Corrections of Self-Selection Bias in Crash Causality Study: An Application on All-Red Signal Control

Sabyasachee Mishra[1*], Xiaoyu Zhu[2]

[1]Department of Civil Engineering, University of Memphis, 112D Engineering Science Building, 3815 Central Avenue, Memphis, TN 38152,United States

[2]National Center for Smart Growth Research and Education, University of Maryland, College Park, MD 20742, United States

[*] Corresponding author. Tel.: +1 901 678 5043; fax: +1 901 678 3026
E-mail address: smishra3@memphis.edu (S. Mishra), shuxy03@gmail.com (X. Zhu)

Abstract

All-Red (AR) interval is designed as a method of clearance interval to safely clear vehicles that enter the signalized intersection. The provision of AR is generally expected to reduce the occurrence of crashes, though there are situations that AR is not proved to be effective because it is used at intersections with a higher potential for crashes. This controversial result however, does not indicate that the AR interval is a contributing cause of crashes. Therefore, the self-selection bias of signal designs needs to be corrected when estimating their effect in improving safety.

To address the selection-bias problem at signalized intersections, a Heckman two-stage approach is adapted. First, a probit model is developed to explain the interrelationship between the AR interval and highway geometry, traffic volume, and environmental variables. Second, the selection bias term (or Heckman correction) is included in the second stage to build two negative binomial models for locations with and without an AR interval. Three-year crash data on urban signalized intersections in the Detroit metro area is used to validate the proposed models. The results show that the impact of AR intervals is 51.7% reduction in total crashes on the treated intersections, and 42.5% on a random intersection.

Key words: Self-selection bias, Heckman Two-step correction model, All-Red Interval, Probit Model

1. Introduction

The signalized intersection is a critical subject in traffic control and operation. According to the National Highway Traffic Safety Administration's (NHTSA) Traffic Safety Facts 2009 Report, there were more than 2.2 million reported intersection-related crashes, resulting in 7,043 fatalities (20.8% of total fatalities). Among the intersection crashes, 1.06 million (around 50%) crashes took place at traffic-signalized intersections (NHTSA, 2009). To improve intersection safety, crash causation studies are a popular approach to investigate the impact of traffic, geometry, and environmental factors.

To investigate the causality effect in reducing the occurrence of intersection crashes, the issue of a self-selection bias needs to be addressed first, especially when considering the intersection related signal control. Self-selection bias appears where observations in the data select themselves into a group (called "treated"), causing a biased sample. It arises where the characteristics of the observations which cause them to select themselves in the group create abnormal or undesirable conditions in the group (Heckman 1979). Obviously "treated" observations should be different than the rest of the population. Therefore, we are not able to divide the part of outcome related to the causal relationship from the part due to the self-selection into treated and untreated group. For example, the signal design (such as All-Red (AR) clearance interval) depends on the size and the complexity of the intersection (McGee et al. 2012; Srinivasan et al. 2011). According to Manual of Uniform Traffic Control Devices (MUTCD), the AR intervals are applied at intersections with more complex traffic and geometry design or higher potential for conflicts, where there are higher potential of crash occurrence (MUTCD, 2009). Therefore, the number of crashes at intersections with AR may be higher than other similar intersections. On the other hand, the AR interval is expected reduce the crash occurrence rate. To examine the causality effect of AR on crash reduction, the sample selection issue that AR is required and more applied in the intersections with more crashes should be addressed and captured. Due to the bias, the scenarios where intersections with AR have more crashes than without AR does not mean that AR interval is not effective in crash reduction, because the intersection in question may have already had a higher crash rate which necessitated the use of an AR design. The reduction effect of AR could not be directly achieved by comparing the number of crashes at the two types of intersections. Method to capture selection bias and AR effect at the intersections is the focus in this research.

The current method for traffic control effect on crash reduction is primarily through the use of before and after studies, for example, speeding camera, warning sign, red light running, (Persaud et al. 2005). This experimental design avoids the sample selection bias, but takes a long period of time to collect data and may cause problems in the traffic control because of changes in the design and possible risks. Based on the available data and existing design, statistical methods are generally applied to measure the traffic control effect. Then, this self-selection issue occurs but is commonly ignored in most of the current research. Raised by Heckman et al. (2001), a two-step procedure method can measure the selection bias and capture the target treatment effect, such as school enrollment selection (Heckman et al., 2001) and residential location choice (Zhou and Kocckelman, 2007).

