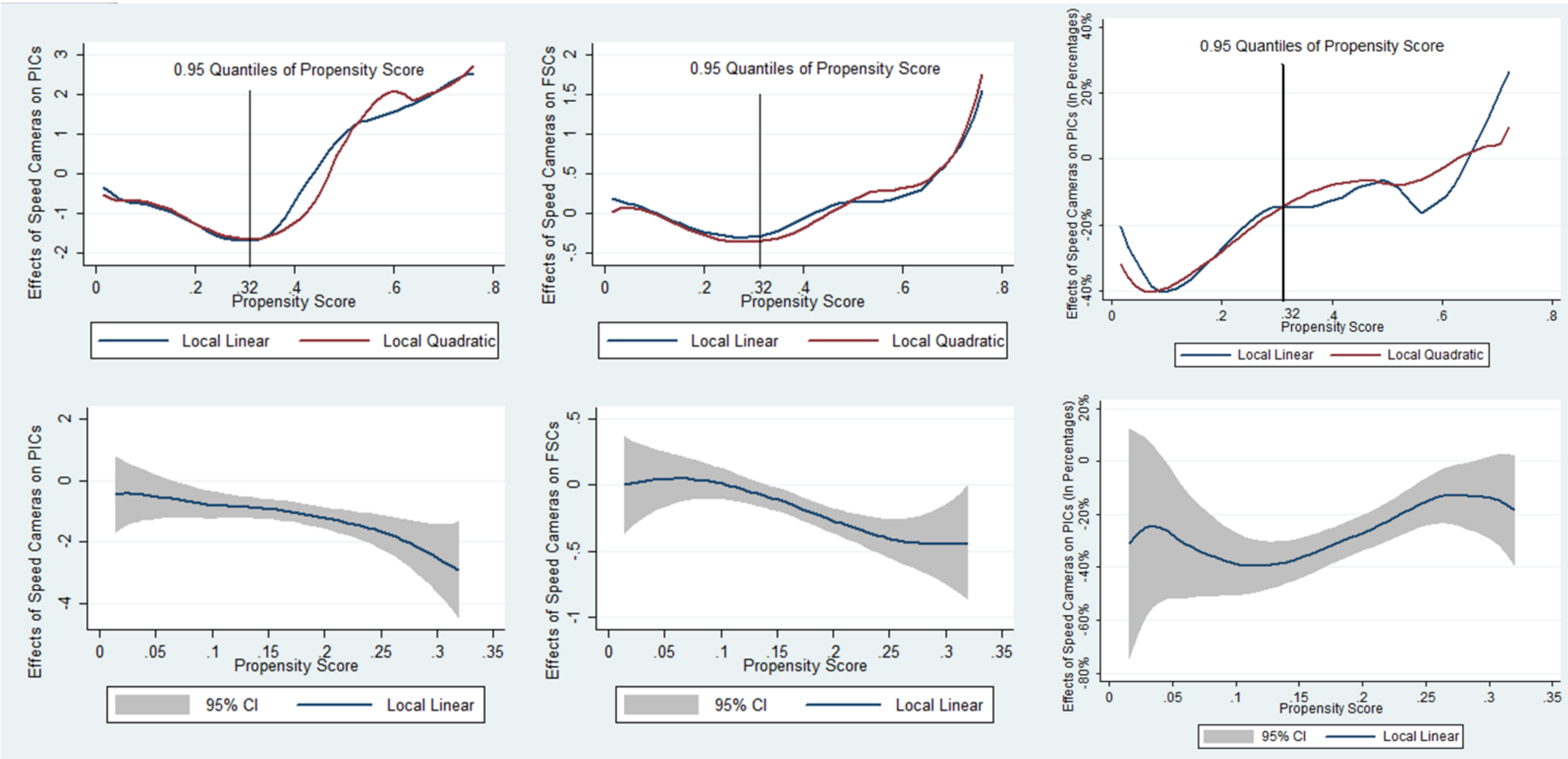

FIGURE 3 HTE of speed cameras by propensity scores

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 factor affecting camera sites selection is historical casualties (Gains et al., 2004; DfT, 2005). It is worth noting that the rules for proposed fixed speed camera sites are slightly different in the handbook published by DfT (2005). Most rules are consistent with the criteria described in the four-year report (Gains et al., 2004). 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" (DfT, 2005). 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 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 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 possible 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 included in the sample. The idea is to compare the treatment effects between these two groups of treated sites. If the treatment is more effective at the sites not meeting the criteria, then we may conclude that the criteria for selecting camera sites are not optimized. 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 4, 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 4 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, the third column of Figure 4 shows the distribution of treatment effects for camera sites meeting and not meeting the criterion (risk values>=22). The horizontal axis indicates the effects of speed cameras on collisions, while the percentages are shown on the vertical axis. It can be seen that the pattern for camera sites with risk values greater than 22 is more skewed left, indicating a better performance 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.

