A Reliability-Based Analysis of Bicyclist Red-Light Running Behavior at Urban Intersections

Mei Huan 1 and Xiaobao Yang 2

1 School of Economics and Management, Beijing Institute of Graphic Communication, Beijing 102600, China
2 MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

Correspondence should be addressed to Xiaobao Yang; yangxb@bjtu.edu.cn

Received 15 July 2014; Revised 28 September 2014; Accepted 29 September 2014

1. Introduction

Cycling is one of the most popular modes of transportation in some Asian developing countries, such as Cambodia, Vietnam, and China. Even in developed countries, cycling travel is recognized as low energy consumption and healthy to the users and it does not damage the health of others. US Census Bureau’s American Community Survey (ACS) reports twice as many daily bicycle commuters in 2009 as in 2000 [1]. At the same time, the data from the Canadian Census reveals a 42% increase in the number of daily bicycle commuters between 1996 and 2006 [2]. In March 2009, half (50%) of all Australian households had at least one working bicycle. Of these households, two-thirds (66%) had two or more bicycles in working order [3].

However, the growing popularity of bicyclists also entails safety concerns as observed in accident statistics. In 2010, there were 4616 bicyclists dead and 14,283 seriously injured in road accidents in China, representing 7.1% of all traffic fatalities and 5.6% of injuries [4]. One typical type of rule violation behavior is violation behavior at red period. Because of the poor law enforcement and peoples’ low safety awareness, violation behavior at red period is rather prevalent and represents a substantial safety problem in Chinese urban intersections.

Bicyclist safety has been a key topic in traffic safety research. Many studies were conducted on the issues about safety perception and evaluation of cycling [5, 6]. In addition, some researchers mainly focused on bicyclists’ injury and accident analysis in bicycle-car collisions [7]. However, the literature review suggests that very little has been done on the red-light crossing behavior of bicyclists. In recent years there have been only a few empirical studies on investigating bicyclist red-light running behavior. For instance, Johnson et al. used video recordings to analyze the violating rate, behavioral characteristics and risk factors of noncompliant

This paper describes the red-light running behavior of bicyclists at urban intersections based on reliability analysis approach. Bicyclists’ crossing behavior was collected by video recording. Four proportional hazard models by the Cox, exponential, Weibull, and Gompertz distributions were proposed to analyze the covariate effects on safety crossing reliability. The influential variables include personal characteristics, movement information, and situation factors. The results indicate that the Cox hazard model gives the best description of bicyclists’ red-light running behavior. Bicyclists’ safety crossing reliabilities decrease as their waiting times increase. There are about 15.5% of bicyclists with negligible waiting times, who are at high risk of red-light running and very low safety crossing reliabilities. The proposed reliability models can capture the covariates’ effects on bicyclists’ crossing behavior at signalized intersections. Both personal characteristics and traffic conditions have significant effects on bicyclists’ safety crossing reliability. A bicyclist is more likely to have low safety crossing reliability and high violation risk when more riders are crossing against the red light, and they wait closer to the motorized lane. These findings provide valuable insights in understanding bicyclists’ violation behavior; and their implications in assessing bicyclists’ safety crossing reliability were discussed.
commuter cyclists at urban intersections in Melbourne, Australia [8]. Furthermore, Johnson et al. investigated the behavioral, attitudinal, and traffic factors contributing to red-light infringement by Australian cyclists using a national online survey [9]. Wu et al. used logistic model to analyze how gender, age, and conformity behavior affected the cyclist’ violation and classify red-light crossing behavior to three types: obeying the rules, risk-taking, being opportunistic [10]. Zhang and Wu considered the effects of sunshields on red-light running behavior of cyclists and electric-bike riders by using the logistic model [11].

Most of the existing researches on cyclists’ crossing behavior used logistic model to analyze the violating probability and its influence factors. In this paper, we use proportional hazard models to study the reliability of bicyclist safety crossing behavior at the signalized intersection. Hazard model is a common topic in many areas including biomedical, engineering, and social sciences. In the transportation field, it has been applied to describe a number of time-related issues including activity duration [12–14], traffic accidents [15], automobile ownership [16, 17], and vehicle travel time [18, 19]. In addition, Hamed used parametric hazard models to investigate pedestrians’ red-light crossing behavior in Jordan [20]. Tiwari et al. applied nonparametric duration model to study the pedestrians’ crossing behavior at signalized intersections in Indian [21]. Guo et al. used semiparametric and parametric hazard models to analyze the safety crossing reliability of pedestrians in China [22]. Huan and Yang applied the Cox hazard model to estimate the waiting endurance times of electric two-wheelers in China [23]. However, no studies addressed bicyclists’ crossing reliability at signalized intersections.

