Accident Analysis and Prevention 85 (2015) 58–65

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Exploring the influential factors in incident clearance time: Disentangling causation from self-selection bias

Chuan Ding[a], Xiaolei Ma[a,*], Yinhai Wang[b], Yunpeng Wang[a]

[a] School of Transportation Science and Engineering, Beijing Key Laboratory for Cooperative Vehicle Infrastructure System and Safety Control, Beihang University, Beijing 100191, China
[b] Department of Civil and Environmental Engineering, University of Washington, Seattle 98195, United States

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 29 August 2015
Accepted 29 August 2015
Available online 12 September 2015

Keywords:
Freeway incident
Response time
Clearance time
Self-selection bias
Treatment effect

ABSTRACT

Understanding the relationships between influential factors and incident clearance time is crucial to make effective countermeasures for incident management agencies. Although there have been a certain number of achievements on incident clearance time modeling, limited effort is made to investigate the relative role of incident response time and its self-selection in influencing the clearance time. To fill this gap, this study uses the endogenous switching model to explore the influential factors in incident clearance time, and aims to disentangle causation from self-selection bias caused by response process. Under the joint two-stage model framework, the binary probit model and switching regression model are formulated for both incident response time and clearance time, respectively. Based on the freeway incident data collected in Washington State, full information maximum likelihood (FIML) method is utilized to estimate the endogenous switching model parameters. Significant factors affecting incident response time and clearance time can be identified, including incident, temporal, geographical, environmental, traffic and operational attributes. The estimate results reveal the influential effects of incident, temporal, geographical, environmental, traffic and operational factors on incident response time and clearance time. In addition, the causality of incident response time itself and its self-selection correction on incident clearance time are found to be indispensable. These findings suggest that the causal effect of response time on incident clearance time will be overestimated if the self-selection bias is not considered.

1. Introduction

The duration of traffic incidents increase the occurrence likelihood of secondary incidents, and is likely to induce more severe traffic congestions (Zhang and Khattak, 2010, 2011). Therefore, it is of great importance to alleviate the impact of incidents by reducing incident duration for traffic management agencies. To achieve this goal, identifying the influential factors related to incident duration is considered as an effective countermeasure. The Highway Capacity Manual (TRB, 1994) defined incident duration into the following four phases: incident detection time, response time, clearance time, and traffic recovery time. Among them, incident clearance time and response time are two key measures to demonstrate traffic incident management (TIM) performance. Researchers have found that the traffic incident can be quickly cleared up if a prompt incident response is taken (Hou et al., 2014; Nam and Mannering, 2000). However, association does no necessary imply causality. It is likely that higher priorities (e.g. more manpower and equipment) are given to those incidents with short response time, while these incidents may share similar characteristics that inherently require less response time, and consequently lead to short clearance time. In this case, self-selection bias arises-individuals select themselves into groups with nonprobability sampling (Cao, 2009). This is to say that the effort of swift incident response may be exaggerated. Especially when other influential factors such as weather and traffic condition are involved, disentangling the interaction between endogenous and exogenous variables on incident clearance time becomes difficult.

To address this issue, this study applied endogenous switching model to disentangle the effect of incident response time and the effect of self-selection on incident clearance time. The primary contributions of this study lie in the following aspects: firstly, the self-selection bias in incident response time is tested and corrected through an endogenous switching modeling framework, and the extent to which self-selection contributes to the connection

* Corresponding author.
E-mail addresses: cding@buaa.edu.cn (C. Ding), xiaolei@buaa.edu.cn (X. Ma), yinhai@uw.edu (Y. Wang), ypwang@buaa.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.aap.2015.08.024
0001-4575/

between incident response time and clearance time is quantified; secondly, significant factors affecting incident response time and clearance time are captured, including incident, temporal, geographical, environmental, traffic and operational attributes. That is to say, for the increase of incident clearance time, what proportion is due to effect of self-selection and what proportion is caused by incident response time itself are calculated using Heckman's treatment effect theory (Mishra and Zhu, 2015). This study provides insightful evidence for traffic incident management agencies to adopt the most targeted countermeasure for incident duration minimization, and optimize the limited incident response resources.

The remainder of this paper is organized as follows. The methodology, an endogenous switching modeling framework with treatment effect analysis, is introduced in Section 2. In the following section, the data sources and description are summarized. Then the proposed model is applied and empirical model results are discussed in Section 4. Finally, conclusions and future studies are provided in the last section.

