Quantifying the causal effects of 20 mph zones on road casualties in London via doubly robust estimation

Haojie Li[a,*], Daniel J. Graham[b]

[a] *School of Transportation, Southeast Univeristy, China*
[b] *Centre for Transport Studies, Imperial College London, UK*

Abstract

This paper estimates the causal effect of 20 mph zones on road casualties in London. Potential confounders in the key relationship of interest are included within outcome regression and propensity score models, and the models are then combined to form a doubly robust estimator. A total of 234 treated zones and 2844 potential control zones are included in the data sample. The propensity score model is used to select a viable control group which has common support in the covariate distributions. We compare the doubly robust estimates with those obtained using three other methods: inverse probability weighting, regression adjustment, and propensity score matching. The results indicate that 20 mph zones have had a significant causal impact on road casualty reduction in both absolute and proportional terms.

*Corresponding author
Email address: `h.li@seu.edu.cn` (Haojie Li)

Keywords: 20 mph zones, Doubly Robust Methods, Causal Analysis

1. Introduction

It is widely thought that a reduction in vehicle speeds can reduce the severity of road casualties and decrease the number of traffic collisions (Soole et al., 2013; Elvik et al., 2004; Elvik, 2009). There are a number of policy interventions that can be used by governments to reduce traffic speeds in the hope of improving road safety. An example of such measurse is designation of 20 mph zones, which are widely applied in the UK particularly in residential areas.

While several studies have been undertaken to analyze the impact of 20 mph zones on various outcome of interest, there remains uncertainty regarding the causal effects of 20 mph zones on road safety. A major challenge for evaluation lies in constructing viable counterfactual outcomes that can represent what would have happened to "treated" units in the absence of the treatment (i.e. designation of 20mph status). Since counterfactual outcomes cannot be observed, regression-based statistical models are usually used to model them, particularly via before-after and time-series methods (e.g. Webster and Layfield, 2003; Grundy et al., 2009). The validity of such methods relies on their ability to control for confounders, which are a set of risk factors for the outcome of interest that are also correlated with treatment assignment. The estimator of treatment effects is consistent and unbiased only if the confounders are properly accounted for. This critical issue, however, is inadequately justified in previous studies.

This research contributes to the literature by tackling the issue of confounding using a doubly robust (DR) estimator and subsequently uses this method to evaluate the effect of 20 mph zones on road casualties in London. The DR approach combines outcome regression (OR) and propensity score (PS) models to obtain an estimator which is consistent and asymptotically unbiased so long as at least one of the component models (i.e. OR or PS) is correctly specified. It thus provides two opportunities for valid treatment effect estimates which is useful in situations when the quality of data or knowledge about the underlying processes is not uniform. The DR method has been used routinely to estimate causal treatment effects in other areas of science such as medicine and epidemiology, but, to the best of our knowledge, has not been applied previously in road traffic safety research.

Another key contribution of our paper lies in development of a panel data sets to capture variance in road network characteristics over time. A limitation of previous research on this topic is that road network effects have been assumed static which could lead to biased treatment effect estimates if such characteristics operate as confounders.

This paper is organized as follows. Section 2 reviews previous literature in the field. Methods are described in Section 3 and Section 4. Our results are presented and discussed in Section 5. Conclusions are then drawn in the final section.

2. Literature Review

A wealth of empirical evidence shows a clear relationship between traffic collisions and vehicle speeds. In particular, mean vehicle speeds are found to be positively related with the number and severity of traffic collisions (Elvik et al., 2004; Elvik, 2009). Speed limits specify maximum desirable traffic speeds and these can be used to reduce the number of road traffic casualties. An example of such a measure is traffic calming, which is especially prevalent in residential areas.

Numerous studies have been conducted to evaluate the safety impacts of traffic calming. A meta-analysis by Elvik (2001) investigates the effects on road safety of area-wide urban traffic calming schemes from 33 studies, including research reports from Norway, Sweden, Finland, Denmark, Germany, the Netherlands, Great Britain, France, the United States and Australia. The results show that area-wide urban traffic calming schemes reduce the number of injury accidents by about 15% on average, whilst a 25% reduction in the number of accidents is found for residential streets. Another meta-analysis by Bunn et al. (2003) reviews 16 controlled before-after trials of area-wide traffic calming mainly in high income countries. Their review results also suggest that traffic calming can be effective in reducing the number of traffic crashes. However, previous studies reviewed in these meta-analyses tend to use before-after methods with some defined comparison group, which is not able to fully control for confounding effects, such as selection bias, also known as the regression to mean.

A number of studies have examined the impact of traffic calming in te UK, including 20 mph zones, on road safety, traffic speeds, environmental and health outcomes, amenity, traffic volumes, and inequality (Casanova and Fonseca, 2012; Grundy et al., 2009; Steinbach et al., 2011; Tovar and Kilbane-Dawe, 2013; Webster and Mackie, 1996; Webster and Layfield, 2003; Williams and North, 2013). Webster and Layfield (2003) investigate 78 20mph zones in London applying before-after methods. Allowing for background changes, total and KSI casualties are found to be reduced by 45% and 57% respectively. Grundy et al. (2009) conduct a time series analysis using data of 399 20mph zones in London from 1986 to 2006. Time trend effects are taken into account by using conditional fixed effects Poisson models. The authors also suggest that the RTM effect can be controlled for by dropping data for three, four or five years prior to the implementation of the 20 mph zones.

There are two key issues that have not been fully addressed in previous evaluation studies on the impacts of 20 mph zones. First, the methods used in previous work are mainly before-after control studies. Usually, a control group is employed to estimate the counterfactual outcomes of the treatment group. Ideally control groups should have the same or similar characteristics to those of the treatment group, i.e. the control group must be representative of the treated sites. However, in previous research, insufficient attention has been paid to selection of such control groups. For example, Webster and Layfield (2003) use all unclassified roads in London as control data for roads in 20 mph zones. However, due to selection bias, the characteristics of treated and "control" units defined in this way may differ.

Second, a fundamental assumption required to draw valid causal inference from observational data is that all confounders are measured and represented adequately. Previous studies on 20 mph zones have largely ignored the potential for road casualties to be associated with the road network characteristics. Yet we know from the literature road casualties are significantly associated with road network characteristics, such as road class, road density and the number of nodes, the connectivity and accessibility of the road network, and the curvature of the road network (e.g. Huang et al., 2010; Marshall and Garrick, 2011; Rifaat et al., 2011; Jones et al., 2008; Quddus, 2008). The failure to account for the effects due to road network characteristics in evaluating traffic calming measures can bias estimates of the safety impacts of 20 mph zones. In this paper we develop a detailed panel data set on road network design to address potential confounding from this source.

The doubly robust estimator, originally proposed by Robins et al. (1995), has been described in the statistical literature (Bang and Robins, 2005; Robins et al., 1995; Robins, 1999; Lunceford and Davidian, 2004), and applied extensively in various areas of science. However, it has not yet been used for road safety research although in our view it has great potential.

3. Methods

The DR estimator combines PS and OR models developed using insights from the potential outcomes framework for causal inference. In this section we first introduce the potential outcomes framework and draw attention to its relevant assumptions. We then discuss how a doubly robust estimator of causal effects can be obtained by combining outcome regression and propensity score models.