This research is the first attempt to apply Heckman two-step method to traffic safety research and examine the AR effect on crash reduction and its selection bias. In this paper, we applied the Heckman two-step method to quantify the selection bias and the effect of AR interval at a random intersection and a treated intersection (design with AR).

Sample selection issues arise when a researcher is limited to information on a non-random sub-sample of the population of interest. Specifically, when observations are selected in a process that is not independent of the outcome of interest, selection effects may lead to biased inferences regarding a variety of different items. In general, Heckman's model is among one of the most widely used approaches for sample selection because of (a) the simplified approach for modeling dummy endogenous variable; (b) appropriate treatment for unobserved selection factors by explicitly using information gained from the model of sample selection; (c) a two-step procedure is simple in its form and addresses sample selection bias well (Green 1981, Maddala 1983). In transportation analysis Heckman sample selection approach includes role of secondary accidents in traffic operations (Zhang, and Khattak 2009); determining relationship between adult's and school children's trip mode (Deka 2013); neighborhood type and selection on driving behavior (Cao 2009); productive commuting time in a university setting (Zhao et al. 2013); willingness to change mode from motorized to non-motorized travel (Ferrer, and Ruiz 2013); and analyzing gender and automobile service for non-work service trips (Vance and Iovanna 2007).

In the next section, the literature on crash occurrence and their associated variables is discussed. Then, the methodology, a two-step procedure combining a binary probit model with a negative binomial model, is utilized to correct self-selection bias. The data for this case study is summarized in the following section and empirical analysis is conducted to examine the signal timing and roadway design effect in reducing the number of crashes. Finally, we conclude with a discussion on the importance of this approach and its useful implications.

2. Literature review

In this section, we include the method and results from previous causality research in traffic safety. Then, the experience of selection bias in safety research is discussed. A summary of literature review is shown in Table 1.

2.1 Causality Study of Traffic Safety

Crash occurrence has long been studied for mid blocks and intersections. Rural intersections, such as Abdel-Aty and Nawathe (2006), and Kim et al. (2007), and urban intersections, such as Lord and Persuad (2003) and Greibe (2003), are both commonly studied. But, there are different focuses between rural and urban areas because of the nature of density of development in the vicinity. Driver behavior changes by the access points in terms of cluster of developments around the intersections. In addition, there are various targets in past studies. For example, Songchitruksa and Tarko (2006) studied the right-angle crashes at signalized intersections. Wang and Abdel-Aty (2008) modeled the detailed left-turn crash occurrence according to conflict patterns. Lee and Abdel-Aty (2005) focused on vehicle-pedestrian crashes only in their study.

Kim and Washington (2006) proved that crashes, particularly those related to signal violations at signalized intersections could be reduced by an AR clearance interval. Inadequate AR interval may cause intersection crashes if there is a deviation from the recommended signal timing practices. A study by Zador et al. (1985) found that AR intervals at signalized intersections are commonly ignored and there is a statistically significant relationship between AR intervals that are too short and an increasing number of crashes. Moreover, some locations with AR intervals are not effective in reducing crashes. A study by the Minnesota Department of Transportation shows that short-term improvement (up to one year before-and-after implementation of an AR clearance interval) is beneficial in reducing intersection crashes related to signal violations. On the other hand, long- term (more than two years before-and-after implementation of the AR clearance interval) research findings indicate that the short-term benefits are not sustained (Souleyrette et al., 2004).

<<Table 1 Here>>

The most frequent and widely accepted approach in past studies is using Poisson and Binomial regression models (e.g., Shankar et al., 1997; Ma et al., 2007). Recently, there have been some improvements based on this approach such as random effect negative binomial applied by Chin and Quddus (2003). Lord et al. (2005) compared the Poisson, Poisson-gamma and zero-inflated regression models in crash occurrence. Considering the exposure rate at different locations, number of crashes could be normalized to capture the crashes per day or crashes per unit length (Karlaftis and Golias 2002). Lord and Mannering (2010) proposed methodological alternatives to analyze crash-frequency data.