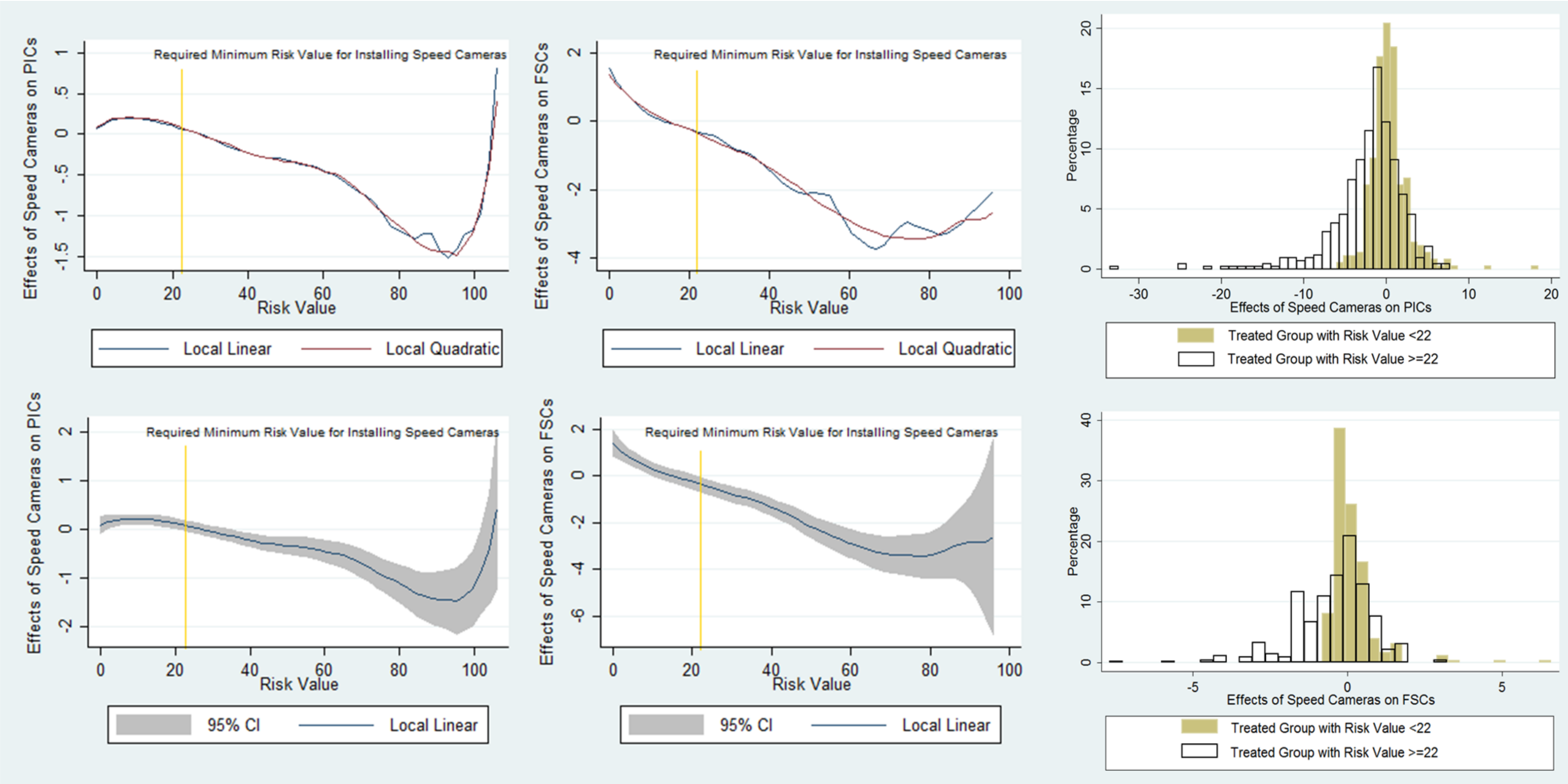

FIGURE 4 HTE on PICs and FSCs by risk values

6. DISCUSSIONS AND CONCLUSIONS

As an important part of the UK Government's ten year road safety strategy, the national safety camera programme is expected to regulate driving speeds and reduce casualties. It would be very rash to conclude that safety cameras are uniformly effective in reducing road casualties. However, a series of issues related to speed cameras have never been touched before, such as how effectiveness varies by site characteristics and under what conditions speed cameras perform most effectively. To date there has been no independent study using advanced state-of-the-art causal methodologies to answer these questions. This study contributes to the literature by applying causal models to estimate heterogeneous treatment effects of speed cameras on road safety.