3.1. Potential outcome framework

In presenting the potential outcome framework, it is necessary to introduce relevant notation. D_i is an indicator of treatment enrolment for individual or unit i. To facilitate understanding, consider only binary treatments. $D_i = 1$, if unit i received the treatment, and 0 otherwise. Let $Y_i(D_i)$ be the potential outcomes for unit i. Therefore, $Y_i(0)$ denotes the level of outcome that unit i would attain if not exposed to the treatment. Likewise, $Y_i(1)$ denotes the level of outcome that unit i would attain if exposed to the treatment. The individual causal treatment effect for unit i can be defined as $\delta_i = Y_i(1) - Y_i(0)$ (Individual Treatment Effect). The fundamental problem of causal inference is that since unit i can be either treated or not, we can only observe one of these two potential outcomes. If unit i is subject to the treatment then $Y_i(1)$ will be realized and $Y_i(0)$ will be an unobservable counterfactual outcome and vice versa.

In simple control studies, such as those described in the literature review above, the average treatment effect on the treated (ATE), $E[Y(1) - Y(0)|D = 1]$, is estimated by taking comparisons of the average outcomes between treated and control units, which can be defined as:

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

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

In the above equation, the term in curly brackets is not zero for most cases due to selection bias, i.e. the treatment assignment is usually associated with the potential outcomes that individuals could attain, with or without being exposed to the treatment.

In randomized experiments, the probability of assignment to treatment does not depend on potential outcomes. That is,

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

Then $E[Y(0)|D = 1] = E[Y(0)|D = 0]$

and therefore

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

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

$= E[Y(1) - Y(0)|D = 1]$ (ATE with randomized assignment) (2)

Equation *(2)* provides an unbiased estimator of ATE. Randomized experiments are straightforward and allow the greatest reliability and validity of statistical estimates of causal effects. Whilst they are a valuable tool for treatment evaluation, it is not always feasible to implement a randomized experiment due to high costs and ethical issues. Consequently, causal analysis with observational data uses models to approximate randomized distinctions. There are two critical assumptions underpinning such studies.

3.2. Assumptions

3.2.1. Unconfoundedness

The validity of causal inferences from observational data crucially relies on the assumption of unconfoundedness. The unconfoundedness assumptions, also known as conditional independence (CIA), 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. The CIA creates a selection process analogous to that of randomized experiments. More generally, the distribution of the counterfactual outcomes for treated and untreated groups are the same. In these circumstances it is possible to infer the counterfactual outcomes and the treatment effect can be estimated by the differences between treatment and control groups. The unconfoundedness can be described as:

$(Y(1), Y(0)) \perp D \mid X, \forall X$

The unconfoundedness assumes that all relevant confounders are observed. This assumption is crucial to making causal inferences in observational studies, but is untestable in practice. The unconfoundednesse assumption is too strong and may not hold when unobserved factors that may influence outcomes are not included in the model. However, this assumption can be relaxed by using the difference-in-difference

(DID) estimator (Heckman et al., 1997). In the DID approach, the dependent variable is the difference between outcomes in pre-intervention and post-intervention periods. Given data from the pre-treatment period, any time-invariant confounder can be controlled for. In addition, the likelihood of satisfying this assumptions can be strengthened by capturing as much information as possible about potential confounders.

3.2.2. Common support

For valid treatment effect estimation it is also required that both treated and untreated units have overlap in the support of the covariate distributions. This is known as the positivity or the overlap condition (Cole and Hernan, 2008). Also, either extremely high or low values of propensity scores can cause problems when inverse weighting by creating large weighted outcome values (Kurth et al., 2006; Emsley, 2007). Similar to the test used in matching approaches, a histogram showing the distribution of propensity scores for both groups can help identify the positivity and avoid the extreme values problem.

3.3. Doubly robust estimation

The DR estimator is described below in the case of binary treatments for Frequentist inference. For a Bayesian treatment of the DR estimator please see

Graham et al. (2015).

3.3.1. Outcome regression

The potential outcome framework can be written in terms of a simple linear regression model:

$Y_i(D) = \alpha + \delta D + \varepsilon_i(D)$ (3)

Where $\varepsilon_i(D)$ is the potential outcomes error term. Linear functions are used for notational simplicity. Other functional forms, such as Poisson, can be used in practice. In this analysis, the model is specified in the DID form to eliminate the influence of time-invariant characteristics. Hence $Y_i(D)$ is defined as the difference between outcomes in pre-intervention and post-intervention periods.

With observational data, the estimator of the average causal effect, δ', can be described as:

$\delta' = E(Y_i|D = 1) - E(Y_i|D = 0)$

$= \delta + E[\varepsilon_i(0)|D = 1] - E[\varepsilon_i(0)|D = 0] + E[\varepsilon_i(1) - \varepsilon_i(0)|D = 1]$ (4)

This estimator of the treatment effect will be biased due to the dependence between the treatment assignment D and the error term ε_i. If all potential confounders $\boldsymbol{X}$ are observed and correctly specified in the regression model, the treatment assignment D is independent of the error term ε_i, $D \perp \varepsilon_i \mid \boldsymbol{X}$. The proper specification of the model, however, can be difficult when multiple potential confounders exist. The propensity score can be used as a single covariate and methods based on the PS, e.g. inverse probability weighting, can be applied in causal analysis.

3.3.2. Inverse probability weighting

Different from the outcome regression methods, the inverse probability weighting (IPW) controls for confounding by using a single index, the propensity score. It is the probability that a unit is selected into the treatment group conditional on observed covariates. The first step when implementing the IPW is to estimate the propensity score. For a binary treatment variable, logit and probit models are usually preferred to a linear probability model, which may generate predictions outside the [0, 1] bounds of probabilities. Logit and probit models usually yield similar results, hence the choice between them is not critical (see further discussion of this point in Smith, 1997). In this paper, a logit model is used:

$P(T = 1 \mid X) = \frac{EXP(\alpha+\beta' X)}{1+EXP(\alpha+\beta' X)}$ (5)

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

The estimator of propensity score, *P'*, can be predicted based on the estimated parameters and observed covariates for both treated and control individuals. Besides matching, another way of using PS to control for confounding is to weight the observed data. The IPW is defined as the inverse of the conditional probability of an individual's actual treatment status.

In observational data the sample is not randomized, but rather one in which individuals from certain subpopulations are over- or under-sampled. The idea is that weighting by the IPW estimator creates a pseudo population in which the distributions of confounders among the treated and untreated are the same as the overall distribution of those in the original total population (Sturmer et al., 2006). This indicates that the potential outcomes are independent of the treatment, which is consistent with the unconfoundedness assumption. The IPW is *1/P'* for the treated and *1/(1-P')* for the untreated. The IPW estimator of the ATE can be modelled as (Lunceford and Davidian, 2004):

$\delta_{IPW} = N^{-1} \sum_i^N (\frac{D_i Y_i}{P'_i}) - N^{-1} \sum_i^N (\frac{(1-D_i)Y_i}{1-P'_i})$ (6)

Similarly, the IPW estimator can be biased if the model for calculating the PS is misspecified.