Hazard model has an advantage in that it allows the explicit study of the relationship between duration time and the explanatory variables. More importantly, hazard model can deal with not only uncensored data but censored data. For example, the exact waiting duration reflecting cyclist endurance cannot be observed if cyclists wait until the permission of traffic rules. This is the very reason why the duration model is chosen to analyze bicyclists’ crossing behavior. The empirical data, which were obtained by video cameras, are modeled by four proportional hazard functions (Cox, exponential, Weibull, and Gompertz distributions). Both crossing reliability and waiting times of bicyclists under various conditions are calculated and the covariate effects are quantified. The finding of this paper can explain when and why bicyclists violate traffic light at the intersection. It is hoped that the results can help to improve the planning and designing of signalized intersections in developing countries.

2. Methods

Reliability is the probability that a system or component will perform its required function under stated conditions for a specified period of time. Accordingly, bicyclists’ crossing reliability can be defined as the probability that a bicyclist obeys the traffic light after a specified waiting time at a signalized intersection.

Let $T$ be a nonnegative random variable representing the waiting duration of a bicyclist in the red-light phase. Let $f(t)$ denote the probability density function of $T$ and let the cumulative distribution be

$$F(t) = \Pr(T \leq t) = \int_0^t f(x) \, dx, \quad t > 0. \tag{1}$$

Mathematically, the reliability function $R(t)$ is the probability that a system will be successfully operating without failure in the interval from time 0 to time $t$,

$$R(t) = \Pr(T > t) = 1 - F(t) = \int_t^\infty f(x) \, dx. \tag{2}$$

In this paper, $R(t)$ is the reliability of bicyclist safety crossing at the red light. It is the probability that a bicyclist who arrives at the intersection in the red-light period will not violate the traffic light in his/her waiting duration time from 0 to $t$.

Another important function in reliability analysis is the failure rate function, or hazard function, $h(t)$. The hazard function is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t \mid T > t)}{\Delta t} = \frac{f(t)}{R(t)}. \tag{3}$$

The hazard function in this paper is the instantaneous rate at which the waiting duration will end in an infinitesimally small time period, $\Delta t$, after time $t$, given that the duration time has lasted to time $t$.

The shape of the hazard function has important implications for the modeling approach. Depending on the underlying event and the duration process, the hazard function may take different shapes. Note that bicyclist crossing behavior at intersections is influenced by various factors. The influential factors can be defined as a vector of explanatory variables, $x = (x_1, x_2, \ldots, x_p)'$. To accommodate the effects of these influential factors, the most commonly used approach to model duration data is the proportional hazard model [12]. Then the proportional hazard form is introduced, which specifies the effects of explanatory variables to be multiplicative on a hazard function

$$h(t) = h_0(t) g(x, \beta), \tag{4}$$

where $g(\cdot)$ is a known function to represent the effects of explanatory variables, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$ is a vector of estimable coefficients for $x$. $h_0(t)$ is the baseline hazard rate at time for covariate vector $x = 0$.

The Cox model is the most commonly used proportional hazard model in which $\exp(\beta' x)$ is used as the function form of the covariate influence. This model defines the hazard rate at time $t$,

$$h(t) = h_0(t) \exp(\beta' x). \tag{5}$$
Combining (3) and (5), the reliability function can be written as
\[ R(t) = \exp \left[ - \int_0^t h(w) \, dw \right] \]
\[ = \left\{ \exp \left[ - \int_0^t h_0(w) \, dw \right] \right\}^{\exp(b'x)} \]
\[ = R_0(t)^{\exp(b'x)}, \]
where \( R_0(t) = \exp[ - \int_0^t h_0(w) \, dw ] \) represents the baseline reliability with \( x = 0 \).

In addition, three parametric proportional hazard models including exponential, Weibull, and Gompertz distributions are used. The common distributions are summarized below.

The exponential distribution is suitable for modeling data with constant hazard. The exponential hazard and reliability functions are
\[ h(t) = \lambda \quad \text{(constant hazard)} \]
\[ R(t) = \exp(-\lambda t). \]

The model is implemented by parameterizing \( \lambda_j = \exp(b_j'X_j) \) when the influenced covariates are considered, implying that \( h_0(t) = 1 \).