Table 1 shows that in general traffic, geometry and environmental factors are taken into account while developing crash causation studies. For instance, Annual Average Daily Traffic (AADT) has proven to be highly correlated with crash occurrence (see Persaud et al., 2004; Kim et al., 2006; and Kim et al., 2007). Left turn lanes are addressed to reduce the number of crashes in several studies by Kim et al. (2006), Oh et al. (2003), and Harwood et al. (2000). Guo et al., (2010) found that the size of the intersection, the traffic conditions by turning movement, and the coordination of signal phase have significant impacts on intersection safety. Wang and Abdel Aty (2006) found intersections with left-turn protection on the minor roadway, with high speed limits on the major roadway, and in high population areas are correlated with high rear-end crash frequencies. But the effect of signal timing on intersection safety is limited in the literature. An investigation of effects of AR interval on crash occurrence is discussed in this paper.

In conclusion, Poisson and negative binomial regression are still preferred methods to model crashes occurrence and prediction. The causal relationship between signal timing of urban intersections and crashes has received limited attention in the literature. In this research, we will use the two-step Heckman correction approach to study the crash occurrence. In addition to controlling for roadway design, the research will focus on signal timing (AR interval).

2.2 Selection Bias in Safety Analysis

Sample selection issues arise when a database is limited to information on a non-random sub-sample of the population of interest. Analyses based on non-randomly selected samples may lead to erroneous conclusions and poor policy decision making. To address this issue, one of the early studies by Heckman suggested using two stage correction method (Heckman 1979). In traffic safety analysis self-selection bias is also studied by a number of researchers. Safety concerns in using such an approach to assign countermeasures have necessitated transportation agencies to incorporate treatments to locations experiencing high-crash frequencies or severities where treatment implementation is not randomized. Statistical causal inference methods may be used to estimate the average treatment effect of a countermeasure from nonrandomized observational data. Such an approach minimizes bias in the estimated treatment effect by mimicking randomization based on observed covariates (Sasidharan and Donnell 2012). In the broader area of traffic safety research selection bias is used in a variety of topics including pedestrian injury severity factors using linked police-hospital data (Tarko and Azam 2011), impact of self-reported injury and non-injury crash data (Hanley and Sikka 2012); cross sectional studies on crash reduction (Davis 2000); safety education for pedestrians for injury reduction (Duperrex 2002); effect of provisional licenses for traffic safety benefits (Jones 1994); safety evaluation shoulder rumble strips on freeways (Griffith 1999); evaluating traffic judgments between young and old adult pedestrians (Oxley et al. 1997); countermeasure selection bias for validating crash models for rural intersections (Oh et al. 2003) and assessment of fatalities resulted from non-use of seat belts (Petridou et al. 1998). Four measures of effectiveness namely average treatment effect

(ATE), treatment on the treated (TT), local average treatment effect (LATE), and marginal treatment effect (MTE) are suggested in literature (Zhou and Kockelman 2007; Heckman et al. 2001).

3. Data

The data for the research includes (a) crash data, (b) traffic data, (c) highway geometry and (d) traffic signal data. Figure 1 shows the locations of all intersections used in the analysis. Crash data are collected from South Eastern Michigan Council of Governments (SEMCOG) for the years 2002 to 2004. The normalized crash rate is defined as the number of crashes divided by the exposure data (logarithm of AADT) and used as the dependent variable in the analysis. The highway geometry and crash data were collected within 250 feet of the intersection. The traffic geometry and signal data are further classified on major and minor approaches. In this study, AR intersection is defined as AR intervals on both approaches. The remaining intersections are categorized as “Non-AR Intersections”. Definitions and descriptive statistics for the variables used in the analysis are presented in Table 2.

<<Figure 1 Here>>

A total of 185 observations are used in the analysis procedure. Total crashes for a three-year period are examined, excluding pedestrian and fixed object crashes. First, we observe the distribution of normalized crash rate for AR and Non-AR intersections are approximately normal. With mean crash rate of 1.99 and 2.12, there is no significant difference between treated and untreated. Most of the variables have similar mean values comparing the AR and non-AR intersections. Major streets and minor streets are different in traffic volume, number of lanes, hazard rate, etc. A number of signal timing variables such as cycle length, amber time, pedestrian crossing time, presence and absence of protected and permitted left turns are considered in the analysis.

<<Table 2 Here>>

4. Methodology

In this section, we formulate the approach based on a classical Heckman two-step procedure (also called as Heckit model), introduced by Heckman (1979) and Heckman et al. (2001). The core idea is to achieve the selection-correction terms from a binary probit model for treatment selection in the first step (selection equation) and then utilize these terms as explanatory variables in regular model estimation in second step (outcome equation). Simple closed-form expressions for treatment parameters: ATE and TT are derived to quantify the effects of individuals' self-selection.