3.3.3. Doubly robust estimator

The doubly robust methods proposed by Robins et al. (1995) combine the outcome regression and inverse probability weighting in one single model. The DR estimator can be expressed as the following equation:

$\delta_{DR} = N^{-1} \sum_i^N [\frac{D_i Y_i}{P'_i} - \frac{(D_i - P'_i)Y'_{i,D=1}}{P'_i}] - N^{-1} \sum_i^N [\frac{(1-D_i)Y_i}{1-P'_i} - \frac{(D_i - P'_i)Y'_{i,D=0}}{1-P'_i}]$ (7)

Where $Y'_i = E(Y \mid D, X)$ is the predicted value from the outcome regression model given *D=0,1* and the baseline covariates ***X***. The two average terms are estimates of the mean potential outcomes, $Y_{X=1}$ and $Y_{X=0}$, if everyone were to be treated and untreated. As a consequence, the difference in means is the effect due to the treatment.

In *equation (7)*, the first terms in each average are the IPW estimators for $E(Y_{X=1})$ and $E(Y_{X=0})$ respectively. The second terms are called augmentations (Funk et al., 2011) as this component is formed by taking the product of two bias terms: one from the PS model and one from the outcome regression model. If either bias term equals zero, then it excludes the other non-zero bias term from the incorrect model. That is the DR estimator will be consistent for the true average treatment effect, if either model is correctly specified. (For more detailed demonstration of the DR property, please refer to the work by Lunceford and Davidian, 2004).

The standard error can be obtained by bootstrapping the whole sequence of regressions, including the estimation of the propensity score. This can be realized in the STATA package *dr* (Emsley et al., 2008). Figure 1 shows the diagram of applying the DR methods to the estimation of treatment effects.

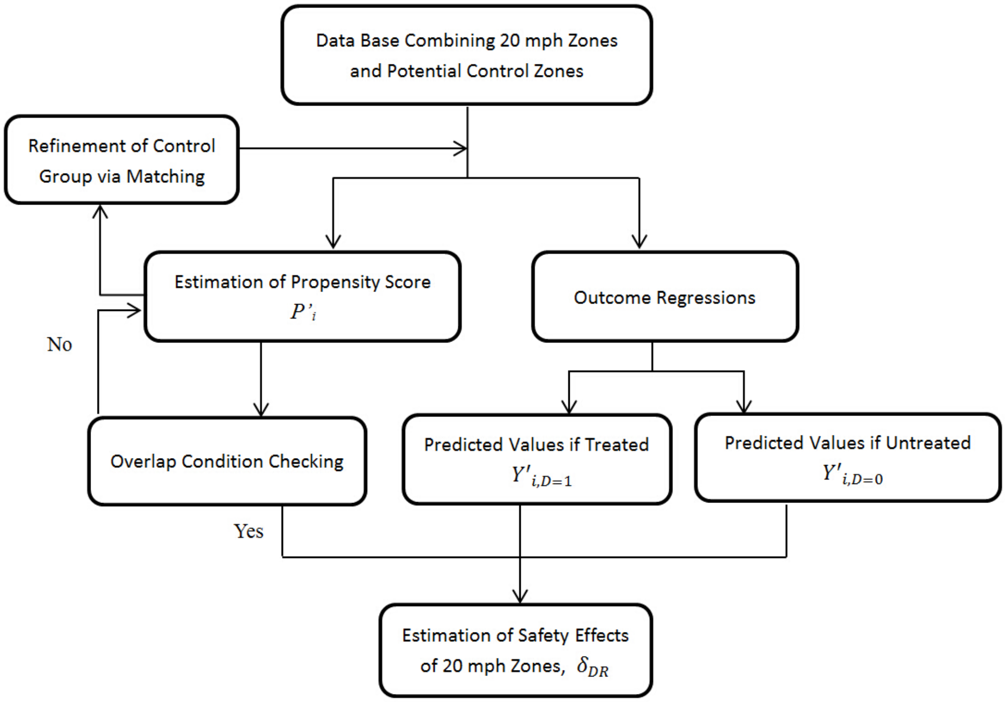

Figure 1: The diagram of the application of the doubly robust method to the evaluation of safety effects of 20 mph zones

4. Data

4.1. Confounders

The validity of the DR methods heavily relies on the "no unmeasured confounders" assumption, which is unfortunately untestable. However, its influence can be lessened by capturing as much information about potential confounders as possible.

Theoretically, covariates that affect the treatment assignment and potential outcomes should be included in the models. In practice, however, selection of such covariates can be complex due to the lack of precise knowledge of the relations among outcomes, treatment and confounders.

Although including additional covariates can increase the precision of the DR estimator (Lunceford and Davidian, 2004), this could generate problems with the common support (Bryson et al., 2002). And although the inclusion of non-significant covariates will not affect the unbiasedness and consistency of the estimates, it can reduce their efficiency, especially with small samples (Augurzky and Schmidt, 2000).

It is also suggested that omitting important covariates can cause serious bias in estimation (Heckman et al., 1997). Rubin and Thomas (1996) 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. (1996) illustrates how the choice of variables included in the propensity score model 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 DR*

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 the DR models.

4.2.1. *Covariates suggested as criterion for 20 mph zones selection*

Although the requirements for 20 mph zones have been prescribed in a number of legislation and regulations, the criterion of selecting 20 mph zones remains unclear. The most relevant document is setting local speed limits by DfT (2013), which provides guidance to highway authorities and local traffic authorities who are considering setting local speed limits, including 20 mph zones. The key factors that should be taken into account in any decisions on local speed limits are shown below:

- History of collisions, including frequency, severity, types and causes;
- Road geometry and engineering (e.g. bends, junctions);
- Road function;

- Presence of vulnerable road users;
- Existing traffic speeds.

According to another report by Steer Davies Gleave (2014), there is considerable variability as to the implementation of 20 mph zones in different authorities. However, most boroughs prioritize areas as 20 mph zones based on collision history, resident requests, and in some cases the presence of schools. Selection of 20 mph zones, therefore, is primarily based on accident history. Pre-treatment accident records are valuable covariates when estimating the DR estimator because they are important predictors of treatment entry and potential outcomes in post-treatment period. The accident data was collected from the STATS 19 data base and was further classified by severity type. The location of an accident was recorded using the British National Grid coordinate system. Each individual accident was located on the map using Geographical Information System (GIS) software, such as MapInfo and Arcmap.

Existing traffic speeds, such as the 85^{th} percentile speed and percentages of vehicles over the speed limit are not normally publicly available for all sites on UK roads, however. We address this issue by randomly selecting a large sample of potential control zones within London area. In doing so, it is expected that both treated and untreated zones are observed at every level of pre-treatment traffic speeds, so the overlap condition is met.

To account for the impacts of the presence of vulnerable road users on the decision to implement 20 mph zones, the Index of Multiple Deprivation (IMD) is obtained from the office for the Deputy Prime Minister. The Index of Multiple Deprivation integrates data on the following seven deprivation domain indices into one overall deprivation score: income, employment, housing and services, health, education, crime and environment.