2. Literature review

The vast majority of previous studies focused on modeling incident duration based on traffic condition, incident characteristics and road structure. These methods can be generally grouped into three categories: regression-based approaches, artificial intelligence-based approaches, and survival analysis. Regression-based approaches include linear regression technique (Garib et al., 1997; Valenti et al., 2010) as well as time sequential model (Khattak et al., 1995; Qi and Teng, 2008), and were widely utilized for simplicity and effectiveness. To further improve the prediction accuracy of incident duration, artificial intelligence-based approaches were then adopted to capture the non-linear relationship between incident duration and its influential factors. Typical applications include artificial neural networks (Wang et al., 2005; Wei and Lee, 2007; Lopes et al., 2013), fuzzy logic models (Kim and Choi, 2001), naive Bayesian classifier (Boyles et al., 2007). Considering the stochastic nature of incident duration, probability distribution can better qualify incident duration with uncertainty and time dependency (Heckman and Borjas, 1980). This triggers the survival analysis, where hazard function can represent the conditional probability of incident being cleared given that the incident has lasted for a certain time period. A number of studies resorted to the hazard-based models for depicting incident duration distribution with associated explanatory factors (Jones et al., 1991; Nam and Mannering, 2000; Qi and Teng, 2009; Chung, 2010; Hojati et al., 2013; Wang et al., 2013; Hou et al., 2014; Li et al., 2015).

Nevertheless, most of previous studies neglected the dependency (i.e. correlation between explanatory variables and error term) and causality (i.e. self-selection bias) among explanatory variables when modeling incident duration (Weng et al., 2015), where their primary research focuses is to predict rather than interpret incident duration. This is partly because the computational intelligence techniques (e.g. neural network, fuzzy logic) lack sufficient model explanatory power to establish input-output relationship and causality (Karlaftis and Vlahogianni, 2011). One notable phenomenon is the interaction between different incident duration phases (i.e. notification time, response time, clearance time and traffic recovery time). For instance, both endogeneity and select-selection effects can be observed in modeling incident clearance time and incident response time (Nam and Mannering, 2000; Lee and Fazio, 2005): When the incident response team (IRT) is notified of a severe incident with fire or fatality, a higher priority is commonly granted in this scenario. This causes early arrival of response team with additional personnel to the incident site for prompt rescue, and further shortens the incident clearance time. This is to say, the reduction of incident clearance time may not be attributed to other relevant variable changes (e.g. traffic condition, incident type, etc.) but the self-selection of response time. In other words, self-selection causes biased sample selection problem where incidents with short response time select themselves for short clearance time (Zhou and Kockelman, 2009; Mokhtarian and Cao, 2008; Cao, 2009). On the other side, response time can be viewed as an endogenous variable. This is because the incident response time may cause changes for clearance time, while both the response time and clearance time are simultaneously affected by other factors such as weather, traffic condition and incident type. Therefore, it is difficult to judge whether the increase of clearance time is incurred by excessive response time or other contributing factors when establishing a statistical relationship for incident clearance time. Both self-selection bias and endogeneity bias intertwine with each other, and complicates the interpretation of incident clearance time. Failure to eliminate these biases may produce a misleading and invalid result (Hanley and Sikka, 2012). Therefore, it is desirable to identify the scales of the differences in incident duration owing to self-selection or other influential factors.

3. Methodology

The influence of incident response time on incident clearance time has been investigated, assuming that the incident response time is an exogenous attribute (Hou et al., 2014). However, the incident response time should be viewed as an endogenous variable in the incident clearance time model since the incident response time is determined by other variables, and these variables also influence the incident clearance time. For example, a serious incident with a short clearance time is more likely to be that which has a short response time to save the injured persons, rather than owing to good weather condition. In this case, the relationship between incident clearance time, response time and other contributing factors should be carefully studied. When the endogeneity bias is eliminated, the question remains whether a quick incident response can save more incident clearance time over the entire sample of incident. This is to investigate the total influence of incident response time on clearance time by quantifying the degrees of clearance time reductions due to self-selection or response time itself. To answer this question, endogenous switching modeling framework is applied because it can account for endogeneity of incident response and sample selection problem simultaneously. In this study, the endogenous switching model is utilized to explore the influential factors in incident clearance time, and is described in the following section. Detailed information can be found in the study conducted by Heckman et al. (2001), Lokshin and Sajaia (2004), and Alene and Manyong (2007).

3.1. Model specification

In the first step of the endogenous switching modeling approach, the incident response time is modeled by the binary (i.e. short response time and long response time) dependent variable method. In the second stage, equations for incident clearance time are then estimated separately for each group (i.e. short and long response time groups), conditional on the incident response time decision. Let D denotes the incident response time decision (with $D = 1$ being in a short response time group):

$$D = \begin{cases} 1 & \text{if} \quad \alpha Z + \varepsilon_D > 0 \\ 0 & \text{if} \quad \alpha Z + \varepsilon_D < 0 \end{cases} \tag{1}$$