As mentioned above, this two-step model contains: (1) Probit selection equation; and (2) outcome equation, negative binomial regression. The two-step Heckman correction (Heckit) model is formulated as the following discussion:

(i) Probit selection equation is formulated and computed as

$$D_i = \begin{cases} 1 & if\ D_i^* \geq 0 \\ 0 & otherwise \end{cases} \tag{1}$$

$$D_i^* = Z_i\theta + \varepsilon_i^D \qquad 2)$$

where,

D_i denotes the observed treatment decision (AR or Non-AR) of the intersection i;($D_i = 1$ for intersections with AR and $D_i = 0$ for intersections without AR);

D_i^* is the latent variable contributed by the intersection related characteristics;

Z_i is a row vector of explanatory variables for intersection i;

θ is the coefficient of corresponding variable effect to be estimated;

ε_i^D is the error term for unobserved variables, which is normally distributed.

(ii) Use $\hat{\theta}$ obtained from (i) to compute the selection-correction terms, which are also defined as inverse Mill's ratio.

$$E\left(\varepsilon^D \middle| Z_i\hat{\theta}, D_i = 1\right) = \frac{\phi(Z_i\hat{\theta})}{\Phi(Z_i\hat{\theta})} = \lambda(Z_i\hat{\theta}) \tag{3}$$

$$E\left(\varepsilon^D \middle| Z_i\hat{\theta}, D_i = 0\right) = -\frac{\phi\left(Z_i\hat{\theta}\right)}{1 - \Phi\left(Z_i\hat{\theta}\right)} = -\lambda(-Z_i\hat{\theta}) \tag{4}$$

where,

$\phi(\cdot)$ is the probability density function

$\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution,

$\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ is the inverse Mill's ratio.

(iii) Estimate the outcome equation of the two-step model. Under the assumption that the error terms of selection equation and outcome equation are jointly normally distributed, estimate two regression models for groups with treatment and without treatment, respectively, with the appropriate selection-correction terms obtained from (ii).

$$y_i^1 = X_i\beta^1 + \varepsilon_i^1 = X_i\beta^1 + \rho_1\sigma_1\lambda(Z_i\hat{\theta}) + \eta_i^1 \tag{5}$$

$$y_i^0 = X_i\beta^0 + \varepsilon_i^0 = X_i\beta^0 - \rho_0\sigma_0\lambda(-Z_i\hat{\theta}) + \eta_i^0 \tag{6}$$

where,

y_i^1 and y_i^0 are the response variables, representing the number of crashes divided by log (AADT) at AR and non-AR intersections respectively;

X_i is the predictors, including intersection design, traffic and signal control characteristics for intersection i;

β^1 and β^0 are the coefficients to be estimated;

$\lambda(Z_i\hat{\theta})$ and $-\lambda(-Z_i\hat{\theta})$ are the correction terms (inverse Mills ratio) calculated from (ii), and these terms are used as the predictors in regression;

η_i^1 and η_i^0 are vectors of residuals from the regression, $E(\eta^1) = E(\eta^0) = 0$;

ε_i^1 and ε_i^0 are error terms for unobserved variables with the assumption of 3-dimensional normal distribution with ε^D in selection equation (2)

$$\begin{pmatrix} \varepsilon^D \\ \varepsilon_i^1 \\ \varepsilon_i^0 \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_1\sigma_1 & \rho_0\sigma_0 \\ \rho_1\sigma_1 & \sigma_1^2 & \rho_{12}\sigma_1\sigma_0 \\ \rho_0\sigma_0 & \rho_{12}\sigma_1\sigma_0 & \sigma_0^2 \end{pmatrix}\right). \tag{7}$$

The observed outcome y_i is denoted as

$$y_i = D_i y_i^1 + (1 - D_i) y_i^0. \tag{8}$$

(iv) Use the estimation result of $\hat{\beta}^1, \hat{\beta}^0, \widehat{\rho_1\sigma_1}$, and $\widehat{\rho_0\sigma_0}$ obtained from step (iii) and $\hat{\theta}$ from step (i) to obtain point estimates of treatment parameters ATE and TT.

$$ATE(x) = x\left(\hat{\beta}^1 - \hat{\beta}^0\right) \tag{9}$$

$$TT(x, z, D[z] = 1) = x\left(\hat{\beta}^1 - \hat{\beta}^0\right) + (\widehat{\rho_1\sigma_1} - \widehat{\rho_0\sigma_0})\frac{\phi(z_i\hat{\theta})}{\Phi(z_i\hat{\theta})} \tag{10}$$

where,

$ATE(x)$ is the average treatment effect (ATE) evaluated for a random sample.