4.2.2. Covariates suggested as important factors affecting road casualties

Notwithstanding the covariates discussed above there are areas not meeting the criteria (e.g. the collision history) which may still be selected as 20 mph zones for one or more of the other reasons, such as community concern and engineering factors (DfT, 2013). In other words, there are unknown factors that affect treatment assignment but are not explicitly described in the criterion for 20 mph zones selection. As suggested by Rubin and Thomas (1996) and Brookhart et al. (2006), unless there is consensus that the covariate is unrelated to treatment participation, covariates that affect the outcome should be included in the model, because they decrease the variance of the estimated treatment effect without increasing bias. Hence covariates suggested as important factors for analyzing road casualties are also considered.

One constraint in previous research on 20 mph zones is that no longitudinal or panel data of road network characteristics has been employed. The statistical relationship between road casualties and the characteristics of a road network has been investigated in the literature as described above, showing in general statistically significant effects. In this study, information regarding the road network was obtained from Ordnance Survey (OS) Meridian, which is a vector map dataset of Great Britain at a scale of 1:50000. This dataset is updated annualy and is collected for study period excluding for 2005 due to data availability. A set of variables is extracted from Meridian data set to describe the characteristics of the road network at zone level.

- Traditional road network characteristics. The length, as well as the density, of the road network is calculated according to road class, e.g. A road, B road, Minor road. Road network nodes are defined as meeting points of two or more roads. The total number and density of nodes is also calculated.

- Connectivity and accessibility of the road network. It has been suggested that the degree of connectivity and accessibility of a road network can influence the number of crashes (Marshall and Garrick, 2011). The measure used in this study is the link-to-node ratio, which is calculated by dividing the number of links by the number of nodes. A high link-to-node value indicates a more connected road network than one with a low link-to-node value. A node with only one link, also known as a dead end, is usually associated with a residential area. The density of dead ends is used in this study as a measure of the accessibility of a network.

- Curvature of the road network. Road curvature has been suggested as an important factor influencing road casualties (Jones et al., 2008; Quddus, 2008). The literature indicates that straighter roads have more crashes than roads with more bends. The variable used in this research to measure curvature is the number of vertices per km. The number of vertices are obtained using ArcGIS and divided by the road length in each zone.

Previous research has also suggested an association between road traffic crashes and socio-demographic characteristics, such as employment, deprivation and land use (Wier et al., 2009; Dissanayake et al., 2009; Graham and Stephens, 2008). In particular, a positive relationship has been found in relation to the size of the population and the level of employment, which implies that more casualties may occur in areas with more residents and job opportunities. To consider this effect, the data for population and employment, as well as the information of land use was obtained from the Office for National Statistics (ONS). In summary, the covariates that we included in the DR model are shown in Table 1. All of these covariates are included in both outcome regression and propensity score models.

Table 1: Covariates included in the doubly robust model

Covariates	Description
KSI (baseline)	Killed and seriously injured casualties in three years before the intervention
Slightly injured (baseline)	Slightly injured casualties in three years before the intervention
A roads (%)	Percentage of A roads
B roads (%)	Percentage of B roads
Minor roads (%)	Percentage of minor roads
IMD	The index of multiple deprivation
Domestic (%)	Percentage of domestic buildings, e.g. residential area
Non-domestic (%)	Percentage of non-domestic buildings, e.g. business and office district area
Green space (%)	Percentage of green spaces and gardens
Population density	Residential population per m^2
Employment density	Number of employees per m^2
Ratio of Emp to Non-Emp	Ratio of employment to non-employment
Density of dead ends	Ratio of nodes with only one link to all nodes
Links per node	Ratio of road links to nodes
Vertices density	Number of horizontal vertices per km

4.3. Sample size

Figure 2 shows the map of 20 mph zone and control zones in London. The 20 mph zones in the dataset cover a large period of time of 1989-2007. Due to data restrictions, only 20 mph zones established between 2002 and 2007 are included in the treatment group. Besides 234 20mph zones, a total of 2844 potential control zones were selected randomly within London area. It is possible that the implementation of 20 mph zones may have impacts on neighboring zones, so zones within 150 meters of each 20 mph zone are not included in the potential control group (Grundy et al., 2009). To ensure that three years data before and after are available for all 20 mph zones. The STATS 19 data used for this analysis includes road accidents in the UK from 1999 to 2010.

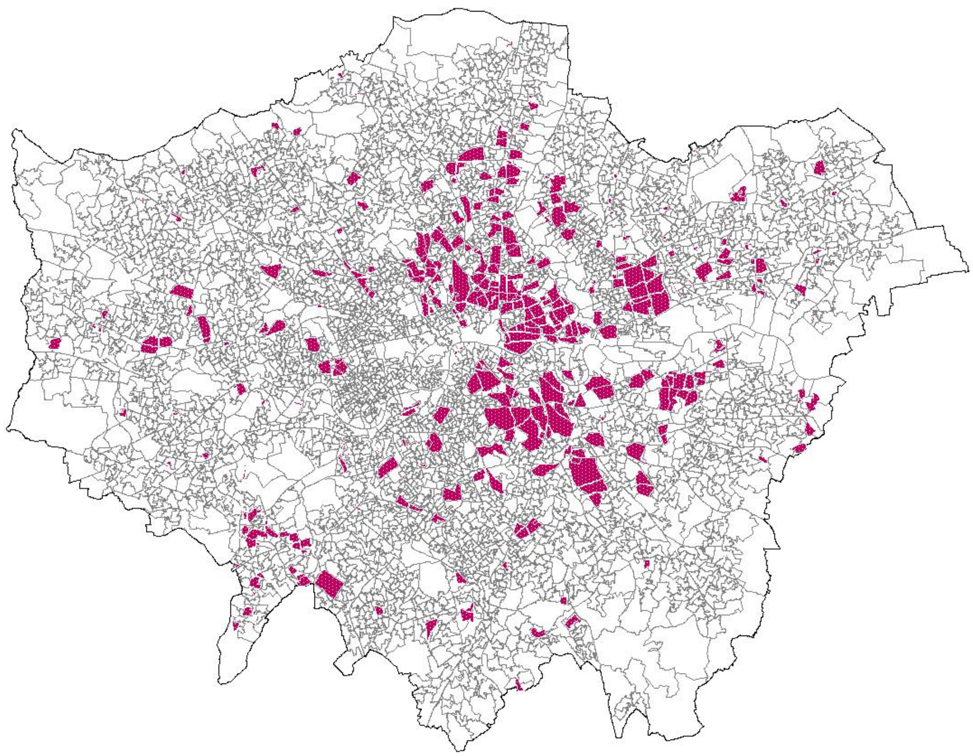

Figure 2: Map of 20 mph zones in London with Lower Layer Super Output Areas Boundary

5. Results

In this section, the safety effects of 20 mph zones are investigated using the DR method. The DR method combines two separate models: the propensity score model and outcome regression model. As discussed earlier, the optimal practice is to use the same set of covariates in both the propensity score and outcome regression models. The regression results from both models are presented, followed by the estimation of 20 mph zones effects using the DR method.