The Weibull distribution is suitable for modeling data with monotone hazard rates that either increase or decrease exponentially with time. The Weibull hazard and reliability functions are
\[ h(t) = \lambda \gamma t^{\gamma-1}, \]
\[ R(t) = \exp(-\lambda t^{\gamma}). \]

The model is implemented by parameterizing \( \lambda_j = \exp(b_j'X_j) \), implying that \( h_0(t) = \gamma t^{\gamma-1} \), where \( \gamma \) is an ancillary parameter to be estimated from the data. Note that the hazard reduces to exponential if \( \gamma = 1 \).

The Gompertz regression is parameterized only as a proportional hazard model. The Gompertz distribution implemented is the two-parameter function, with the following hazard and reliability function:
\[ h(t) = \lambda \exp(\gamma t), \]
\[ R(t) = \exp\left(-\lambda \exp(\gamma t-1)\right). \]

The model is implemented by parameterizing \( \lambda_j = \exp(b_j'X_j) \), implying that \( h_0(t) = \exp(\gamma t) \), where \( \gamma \) is an ancillary parameter to be estimated from the data.

The Cox hazard model can be estimated by maximum partial likelihood method. And parametric hazard model can be estimated by maximum likelihood method. The detailed estimation can refer to Lee and Wang [24].

The overall goodness of fit of the model estimation is determined by the likelihood ratio (LR) statistics, which is specified as
\[ X_L = -2 \left[ \log L(\beta_0) - \log L(\hat{\beta}) \right], \]
where \( \log L(\beta_0) \) is the log (partial) likelihood for null model with all the regression coefficients being set as zero and \( \log L(\hat{\beta}) \) is the log (partial) likelihood at convergence with \( p \) regression coefficients.

**3. Data**

**3.1. Site Characteristics.** The site survey was conducted at selected signalized intersections near Jiaotong University in Beijing, China. The survey periods included peak hour (7:30 a.m.–9:30 a.m.) and off-peak hour (10:00 a.m.–4:00 p.m.) in September, 2013. Data collections were done by placing video cameras at each location. The survey area covered the zebra crossing and a part of traffic lanes so that riders’ crossing behavior and the corresponding traffic conditions can be monitored clearly. Figure 1 gives a photograph at Z-T (Zaojunmiao Road-Third Loop North Road, observed direction: west-east) intersection to illustrate the crossing behavior of bicyclists and survey process.

In addition, some additional attentions and explanations are needed for the site survey.

(a) The signals were old traditional person heads so that the influence of type of signals could be neglected. The selected sites had similar characteristics which involved geometric conditions, traffic conditions, and traffic control.

(b) The survey was conducted in good weather and the absence of pointsmen. Bicyclists were unobtrusively observed.

**3.2. Data Collection and Processing.** To record bicyclists’ waiting durations, the whole red-light period of a signal cycle was observed as a data collection unit. Only the bicyclists who arrived in the red-light period were defined as valid sample. The waiting duration was from the time a bicyclist arrived at the crossing location to the time he/she began to cross. It can be classified into two kinds: uncensored data and censored data. The uncensored data was defined as the waiting duration which ended within the red-light period (violating crossing). Otherwise, the waiting duration was called the censored data as long as it ended within the green light period (normal crossing). For the censored data, it is unknown what is the exact waiting duration which can reflect the endurance of waiting time for bicyclists.
3.3. Covariate Selection. The covariate selection takes into account the previous researches and arguments regarding the effects of the exogenous variables and human factors on bicyclists’ violation behavior. Two broad sets of variables are considered as covariates: personal characteristics and traffic conditions. Personal characteristics involve age and gender. The selected covariates of traffic conditions can determine the effects on the waiting time and traffic volume. The practical effects on waiting behavior and the feasibility of data acquisition are considered in the covariate selection. The following covariates, as shown in Table 1, are adopted to construct the reliability model.

4. Results

4.1. Descriptive Statistics. 619 crossing events of bicyclists at the red-light phase were observed by video recordings. Of the valid observations, 320 (51.70%) bicyclists violated the traffic regulations. The average waiting time of all samples was 40.85 seconds. The average waiting time of the violating crossing was 28.06 seconds, while the average waiting time of the normal crossing is 54.53 seconds. There are 15.48% of bicyclists with negligible wait time; their waiting times are less than one second. The maximum waiting duration was 182 seconds while the minimum was 0 second. The latter means people cross the street without any wait. This descriptive statistic cannot reflect the exact crossing reliability due to the neglect of the censored data. The estimation of the crossing reliability with censored data will be discussed later.