$TT(x, z, D[z] = 1)$ is the expected change caused by the treatment for those selected to be treated (TT).

ATE represents the treatment effect on the crash rate for a random intersection and TT is the treatment effect for the treated intersections. ATE or TT >0 means intersections with AR have more crashes than without AR, and vice versa. A comparison of ATE and TT values can evaluate the part of causality effect and the part of the effect caused by the self-selection. Implications of ATE and TT are discussed in the results section.

5. Results

The estimation is conducted in the open source statistics software R. The results for two steps are displayed in Table 3. In this section, we first discuss the results of the probit model along with a description of its corresponding significant variables. Second, the negative binomial regressions for two groups are discussed and explained. The treatment effect of an AR interval is then compared and explained.

The first step analyzes the choice of AR design on both major and minor approaches using a binary probit model. The positive sign of “Estimate” means the variable is positively correlated with AR, vice versa. For example, the intersections with AR are less likely to have Left Turn Protection on Major approach or it could be interpreted as intersections without left turn protection are more likely to have AR intervals. This is because at the permissive left turn locations, vehicles make a left using the gap of opposite through traffic, especially the last few seconds of green and yellow interval. Therefore, AR is designed to clear the left turn traffic in the intersection. The size of the intersection is also an important factor related with AR design.

From the intersection geometry data, there are information for the number of lanes, median width, median type, and shoulder width for both major and minor approaches. The total width for both major and minor are also approximated from the above (shown in Table 2). Among all the highway geometry variables, the existence of medians at the major approach is highly correlated with AR interval. With a median at the major approach, the intersection is more likely to have AR interval for both directions. A major street speed limit, for instance less than 30 miles per hour, increases the probability of a required AR interval. This is not consistent with the opinion that AR should be designed at the locations with higher speed. However, the intersections analyzed in this study are located in an urban Central Business District area where traffic volume is very high and speed limit is mostly 30mph and 35mph. The locations with lower speed limit are generally related to narrower roadways with more conflicts. In addition, other approaches for traffic information, such as using the speed limit as a continuous variable, do not provide a significant outcome.

Table 3 also shows negative binomial model results for (1) AR Model, and (2) Non-AR Model. A set of common variables (number of lanes and amber time period) are used for both models. The outcome is the crash rate defined as number of crashes at the intersection divided by logarithm of Average Daily Traffic (ADT). The traffic variables, such as ADT on two approaches, are not significant. Speed limit is also examined as continuous covariates or categorical factors, but is also not significant. It is because that the speed limit in the sample is not quite high (mostly under 35mph) comparing with freeway where accidents are more related with speeding. The significant geometry variable in the outcome equation is number of lanes on major approach for both AR and Non-AR intersections. Other variables shown in Table 2 are proved to be not significantly correlated with crash rate in this study. The crash rate increases with more lanes on major approach, which also represents higher crash rate at a larger intersection. Signal design effect is also examined in outcome equations, including left turn phasing, protection or not, and amber time. Only the amber time period is highly related with the crash rate for Non-AR intersections, but not effective at AR locations. Locations without AR have to clear the traffic and reduce conflicts during yellow intervals, but the vehicles also tend to rush during the yellow interval. The result implicates that at the Non-AR intersections, longer amber period on the major approach may cause more crashes. In this case study, the R-square does not indicate a good degree of fit, as also found in two-stage Heckman correction approach for self-section bias literature (Zhou and Kocckelman, 2007). The lower R-square is limited by the sample size and the related variables but does not impact the measurement of AR effect.