5.1. Propensity score estimation

The first step in the doubly robust method is to estimate the propensity scores. The logit and probit models are usually used in the PS model and give similar results. In this study, the logit model is preferred due to a higher BIC value. The estimation model shows a low Pseudo R-squared value. As Westreich et al. (2011) emphasized, however, the primary purpose of PS model is not to predict treatment assignment, but to balance covariates in order to control for confounding. Previous studies (Brookhart et al, 2006; Myers et al., 2011; Austin, 2009) have also shown that better predictive performance does not improve the balance of risk factors for the outcome. It is recommended to use measures of covariate balance to evaluate PSM models (Austin, 2009; McCaffrey et al., 2004).

The logit model is regressed on the covariates, which could influence both the treatment assignment and the potential outcomes. The covariate "Minor roads (%)" is dropped due to multicollinearity. Table 2 shows that most covariates are significantly related to the treatment assignment, indicating that they are important in predicting the possibility of being treated. Specifically, 20 mph zones are more likely to be implemented in areas with a higher historical record of slightly injured casualties, which is consistent with the guidelines by DfT (2013). However, the relation between the number of KSI in baseline years and the propensity score is not significant. This indicates that slightly injured casualty is more predominant when making decisions on 20 mph zones. In addition, deprived areas have substantially more 20 mph zones, which is consistent with previous findings (Rodgers et al., 2010). In terms of land use, the propensity score is found to be negatively related to the percentages of non-domestic buildings and green space. In addition, areas with higher density of residential population and lower density of employees have higher propensity of being selected as 20 mph zones. Regarding the characteristics of road network design, only the covariate links per node has a significant impact on treatment assignment, which is not surprising, because they are assumed to have more impacts on road casualties.

As discussed in the previous section, a total of 2844 potential control zones were selected randomly within the London area. However, the characteristics of treated and potential control zones may differ in the absence of any treatment. Only untreated zones with similar characteristics to those treated can be used to approximate the counterfactual outcomes of the 20 mph zones. So before proceeding to the doubly robust estimation, the control group need to be refined via matching, which can improve the balance of characteristics between treated

Table 2: Propensity score model

	Coef.	Std. Err.	z	$P > \|z\|$
KSI (baseline)	—			
Slightly injured (baseline)	0.007	0.002	3.19	0.001
A roads (%)	0.884	0.288	3.07	0.002
B roads (%)	1.148	0.391	2.94	0.003
IMD	0.015	0.004	4.05	< 0.001
Domestic (%)	—			
Non-domestic (%)	-4.581	1.536	-2.98	0.003
Green space (%)	-0.735	0.386	-1.9	0.057
Population density	173.455	42.180	4.11	< 0.001
Employment density	-243.780	82.203	-2.97	0.003
Ratio of Emp to Non-Emp	0.252	0.094	2.68	0.007
Density of dead ends	—			
Links per node	-0.094	0.019	-5.440	< 0.001
Vertices density	—			
Pseudo R Square: 0.31		BIC: 2145.4		

and control groups. Table 3 shows the t-test of differences in covariate means before and after radius matching (caliper=0.05). It can be seen that the characteristics between the treated and original control groups are imbalanced. Matching is subsequently used to refine the control group and the bias due to differences in observable characteristics is reduced as shown in table 3. The sample size of control group is now refined to 1415.

Next, we check the distributions of propensity scores for both groups. The histograms in figure 3 show that the propensity scores have similar ranges across the two groups and overlap very well, indicating the overlap assumption is plausible. It is also worth noting that the propensity scores do not have either extremely high or extremely low values, which can cause problems when inverse weighting by creating large, weighted outcome values (Kurth et al., 2006).

Table 3: T-test of covariate means pre- and post-matching

Covariate	Unmatched Matched	Mean Treated	Control	%reduct bias	t-test t	$p > t$
KSI	U	5.375	5.715		-0.35	0.726
	M	5.375	5.516	58.7	-0.19	0.851
Slightly injured	U	36.088	34.203		0.33	0.742
	M	36.088	37.013	50.9	-0.19	0.852
A roads (%)	U	0.092	0.059		5.16	< 0.001
	M	0.092	0.084	75.4	0.69	0.488
B roads (%)	U	0.046	0.024		5.35	< 0.001
	M	0.046	0.042	84.8	0.38	0.705
M roads (%)	U	0.862	0.917		-7.26	< 0.001
	M	0.862	0.874	79.1	-0.8	0.423
IMD	U	30.558	22.871		9.61	< 0.001
	M	30.558	30.294	96.6	0.23	0.820
Domestic (%)	U	0.137	0.124		3.75	< 0.001
	M	0.137	0.136	94.5	0.14	0.885
Non-domestic (%)	U	0.077	0.060		3.63	< 0.001
	M	0.077	0.075	85.6	0.38	0.702
Green space (%)	U	0.356	0.423		-4.86	< 0.001
	M	0.356	0.358	97.1	-0.11	0.909
Population density	U	0.009	0.007		10.53	< 0.001
	M	0.009	0.009	98.8	-0.07	0.941
Employment density	U	0.004	0.003		8.17	< 0.001
	M	0.004	0.004	96.3	-0.2	0.842
Ratio of Emp to Non-Emp	U	2.031	2.197		-4.4	< 0.001
	M	2.031	2.045	91.5	-0.23	0.817
Density of dead ends	U	0.055	0.053		0.26	0.795
	M	0.055	0.065	-467.8	-0.86	0.393
Links per node	U	3.030	3.026		0.2	0.844
	M	3.030	3.017	-243.9	0.4	0.687
Vertices density	U	19.275	17.816		2.2	0.028
	M	19.275	19.149	91.4	0.14	0.891

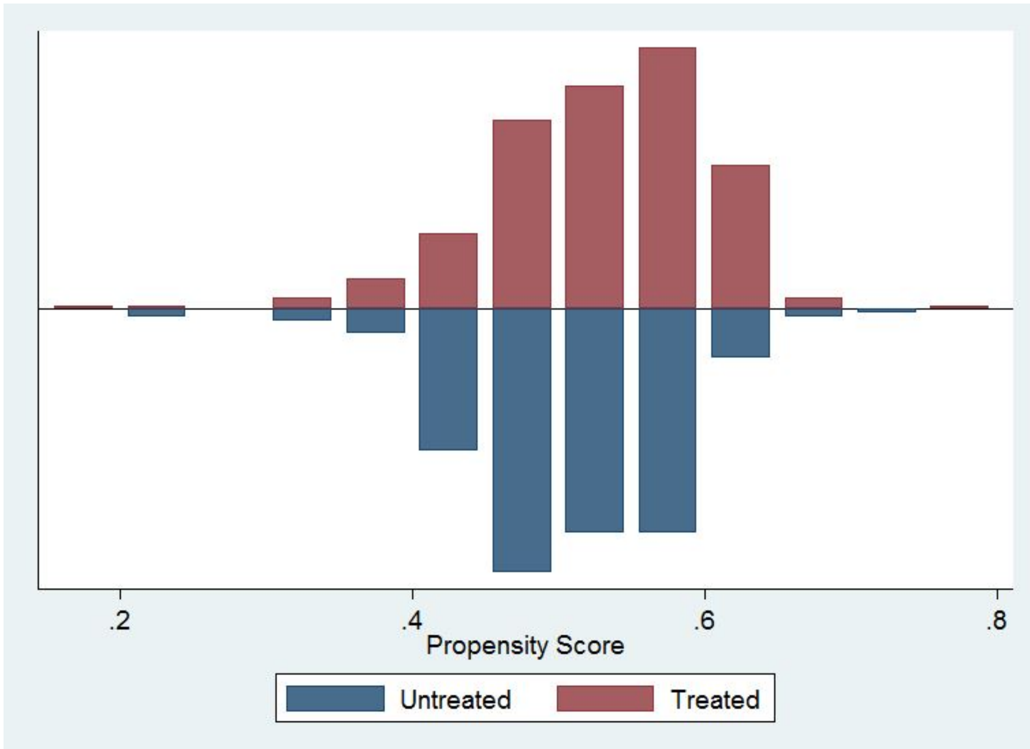

Figure 3: Propensity score distribution by treatment status

5.2. *Outcome regression models*

We apply generalized linear regression models using the pre-treatment covariates listed in Table 1 as predictors. The STATS 19 data classifies the casualty by severity (KSI and slightly injured) and by types (Cycle-, Pedestrian-, and Motor-related). Regression models for outcomes are fitted on the predictors for the treatment and control groups separately, and the predicted values will be obtained for the whole population.