4.2. Parameter Estimated Results. Table 2 shows the estimated parameters of proportional hazard models in predicting safe crossing reliability of bicyclists. Four models show perfect overall goodness of fit indicated by the likelihood ratio values. Among them, Cox proportional hazard model is considered best to describe bicyclists’ crossing reliability. The hazard ratios of four models are not time-dependent. Therefore, the ith covariate changes by one unit and the risk changes by \( \exp(\beta_i) \times 100\% \).

From the estimated results, gender has low significant level. Most of other covariates are statistically significant at the 0.05 level of significance. These covariates are significantly related to safety crossing reliability. A positive parameter has a positive effect on the violation hazard and a negative effect on bicyclists’ crossing reliability (e.g., CN, AGE, and WP). Therefore, an increase in the corresponding covariates can increase the risk of bicyclists’ red-light running and decrease their safe crossing reliability. For example, a bicyclist is more likely to violate traffic rules if more bicyclists are crossing against the red light when he/she approaches the intersection. Similar results were found in pedestrians’ crossing behavior. For example, Rosenbloom indicated individual pedestrians were more likely to cross against the red light than pedestrians who were in a groups [25]. Bicyclists who wait closer to the motorized lane have higher likelihoods of red-light running and lower safety crossing reliabilities. The possible reason is that most bicyclists who wait near the motor lane are more likely to be risk-takers. These risk-takers are more apt to cross against the red light than the bicyclists who wait at the appropriate area. Older bicyclists have higher violation risks and lower crossing reliabilities. It is partly because middle-aged and old bicyclists may be considered to have higher cycling skills, which lead to having a higher violation risk.

On the contrary, the negative effect means that an increase in the corresponding covariates can increase the risk rate and reduce the safe crossing reliability (e.g., MV). A bicyclist is likely to wait to reduce the collision risk in the presence of heavy traffic. Similar results were found in pedestrians’ crossing behavior [22] and riders’ red-light running behavior [10].

4.3. Crossing Reliability Analysis of Bicyclists. Bicyclists’ crossing reliabilities calculated by four proportional hazard models are shown in Figure 2. All of the probability curves present a monotonically decreasing trend, which means the crossing reliability decreases with the increasing waiting time. That is to say, bicyclists were more likely to infringe on the traffic lights as their waiting times increase. In addition, the curves of the Cox and Weibull distributions have a rapidly decreasing trend in the early part of the waiting time. It is consistent with the fact that about 15.5% of bicyclists are at high risk of violation and low waiting time to cross against the red right in our sample. Therefore, the exponential and Gompertz distributions have a low adaptability to describe bicyclists’ crossing reliability.

Figure 3 gives the risk rate curve calculated by four proportional hazard models. The constant risk in the exponential model reflects the fact that the risk of violating behavior would not change with the elapse of waiting time. Therefore, the exponential model is not appropriate for describing the crossing reliability of bicyclists. On the other hand, the curves of the Cox, Weibull, and Gompertz distributions show a clear downward trend over waiting time, reflecting a negative dependence. It means that the likelihood of terminating the waiting time decreases with elapsed waiting time. It is consistent with the fact that a part of bicyclists are generally nonrisk-takers who can obey the traffic rules after waiting a long time.

4.4. Reliability Function of Crossing Behavior. Once bicyclists violate traffic light, such a crossing behavior is defined as the occurrence of a risk event. Bicyclists’ crossing reliability is the probability that a bicyclist obey the traffic light after a specified waiting time at the intersection. The Cox and Weibull hazard models give the better description of bicyclists’ crossing behavior than the exponential and Gompertz models. Taking the Weibull model, safe crossing reliability of bicyclists using our sample data can be written as

\[
R(t) = \exp \left( -\lambda t^\gamma \right) = \exp \left( -t^{\gamma} \exp \left( \beta' X \right) \right)
\]

\[
= \exp \left( -t^{0.438} \exp \left( 0.248 \text{AGE} + 0.636 \text{WP1} + 1.950 \text{WP2} + 0.062 \text{VN} - 3.774 \right) \right).
\]

(II)
Table 1: Covariates selection and explanation.