<<Table 3 Here>>

Above is the two step model that corrects the selection bias and tests the variables related with crash rate. While, to examine the effect of AR on crash rate, we need to conduct the step (iv) in the methodology section. ATE and TT parameters are estimated using equation (9) and (10). The magnitude of treatment effects is shown in Table 4. Given the average crash rate at the CBD area is 2.02, ATE of -1.49 represents that a random intersection would have 1.49 higher crash rate if converting AR to Non-AR, or the AR effect on a random intersection is reducing the crash rate from 3.51 (2.02+1.49) to 2.02, which is a 42.5% (1.49/3.51) reduction. The TT value shows that for intersections already designed with AR the crash rate will increase by 2.13 if it is Non-AR. Given the average crash rate at AR of 1.99, the impact of AR intervals on the intersections with treatment is reducing the crash rate from 4.12 (=1.99+2.13) to 1.99, which is 51.7% reduction. From the above values, we observe that the average of crash rate for AR intersections is lower than average for the whole sample, because the crash rate for Non-AR is higher than AR. The treated locations (designed with AR) are self-selected with a higher initial crash rate as 4.12, comparing with a random average of 3.51. The reduction effect of AR on the crash rate (51.7%) is much higher for the self-selected locations with AR treatment, comparing with a random location with 42.5% reduction. Based on the magnitude of the ATE and TT, the impact of AR on crash rate decrease for a random intersection is approximately 70% (1.49/2.13) of the treated effect versus non-treated. Therefore, the self-selection effect of treatment counts 30% within the decrease among the intersection with AR versus Non-AR.

<<Table 4 Here>>

6. Conclusions

In this paper, we propose a two-stage approach to examine the effect of AR signal design in reducing the number of crashes. In the first stage, a probit model demonstrates the interdependence between AR and other geometry variables. In the second stage, negative binomial models analyze the crash causal factors including a self-selection bias correction term. Additionally, two treatment parameters (ATE and TT) are estimated to examine the effect on AR for all intersections, and treatment intersections. Three years of crash data are used for analysis of selected signalized intersections in the Detroit Metro area consisting of 185 samples. Within this dataset there are 137 treated samples and 48 untreated samples.

The probit model results show AR interval design is negatively correlated with protected left turns on major streets, and positively correlated with median width, and lower speed limit on major streets. In the negative binomial model for the both data sets (AR and non-AR), number of lanes and amber time period are correlated with crash rate. The impact of AR intervals is a 51.7% reduction on the treated intersections, and a 42.5% reduction on a random intersection. The impact of AR for a random intersection is approximately 70% of the treated effect versus non-treated. And the self-selection effect of treatment counts 30% within the decrease among the intersection with AR versus Non-AR.

In addition to correcting the self-selection bias of an AR interval, this approach provides methodological implications to analyze other variables with selection bias. For some of before-after studies in traffic safety, researchers need to collect crash data for a long duration before and after an event to evaluate a treatment. It is not always appropriate to do the before-after experiment considering the risk of crash caused by the experiment. With method developed in this paper, researchers do not need to wait for the outcome with treatment and risk to apply the treatment. Based on the current sample, with and without treatment, the treatment effect could be calculated with correction of selection bias and control of other variables. Furthermore, it should be noted that selection bias estimation and correction rely on the assumptions of a joint normal distribution, and may not be robust for departures from this distributional assumption (Heckman et al. 2001). More robust approaches tend to be nonparametric in nature. Future research should incorporate nonparametric distribution in the estimation of selection bias terms.

7. References

Abdel-Aty, M., and Wang, X., 2006. Crash Estimation at Signalized Intersections Along Corridors: Analyzing Spatial Effect and Identifying Significant Factors, *Transportation Research Record: Journal of the Transportation Research Board, No. 1953*, TRB, National Research Council, Washington, D.C., 98-111.

Abdel-Aty, M., Keller, J., and Brady, P. A., 2005. Analysis of Types of Crashes at Signalized Intersections by Using Complete Crash Data and Tree-Based Regression. *Transportation Research Record: Journal of the Transportation Research Board, No. 1908*, TRB, National Research Council, Washington, D.C., 37-45.

Abdel-Aty, M., and Nawathe, P., 2006. A novel approach for signalized intersection crash Classification and prediction. *Journal of Advances in Transportation Studies information* 9, 67-80.

Cao, X. (2009). Disentangling the influence of neighborhood type and self-selection on driving behavior: an application of sample selection model. Transportation, 36(2), 207-222.

Chin, H. C., and Quddus, M. A., 2003. Applying the random effect negative binomial model to examine traffic accident occurrence at signalized intersections. *Accident Analysis & Prevention* 35(2), 253-259.

Davis, G.A., 2000. Accident reduction factors and causal inference in traffic safety studies: a review. Accident Analysis & Prevention, 32(1), pp.95–109.