Table 4 shows the regression results for different casualty severities and types by groups. Most covariates are significantly associated with the number of casualties for both treated and control groups. Specifically, the number of KSI and slightly injured casualties in baseline years are positively related to the casualty number after the treatment. The density of residential population and employees are used to control for traffic exposure within each zone. Most models show there are positive effects from the level of population and employment. This implies that more casualties may occur in zones with a higher density of residents and job opportunities. Socio-economic deprivation has previously been shown to be positively related to road traffic casualties (Graham and Stephens, 2008), and this has been confirmed by the results of this study which indicate that IMD scores have positively effects on all types of casualties.

We further investigate the effects of land use characteristics on casualties. Three main types of land use are examined: domestic, non-domestic and green space. The non-domestic area studied in this paper includes office district area and business area, such as large trade area, warehousing, and wholesaling. The results suggest higher percentages of domestic and non-domestic areas are associated with more casualties, whilst there are fewer casualties in areas with higher percentages of green space. This is consistent with previous findings (Pulugurtha et al., 2013), which suggest that land use characteristics such as residential and business areas are generally high traffic activity generators.

Regarding road network characteristics, as suggested in many other studies (e.g. Huang et al., 2010), road density is positively associated with road casualties at all types and severity levels. Two covariates were used as indicators of road network connectivity: the links per node and the number of nodes with one link (Chin et al., 2008). It can be hypothesized that areas with a better-connected road network will have more casualties, because pedestrians, cyclists and motor vehicles have better accessibility and total traffic activities tend to be more frequent. The results indicate that an increase in links per node is associated with an increase in the casualty numbers for all severities. Lower densities of nodes with one link, also known as dead ends, usually indicate limited access to streets. The results show that higher densities of dead ends are associated with fewer casualties. The results also suggest that road networks with a greater degree of horizontal curvature, i.e. more vertices per km, are associated with fewer casualties. This result is consistent with previous findings (e.g. Jones et al., 2008; Quddus, 2008). The mechanisms for this could be complex, however, one possible reason is that vehicles have lower speeds when passing curving road sections.

The adjusted R-square values are estimated to show the degree to which the outcome regression models fitted the treatment and control groups are appropriate to prediction. The adjusted R-square values are more than 60% for most models and are 70% for slightly injured models. The high values suggest the predicted outcomes from the regression models are valid.

Table 4: Outcome regression models

Covariates	Slightly Injured Treatment=1	Slightly Injured Treatment=0	Killed and Seriously Injured Treatment=1	Killed and Seriously Injured Treatment=0	Cycle-related Casualties Treatment=1	Cycle-related Casualties Treatment=0	Pedestrian-related Casualties Treatment=1	Pedestrian-related Casualties Treatment=0	Motor-related Casualties Treatment=1	Motor-related Casualties Treatment=0
KSI	0.049***	0.004***	0.043***	0.013***	0.069***	0.010***	0.046***	0.001	0.032***	0.004**
Slightly injured	0.003***	0.010***	0.005***	0.009***	2.0E-04	0.009***	0.004***	0.010***	0.006***	0.011***
A roads (%)	1.353***	1.373***	1.050***	1.402***	1.282***	1.542***	1.054***	1.066***	1.474***	1.483***
B roads (%)	1.209***	0.552***	0.664	1.076***	0.673***	0.889***	1.467***	0.581***	1.371***	0.490***
IMD	0.029***	0.004***	0.033***	0.002	0.023***	-2.0E-04	0.026***	0.009***	0.041***	0.005***
Domestic (%)	4.636***	0.223	1.535	-0.318	0.672	-2.288***	6.819***	3.688***	6.010***	0.638**
Non-domestic (%)	0.600	0.389***	1.078	-0.198	0.156	0.740***	2.394***	2.345***	-0.272	-1.713***
Green space (%)	-0.230*	-0.061	0.012	-0.339***	-1.395***	-1.052***	0.260	-0.221**	0.427	0.429***
Population density	44.9***	32.4***	59.7	40.9**	177.3***	40.7***	54.4*	2.3	-5.5	29.6***
Employment density	48.8***	71.3***	118.5**	114.0***	423.4***	191.8***	143.9**	-14.9	132.5***	-6.7*
Ratio of Emp to Non-Emp	0.067	0.013	0.138	0.058	-0.256	0.172***	0.228	0.058*	0.318***	-0.058
Density of dead ends	-1.970***	-0.838***	-3.875***	-0.625*	-3.295***	-2.051***	-2.011***	-0.196	-1.709***	-0.405**
Links per node	0.207**	0.038	0.016***	0.189***	0.516***	0.212**	-0.117	0.031	0.093	0.222***
Vertices density	-0.013***	-0.004***	-0.015**	-0.002	-0.010**	-0.003**	-0.016***	-0.002*	-0.014***	-0.005***
R-square	0.704	0.696	0.545	0.516	0.660	0.669	0.619	0.587	0.623	0.583
Obs	234	1415	234	1415	234	1415	234	1415	234	1415

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

5.3. *Effects of 20 mph zones on road casualties*

Given the satisfaction with the component models, propensity score and outcome regression as discussed in the above sections, now we proceed to estimate effects of 20 mph zones using the doubly robust method. For comparison, three other methods are also applied: inverse probability weighting, regression adjustment and propensity score matching. Besides inverse probability weighting, another application of propensity score is matching. The basic idea is to match each treated unit to untreated units with similar propensity scores. Conditional on the propensity score, differences in observed outcomes between the two groups can be solely attributed to the intervention impacts. The matching algorithm used in this paper is radius matching (caliper=0.05). For detailed discussion of matching algorithms, please refer to the work by Heinrich et al. (2010). Table 5 presents the estimations of the safety effects of 20 mph zones by casualty types and severities.

The 20 mph zones consistently have a significant impact on reducing casualties in both absolute number and percentages. The results are very similar for all four methods, with a reduction in slightly injured casualties of around 1.7 (10% in percentage), and KSI of around 0.73 (24% in percentage) respectively. The number of pedestrian-related casualties decreases by 0.85 (21% in percentage), which is significant at the 99% level for all four methods. In terms of motor-related casualties, only the absolute number of casualties is found to be significantly reduced by 1.5, whilst this effect is not significant when estimated in percentage. No significant effects of 20 mph zones are found on cycle-related casualties. The similar results from four methods increase confidence in the doubly robust method.