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. It shows that the reduction in personal injury collisions ranges from 10% to 40% whilst the average effect is 25.9%, indicating that the effects of speed cameras are not uniform across camera sites and are dependent on site characteristics, as measured by propensity scores. Hence a treatment decision based on the average treatment effect for the entire population can be misleading. Furthermore, since speed cameras are usually costly, it is desirable that the programme is operated in a way that maximizes effectiveness with limited resources. By revealing patterns of HTE, policy makers can install speed cameras at sites most likely to benefit from the treatment, so as to improve the cost-effectiveness of the programme. Although the cost-effectiveness is not evaluated, the methodology introduced and estimation results presented in this paper have laid a foundation for future study.

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 effects on road casualties due to the introduction of speed cameras in the UK. This paper introduced the PSM method to account for these two issues. The results show that the characteristics of the treatment and control groups are well balanced after matching. Therefore, the authors suggest that the selection of such a control 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.

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. The distributions of camera effects between these two groups are compared. In general, it is found that camera sites which meet the criterion perform better in reducing casualties. Only 57 percent of the treated sites in the sample, however, meet the requirement for risk values. As we discussed earlier, although 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. What remains unclear is that whether the current criteria are optimized, e.g. the most cost-effective. Although this is beyond the scope of this study, it could be an interesting topic for future research.

Acknowledgment

This work was supported by the Funds for International Cooperation and Exchange of the National Natural Science Foundation of China (Grant No.5151101143), and the Natural Science Foundation of Jiangsu Province (Project No.BK20150615).

References

1. Abadie, A., and G. Imbens. Large Sample Properties of Matching Estimators for Average

Treatment Effects. Working Paper, Harvard University, 2004.

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

3. Allsop, R., 2010. The Effectiveness of Speed Cameras- A review of evidence. RAC Foundation.

4. Augurzky, B. and C. M. Schmidt. The Propensity Score – A Means to an End. IZA Discussion paper, No. 271, Bonn, 2000.

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

6. Brand, J. E., and D. Davis. The Impact of College Education on Fertility: Evidence for Heterogeneous Effects. Demography, 48, 2011, pp. 863-887.

7. Brookhart, M. A., S. Schneeweiss, and K. J. Rothman, et al. Variable selection for propensity score models. Am J Epidemiol, 163(12), 2006, pp.1149-1156.

8. 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.

9. 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.

10. 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.

11. 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.

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

13. Department for Transport. Handbook of Rules and Guidance for the National Safety Camera Programme for England and Wales for 2006/07. 2005.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. Holland, P. Statistics and Causal Inference (with discussion). Journal of the American Statistical Association, Vol. 81, 1986, pp. 945-970.

20. 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.

21. 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.

22. 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.

23. 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).

24. 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.

25. 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.

26. 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.

27. Rubin, D. B., and N. Thomas. Matching Using Estimated Propensity Scores: Relating Theory to Practice. Biometrics, 52, 1996, pp. 249-264.

28. 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.

29. 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.

30. Smith, H. Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies," Sociological Methodology, 27, 1997, pp. 325-353.

31. Westreich, D., S. R. Cole, and M. J. Funk, et al. The role of the c-statistic in variable selection for propensity score models. Pharmacoepidemiol Drug Saf. 20(3), 2011, pp. 317-320.

32. Xie, Y., J. E. Brand, and B. Jann. Estimating Heterogeneous Treatment Effects with Observational Data. Sociological Methodology, 42, 2012, pp. 314-347.

Appendix

Glossary

AADT: Annual Average Daily Traffic

ATE: Average Treatment Effect

CIA: Conditional Independence Assumption

FSC: Fatal and Serious Collisions

HTE: Heterogeneous Treatment Effect

PIC: Personal Injury Collisions

PSM: Propensity Score Matching

RTM: Regression to Mean