Deka, D. (2013). An explanation of the relationship between adults' work trip mode and children's school trip mode through the Heckman approach. Transport Geography, July 2013, 54-63.

Duperrex, O., 2002. Safety education of pedestrians for injury prevention: a systematic review of randomised controlled trials. *BMJ*, 324(7346), pp.1129–1129.

Green 1981. "Sample selection bias as a specification Error: Comment." Econometrica 49(3): 795-98.

Greibe, P., 2003. Accident prediction models for urban roads. *Accident Analysis and Prevention 35* (1), 273-285.

Griffith, M., 1999. Safety Evaluation of Rolled-In Continuous Shoulder Rumble Strips Installed on Freeways. *Transportation Research Record: Journal of the Transportation Research Board*, 1665(-1), pp.28–34.

Guo, F., Wang, X., and Abdel-Aty, M.A., 2010. Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis & Prevention*. 42, 84–92.

Ferrer, S., and Ruiz, T. (2013). Assessment of the Degree of Willingness to Change from Motorized Travel Modes to Walking or Cycling. Transportation Research Record, Journal of Transportation Research Board (TRB), Volume 2382 (1), pp. 112-120.

Hanley, P.F. and Sikka, N., 2012. Bias caused by self-reporting distraction and its impact on crash estimates. *Accident Analysis & Prevention*, 49(11), 360-365.

Harwood, D. W., Council, F. M., Hauer, E., and Vogt, A., 2000. Prediction of the expected safety performance of rural two-lane highways, *Report prepared by Midwest Research Institute*, Kansas City, MO, USA.

Heckman, J., Tobias, J.L. and Vytlacil, E., 2001. Four Parameters of Interest in the Evaluation of Social Programs. *Southern Economic Journal*, 68(2), pp.211–223.

Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica* 47(1), 153-162.

Jones, B., (1994). The effectiveness of provisional licensing in Oregon: An analysis of traffic safety benefits. *Journal of Safety Research*, 25(1), pp.33–46.

Karlaftis, M. G., and Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis & Prevention*, 34(3), pp.357–365.

Kim, D. J., and Washington, S., 2006. The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis & Prevention* 38(6), 1094-1100.

Kim, D.J., Washington, S., and Choi, K., 2007. Modeling crash outcome probabilities at rural intersections: application of hierarchical binomial logistic models. *Accident Analysis and Prevention* 39(1), 125-234.

Kim, D.J., Washington, S., and Oh, Jutaek., 2006. Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering, American Society of Civil Engineers* 32 (4), 282-292.

Lee, C., and Abdel-Aty, M., 2005. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. *Accident Analysis and Prevention* 37 (4), 775–786.

Lord, D., and Persuad, B. N., 2003. Accident prediction models with and without trend. *Transportation Research Record: Journal of the Transportation Research Board, No. 1717*, TRB, National Research Council, Washington, D.C., 102-108.

Lord, D., Washington, S. P., and John, N.I., 2005. Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis and Prevention* 37(1), 35-46.

Lord, D., and Mannering, F. 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. ransportation Research Part A: Policy and Practice, 44(5), 291-305.

Ma, J., Kockelman, K.M., and Damien, P., 2007. Bayesian multivariate poisson-lognormal regression for crash prediction on rural two-lane highways. CD-ROM. Transportation Research Board of the National Academics, Washington D.C.

Maddala (1983). "Limited Dependent and Qualitative Variables in Econometrics". Cambridge University Press.

McGee, H., Moriarty, K., Eccles, K., Liu, M., Gates, T., and Retting, R. (2012). *Guidelines for Timing Yellow and All-Red Intervals at Signalized Intersections*. National Cooperative Highway Research Program, Report 731, 978-0-309-25859-3, Transportation Research Board, Washington, DC: National Research Council.

MUTCD, 2009. Manual of Uniform Traffic Control Devices (MUTCD), http://mutcd.fhwa.dot.gov/pdfs/2009r1r2/mutcd09r1r2editionhl.pdf

NHTSA, 2009. Traffic Safety Facts 2009 Report, National Highway Traffic Safety Administration's (NHTSA), http://www-nrd.nhtsa.dot.gov/Pubs/811402.pdf.

Oh, J., Lyon, C., Washington, S., Persaud, B., and Bared, J., 2003. Validation of FHWA crash models for rural intersections: Lessons learned. *Transportation Research Record: Journal of the Transportation Research Board, No. 1840*, TRB, National Research Council, 41-49.