<table>
<thead>
<tr>
<th>Covariate Type</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG (age group)</td>
<td>Continuous variable 1 if under 20, 2 if 20–29, 3 if 30–39, 4 if 40–49, 5 if 50–59, 6 otherwise</td>
</tr>
<tr>
<td>GEN (gender)</td>
<td>Categorical variable 1 if male, 0 female</td>
</tr>
<tr>
<td>VN (violating number)</td>
<td>Continuous variable The number of other bicyclists who violate against the red light after arriving</td>
</tr>
<tr>
<td>WP (waiting position)</td>
<td>Categorical variable 0 if a bicyclist waits behind the bicyclists’ stop line (appropriate), 1 if in the bicycle lane (middle), 2 if close to motorized lane (nearest)</td>
</tr>
<tr>
<td>MV (motor vehicle volume)</td>
<td>Continuous variable Average motor vehicle volume per lane per min at the red-light phase when the bicyclist arrives</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimation in waiting duration model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coef.</th>
<th>P value</th>
<th>Coef.</th>
<th>P value</th>
<th>Coef.</th>
<th>P value</th>
<th>Coef.</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN</td>
<td>0.073</td>
<td>0.546</td>
<td>0.125</td>
<td>0.301</td>
<td>0.126</td>
<td>0.300</td>
<td>0.137</td>
<td>0.260</td>
</tr>
<tr>
<td>AGE</td>
<td>0.245</td>
<td>&lt;0.001</td>
<td>0.302</td>
<td>&lt;0.001</td>
<td>0.248</td>
<td>&lt;0.001</td>
<td>0.286</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>WP middle versus appropriate</td>
<td>0.683</td>
<td>0.003</td>
<td>0.697</td>
<td>0.003</td>
<td>0.636</td>
<td>&lt;0.001</td>
<td>0.679</td>
<td>0.004</td>
</tr>
<tr>
<td>WP nearest versus appropriate</td>
<td>1.919</td>
<td>&lt;0.001</td>
<td>2.172</td>
<td>&lt;0.001</td>
<td>1.950</td>
<td>&lt;0.001</td>
<td>2.126</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CN</td>
<td>0.053</td>
<td>&lt;0.001</td>
<td>0.051</td>
<td>&lt;0.001</td>
<td>0.062</td>
<td>&lt;0.001</td>
<td>0.055</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MV −0.095</td>
<td>0.035</td>
<td>−0.113</td>
<td>0.011</td>
<td>−0.053</td>
<td>0.228</td>
<td>−0.092</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>−5.688</td>
<td>&lt;0.001</td>
<td>−3.774</td>
<td>&lt;0.001</td>
<td>−5.592</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>γ</td>
<td>0.438</td>
<td>−0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>likelihood ratio (LR)</td>
<td>249.77</td>
<td>&lt;0.001</td>
<td>334.48</td>
<td>&lt;0.001</td>
<td>266.89</td>
<td>&lt;0.001</td>
<td>316.26</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The proposed reliability model can capture the covariates’ effects on bicyclists’ crossing behavior at the intersection. Before the applications, however, it is noted that the model should be estimated using the specified field data. Additionally, the explanatory variables should be chosen flexibly according to the specified traffic circumstance.

5. Conclusions

The paper uses the concept of crossing reliability and the methodology of survival analysis to examine bicyclists’ violating behavior at signalized intersections. Four proportional
hazard models by the Cox, exponential, Weibull, and Gompertz distributions are proposed to analyze the covariates’ effects on crossing reliability at urban intersections. The methodology uses a hazard-based duration model structure that is able to predict bicyclists’ crossing reliability at any given traffic conditions. It is proved that the Cox proportional hazard model is the most appropriate for describing crossing behavior of bicyclists. The crossing reliability decreases with the increasing waiting time. About 15.5% of bicyclists are at high risk of violation and without any delay to cross against the red light. In addition, various factors can modify crossing reliability in different degrees and the model can be applied to obtain the temporal shifts in crossing reliability due to changes in traffic conditions. The proposed reliability model can capture the covariates’ effects on bicyclists’ crossing behavior at signalized intersections.

In terms of the future work, research with more datasets is required. Also, more parameters under different situations should be taken into account. In addition, we chose these four proportional hazard models (Cox, exponential, Weibull, and Gompertz) to describe bicyclists’ safety crossing reliabilities in this paper. We will explore accelerated hazard models such as lognormal, logistics, and Gamma distributions for reliability function in future research. Finally, some countermeasures of improving bicyclists’ safety crossing reliabilities should be further explored. It is hoped that these findings may have a better understanding of bicyclists’ crossing behavior and help to plan and design proper facilities for cyclists.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the National Basic Research Program of China (Grant no. 2012CB725401), Beijing Higher Education Young Elite Teacher Project (Grant no. YETP0554), National Natural Science Foundation of China (Grant no. 71131001, 71222101), Fundamental Research Funds for the Central Universities (Grant no. 2014YJ0803), and China Scholarship Council.

References


Submit your manuscripts at http://www.hindawi.com