To investigate the robustness of DR method to model misspecification, we further examine false models by omitting confounders from both regression and propensity score models. The omitted confounders are significant predictors of outcomes but insignificant for propensity score estimation. This is similar to the routine done by Lunceford and Davidian (2004), and Bang and Robins (2005). The DR method should offer protection against the bias due to the misspecification of regression model. The results are shown in Table 5. All false models are distinguished by superscript "#". It can be seen that the false OR models lead to unstable estimates with relatively large standard errors due to the omission of significant covariates, while false DR as well as IPW estimators are consistent with the original results for most models. This shows that the DR method is superior for affording protection against misspecification.

It is worth noting that the effects of 20mph zones on reducing casualties estimated in this paper are smaller than the results from previous studies. For example, the reduction in casualties varies from 22% to 61% according to previous reports (Webster and Mackie, 1996; Webster and Layfield, 2003; Grundy et al., 2009; Steinbach et al., 2011). There are several possible reasons for this. First, is that the implementation period of 20 mph zones investigated in previous studies is from 1991 to 2008, and the effects of 20 mph zone may diminish over time (Grundy et al., 2009). If we focus on 20 mph zones implemented in recent years, the previous findings are more consistent with the ones in this paper. For example, Grundy et al. (2009) used conditional fixed effects Poisson models to estimate the effects of 20 mph zones using the same data. They first used the data from 1987 to 2006 and found significant reduction in casualties and collisions. Their initial findings are much higher than the ones of this study. Then they restricted analyses to 2000-2006, the period with the lowest annual numbers of casualties. The results are very similar to the ones of this study this time. For example, they found that the percentage reductions are 28.4% for KSI and 21.6% for pedestrian injuries, and no significant effect for cyclists. These results are, to a large extent, consistent with our findings. Second, and as discussed earlier, the over-estimation of treatment effects in previous studies could be also due to the selection bias. For example, Webster and Layfield (2003) use all unclassified roads in London as "control" data for roads in 20 mph zones. The characteristics of treated and control zones, e.g. historical records of casualties differ in the absence of the treatment, and the counterfactual outcomes approximated by such “control” zones will be biased. Finally, use of detailed panel data on road network characteristics provides adjustment for sources of confounding that have not been addressed in previous studies.

Table 5: Effects of 20 mph zones on road casualties

	Slightly Injured			Killed and Seriously Injured			Cycle-related Casualties			Pedestrian-related Casualties			Motor-related Casualties		
	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
Doubly Robust	-1.825	0.58	***	-0.727	0.172	***	-0.129	0.419		-0.858	0.263	***	-1.564	0.471	***
Inverse Probability Weighting	-1.762	0.638	***	-0.746	0.182	***	-0.123	0.417		-0.867	0.273	***	-1.518	0.492	***
Regression Adjustment	-1.777	0.583	***	-0.723	0.173	***	-0.11	0.424		-0.842	0.267	***	-1.547	0.47	***
Propensity Score Matching	-1.628	0.708	**	-0.787	0.242	***	-0.567	0.438		-0.68	0.307	**	-1.167	0.62	*
Doubly Robust #	-1.844	0.58	***	-0.645	0.231	***	-0.108	0.412		-0.841	0.271	***	-1.54	0.484	***
Inverse Probability Weighting #	-1.841	0.62	***	-0.639	0.237	***	-0.099	0.416		-0.853	0.282	***	-1.527	0.499	***
Regression Adjustment #	-1.306	0.834		-0.607	0.237	*	-0.01	0.414		-0.689	0.315	**	-1.213	0.584	*
	Slightly Injured (%)			Killed and Seriously Injured (%)			Cycle-related Casualties (%)			Pedestrian-related Casualties (%)			Motor-related Casualties (%)		
	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
Doubly Robust	-0.092	0.04	***	-0.243	0.049	***	-0.027	0.071		-0.217	0.046	***	-0.078	0.062	
Inverse Probability Weighting	-0.066	0.058		-0.255	0.049	***	-0.021	0.074		-0.217	0.047	***	-0.068	0.065	
Regression Adjustment	-0.089	0.042	**	-0.247	0.049	***	-0.026	0.073		-0.211	0.047	***	-0.077	0.061	
Propensity Score Matching	-0.113	0.04	***	-0.224	0.059	***	-0.035	0.067		-0.202	0.053	***	-0.095	0.066	
Doubly Robust #	-0.072	0.041	*	-0.239	0.054	***	-0.019	0.072		-0.223	0.047	***	-0.074	0.057	
Inverse Probability Weighting #	-0.046	0.089		-0.245	0.057	***	-0.028	0.075		-0.221	0.048	***	-0.053	0.07	
Regression Adjustment #	-0.069	0.032	*	-0.215	0.079	*	-0.021	0.073		-0.266	0.081	*	-0.075	0.058	

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

6. Discussion and Conclusions

Several studies have been conducted to evaluate the effects of 20 mph zones on road casualties in the UK. A key issue with causal analysis concerns how the statistical methods employed account for confounding. The ability to draw causal inferences from observational data relies on two properties: correctly specified models and comparability between the treatment and control groups under study. Neither of these issues has been addressed rigorously in previous studies. In this paper, we have applied the doubly robust method which affords us two opportunities for obtaining consistent and asymptotically unbiased causal effect estimates. Given the fact that we rarely know the exact relations among potential outcomes, treatment assignment, and confounding factors; the DR property is useful as it increases scope for satisfying model assumptions in practice. In addition, the propensity score incorporated in the doubly robust method can be used as the criterion when constructing the control group.

Our results show that the 20 mph zones consistently have significant impact on reducing casualties in both absolute number and percentages, especially for KSI and pedestrian-related casualties. Considering the diminishing effects of 20 mph zones over time, our results are consistent with the general conclusions of previous research in this field.

This paper also has two other major findings. First, previous studies rarely look at the criteria for 20 mph zones selection. Although there is considerable variability as to the implementation of 20 mph zones in different authorities, propensity score estimation suggests that the main factors affecting the decisions on 20 mph zones are the historical records of casualties and socio-economic characteristics, e.g. deprivation, land use, and population. Second, by developing a panel data of OS Meridian TM 2, the variation in the road network across time is controlled for in our models. The outcome regression models further show that zones with road network of high connectivity and more bends have more casualties.

There are also some limitations with the analysis presented in this paper. Due to data availability, the effects of 20 mph zones on traffic speeds are not investigated in the model. And population and employment are used instead of traffic volume to reflect the overall traffic activities. Despite these, the results from both the propensity score and outcome regression models suggest that the covariates included are significantly associated with the implementation of 20 mph zones and road casualties.

As suggested by Grundy et al. (2009), the effects of 20 mph zones may diminish over time. A study on temporal heterogeneity of treatment effect would make an interesting question. We also suggest researchers to compare the doubly robust method with other widely used causal methods, such as empirical Bayes for future research.

7. Acknowledgments

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). We would like to thank Chris Grundy at London School of Hygiene and Tropical Medicine for supplying the data of London 20 mph zone.