Oxley, J. Ihsen E, Charlton J, and Day R., 1997. Differences in traffic judgements between young and old adult pedestrians. *Accident Analysis & Prevention*, 29(6), pp.839–847.

Persaud, B, N., Retting, R. A., and Lyon, C. A., 2004. Crash reduction following installation of centerline rumble strips on rural two-lane roads. *Accident Analysis and Prevention* 36(6), 1073-1079.

Persaud, B., Council, F., Lyon, C., Eccles, K., and Griffith, M., 2005. Multijurisdictional Safety Evaluation of Red Light Cameras. *Transportation Research Record: Journal of the Transportation Research Board*, 1922, pp.29–37.

Petridou, E, Skalkidou A, Ioannou N, and Trichopoulos D., 1998. Fatalities from non-use of seat belts and helmets in Greece: A nationwide appraisal. *Accident Analysis & Prevention*, 30(1), pp.87–91.

Sasidharan, L. and Donnell, E.T. 2012. Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: Exploratory analysis using intersection lighting data. *Accident Analysis & Prevention*, In press. http://dx.doi.org/10.1016/j.aap.2012.05.036.

Shankar, V., Milton, J., and Mannering, F.L., 1997. Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis and Prevention* 29 (6), 829-837.

Songchitruksa, P., and Tarko, A.P., 2006. Practical method for estimating frequency of right-angle collisions at traffic signals. *Transportation Research Record: Journal of the Transportation Research Board, No. 1953*, TRB, National Research Council, Washington, D.C., 89-97.

Souleyrette, R. R., O'Brien, M. M., and McDonald, T., 2004. Effectiveness of All-Red clearance interval on intersection crashes. Minnesota Department of Transportation, Report # 2004-26. http://www.lrrb.org/pdf/200426.pdf, accessed May 27, 2012.

Srinivasan, R., Baek, J., Smith, S., Sundstrom, C., Carter, D., Lyon, C., Persaud, B., Gross, F., Eccles, K., Hamidi, A., and Lefler, N. (2011). *Evaluation of Safety Strategies at Signalized Intersections*. National Cooperative Highway Research Program, Report 705, 978-0-309-21345-5, Transportation Research Board, Washington, DC: National Research Council.

Tarko, A., and Tracz., M. 1995. Accident prediction models for signalized crosswalks, *Safety Science*, 19(2-3), 109-118.

Tarko, A. and Azam, M.S., 2011. Pedestrian injury analysis with consideration of the selectivity bias in linked police-hospital data. *Accident Analysis & Prevention*, 43(5), pp.1689–1695.

Vance, C., and Iovanna, R. (2007). Gender and the Automobile: Analysis of Nonwork Service Trips. . Transportation Research Record, Journal of Transportation Research Board (TRB), Volume 2013, pp. 54-61.

Wang, X., and Abdel-Aty, M., 2006. Temporal and spatial analyses of rear-end crashes at signalized intersections. *Accident Analysis & Prevention*, 38, 1137–1150.

Wang, X., and Abdel-Aty, M., 2008. Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. *Accident Analysis and Prevention* 40 (1), 76-88.

Yan, X., Radwan, E., and Abdel-Aty, M., 2005. Characteristics of rear-end accidents at signalizedintersections using multiple logistic regression model. *Accident Analysis & Prevention*, 37(6), 983-995.

Zador, P., Stein, H., and Tarnoff, P., 1985. Effect of clearance interval timing on traffic flow and crashes at signalized intersections. *Institute of Transportation Engineers (ITE) Journal* 55(11), = 36-39.

Zhou, B., and Kocckelman, K.M., 2007. Self-selection in home choice: Use of treatment effects in evaluating the relationship between the built environment and travel behavior. *Transportation Research Record*: *Journal of Transportation Research Board*, *No. 2077*, TRB, National Research Council, Washington, D.C., 54-61.

Zhang, H, and Khattak, S. (2009). What is the Role of Multiple Secondary Incidents in Traffic Operations? Compendium of papers, 88th Annual Meeting of Transportation Research Board, National Academy of Sciences, Washington DC.

Zhao, J., Lung, A., and Zhan, G. (2013). Quantity and Quality of Productive Commuting Time Use in a University Setting: A Heckman Model. Compendium of papers, 92nd Annual Meeting of Transportation Research Board, National Academy of Sciences, Washington DC.