References

Augurzky, B., and C. M. Schmidt, 2000. The propensity score-a means to an end. IZA Discussion paper, No. 271, Bonn.

Austin, P. C., 2009. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat. Med., 28(25): 3083-3107.

Bang, H., and J. M. Robins, 2005. Doubly robust estimation in missing data and causal inference models. Biometrics, 61: 962-973.

Brookhart, M. A., S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn, and T. Sturmer, 2006. Variable selection for propensity score models. American Journal of Epidemiology, 163: 1149-1156.

Bryson, A., R. Dorsett, and S. Purdon, 2002. The use of propensity score matching in the evaluation of labour market policies. Working Paper No. 4, Department for Work and Pensions.

Bunn, F., T. Collier, C. Frost, K. Ker, R. Steinbach, I. Roberts, and R. Wentz, 2003. Area-wide traffic calming for preventing traffic related injuries. Cochrane Database of Systematic Reviews, Issue 1.

Casanova, J. and N. Fonseca, 2012. Environmental assessment of low speed policies for motor vehicle mobility in city centres. Global NEST Journal, 14(2): 192-201.

Chin, G. K. W., K. P. Van Niel, B. Giles-Corti, and M. Knuiman, 2008. Accessibility and connectivity in physical activity studies: The impact of missing pedestrian data. Preventive Medicine, 46: 41-45.

Cole, S. R., and M. A. Hernan, 2008. Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology, 168(6): 656-664.

Department for Transport, 2013. Setting local speed limits.

Dissanayake, D., J. Aryaija, and D.Wedagama, 2009. Modelling the effects of land use and temporal factors on child pedestrian casualties. Accident Analysis and Prevention, 41: 1016-1024.

Elvik, R., 2001. Area-wide urban traffic calming schemes: a meta-analysis of safety effects. Accident Analysis and Prevention, 33: 327-336.

Elvik, R., 2009. The power model of the relationship between speed and road safety Update and new analyses. Institute of Transport Economics, Norwegian Centre for Transport Research, Norway.

Elvik, R., P. Christensen, and A. Amundsen, 2004. Speed and road accidents: An evaluation of the power model. Institute of Transport Economics, Norwegian Centre for Transport Research, Norway.

Emsley, R. A., 2007. Statistical Models of Selection and Causation. PhD thesis, University of Manchester, UK.

Emsley, R., M. Lunt, A. Pickles, and G. Dunn, 2008. Implementing double-robust estimators of causal effects. The Stata Journal, 8(3): 334-353.

Funk, M., D. Westreich, C. Wiesen, T. Sturmer, M. Brookhart, and M. Davidian, 2011. Doubly robust estimation of causal effects. American Journal of Epidemiology, 173(7): 761-767.

Graham, D. J., and D. A. Stephens, 2008. Decomposing the impact of deprivation on child pedestrian casualties in england. Accident Analysis and Prevention, 40: 1351-1364.

Graham, D. J., E.J. McCoy and D. A. Stephens, 2015. Approximate Bayesian inference for doubly robust estimation. Baysian Analysis, page (in press).

Grundy, C., R. Steinbach, P. Edwards, J. Green, B. Armstrong, and P. Wilkinson, 2009. Effect of 20 mph traffic speed zones on road injuries in London, 1986-2006: controlled interrupted time series analysis. BMJ, 2009, 339: b4469.

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

Heinrich, C., A. Maffioli, and G. Vzquez., 2010 A Primer for Applying Propensity-Score Matching. Inter-American Development Bank, Technical Notes, No. IDB-TN-161.

Huang, H., M. A. Abdel-Aty, and A. L. Darwiche, 2010. County-level crash risk analysis in Florida Bayesian spatial modelling. Transportation Research Record: Journal of the Transportation Research Board, 2148: 27-37.

Jones, A. P., R. Haynes, V. Kennedy, I. M. Harvey, T. Jewell, and D. Lea, 2008. Geographical variations in mortality and morbidity from road traffic casualties in England and Wales. Health Place, 14: 519-535.

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

Lunceford, J. K., and M. Davidian, 2004. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in Medicine, 23: 2937-2960.

Marshall, W. E. and N. W. Garrick, 2011. Does street network design affect traffic safety? Accident Analysis and Prevention, 43: 769-781.

McCaffrey, D. F., G. Ridgeway, and A. R. Morral, 2004. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. Psychol. Methods, 9(4): 403-425.

Myers, J. A., J. A. Rassen, and J. J. Gagne, et al., 2011. Effects of adjusting for instrumental variables on bias and precision of effect estimates. Am. J. Epidemiol, 174(11): 1213-1222.

Pulugurtha, S., V. Duddu, and Y. Kotagiri, 2013. Traffic analysis zone level crash estimation models based on land use characteristics. Accident Analysis Prevention, 50: 678-687.

Quddus, M. A., 2008. Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. Accident Analysis and Prevention, 40: 1486-1497.

Rifaat, S. M., R. Tay, and A. Barros, 2011. Effect of street pattern on the severity of crashes involving vulnerable road users. Accident Analysis and Prevention, 43:

276-283.

Robins, J. M., 1999. Robust estimation in sequentially ignorable missing data and causal inference models. In American Statistical Association Section on Bayesian Statistical Science, pages 6-10.

Robins, J. M., A. Rotnitzky, and L. P. Zhao, 1995. Analysis of semiparametric regression-models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association, 90: 106-121.

Rodgers, S., S. Jones, S. Macey, and R. Lyons, 2010. Using geographical information systems to assess the equitable distribution of traffic-calming measures: translational research. Injury Prevention, 16: 7-11.

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

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

Soole, D., B. Watson, and J. Fleiter, 2013. Effects of average speed enforcement on speed compliance and crashes: A review of the literature. Accident Analysis and Prevention, 54: 46-56.

Steer Daveis Gleave, 2014. Research into the impacts of 20mph speed limits and zones. Report.

Steinbach, R., C. Grundy, P. Edwards, P.Wilkinson, and J. Green, 2011. The impact of 20 mph traffic speed zones on inequalities in road casualties in London. Journal of Epidemiol Community Health, 65: 921-926.

Sturmer, T., K. J. Rothman, and R. J. Glynn, 2006. Insights into different results from different causal contrasts in the presence of effect measure modification. Pharmacoepidemiol Drug Saf., 15(10): 698-709.

Tovar, M. and I. Kilbane-Dawe, 2013. Effects of 20mph zones on cycling and walking behaviours in London. Par Hill Research Ltd.

Webster, D. C. and A. M. Mackie, 1996. Review of traffic calming schemes in 20 mph zones. TRL Report 215, Transport Research Laboratory, Crowthorne.

Webster, D. C. and R. E. Layfield, 2003. Review of 20mph zones in London boroughs. TRL Report PPR 243, Transport Research Laboratory, Crowthorne.

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

Wier, M., J. Weintraub, E. H. Humphreys, E. Seto, and R. Bhatia, 2009. An area level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. Accident Analysis and Prevention, 41: 137-145.

Williams, D. and R. North, 2013. An evaluation of the estimated impacts on vehicle emissions of a 20mph speed restriction in central London. Imperial College London, London.