Investigating the Relationship between Traffic Violations and Crashes at Signalized Intersections: An Empirical Study in China

Ze-Hao Jiang,1 Xiao-Guang Yang,2 Tuo Sun,2 Tao Wang,3 and Zheng Yang3

1Department of Construction Management, School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
2Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 200092, China
3School of Naval Architecture, Ocean & Civil Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Zheng Yang; zyang@citscp.com

Received 1 December 2019; Revised 3 March 2021; Accepted 25 March 2021; Published 16 April 2021

Abstract

About 90% of traffic crashes are caused by human factors, within which traffic violations are one of the most typical and common causes. In order to investigate the relationship between traffic violations and traffic crashes, this research targets signalized intersections in two Chinese cities: Yinchuan and Suqian. Thirty-one intersections are selected as the research sites, and additionally, the traffic volume, traffic violation, and traffic crash data of each intersection are collected for one year. A White’s test is conducted to test the homoscedasticity of the data and a multiple linear regression model is employed to investigate the relationship between traffic crashes and violations. The results show the following: (1) although the research sites are located in two different cities, the data is homoscedastic, which suggests that the above result may be statistically stable between different cities. (2) There is a significant multiple linear regression relationship ($R^2 = 0.782$, adjusted $R^2 = 0.716$) between the total number of traffic crashes and traffic violations. Among the chosen 7 independent variables, four are significantly related to the dependent variable, namely, driving commercial vehicle during internship, wrong-way entry, speeding, and traffic-light violation. (3) With the increase of annual average daily traffic (AADT), the number of total crashes goes up; however, the injury-or-fatality rate decreases, which means that intersections with smaller traffic volumes tend to have higher traffic crash severity. Based on the above conclusions, it is possible to conduct more targeted enforcement to improve the safety of intersections.

1. Introduction

The traffic safety situation around the world is still not optimistic. According to the latest statistics of WHO [1], more than 1.2 million people unfortunately pass away due to traffic crashes every year with an additional 20–50 million non-fatal injuries. In China alone, more than 260,000 people died in traffic crashes which accounts for 21.64% of the worldwide figures [1]. It is worth noting that China’s traffic safety indicator is worse than the world average. Recent data show that China’s traffic crash death rate is 1.88 per 10,000 people, while the world average is 1.68 [1].

Because of the complex traffic conflicts and frequent signal changes, signalized intersections have been proven to be one of the most unsafe locations in the road network. In the United States, 43% of traffic crashes and 23% of fatal traffic crashes occur in/near intersections [2]. Similarly, 30% of Chinese traffic crashes also take place in/near signalized intersections [3]. Hence, improving the safety of intersections has been considered as a top priority by federal, state, and local agencies [4, 5]. In order to reduce traffic crashes, it is crucial to assess the safety of intersections, identify the risk factors, and conduct targeted enforcement.

In the transportation system, human factors are one of the decisive factors leading to traffic crashes. A Policy on Geometric Design of Highways and Streets (“green book”) [6] reports that around 70% of traffic crashes are related to human factors. China’s statistics [7] have also reached a similar conclusion; that is, human factors have caused more than 86% of China’s traffic crashes. Traffic violations are the most typical and harmful unsafe driving behaviors;
therefore, investigating the relationship between the traffic violations and crashes is of both theoretical significance and engineering value. If the most common violations with their relative order of frequency leading to traffic crashes can be found and targeted enforcement can be conducted, the number of overall accidents will be reduced.

With the rapid development of video recognition technology in the past ten years, electronic police equipment has been widely installed by Chinese traffic management authorities at urban road intersections. This technology could automatically record the times, locations, and types of traffic violations. Additionally, in order to ensure the justice of punishment, authorities also manually check the accuracy of electronic-police data. After these procedures, this recorded data with an accuracy of 95% lays the foundation for this study.

In this research, we raise the following two research questions:

Question 1. Is there a significant statistical relationship between traffic crashes and traffic violations? What kind of relationship is it?

Question 2. Based on the above conclusion, what kind of targeted enforcement advice can be given?

### 2. Literature Review

The quantitative research on the relationship between unsafe acts and accidents can be traced back to the classic Heinrich’s Law [8]: as many as 95% of all workplace accidents are caused by unsafe acts, and there is a relationship of one serious injury accident to 10 minor injury (first aid only) accidents, to 30 damage causing accidents, to 600 near misses, to uncountable unsafe acts. Although this conclusion is summarized for a factory workplace, Heinrich’s work is still claimed as the basis for the theory of behavior-based safety.

Research on traffic violations mainly focuses on the following two points. The first concern is the impact of traffic enforcement on traffic crashes, namely, “can traffic enforcement reduce traffic crashes?” Research by Bjørnskau and Elvik et al. [9] shows that road users tend to abide by the law if the police are observed, and violate if there are no police around. This means that most attempts at on-site enforcement will not have lasting effects, either on the road-user behavior or on the crashes. However, Retting et al. [10] suggested that electronic police enforcement can generally reduce violations by an estimated 40–50%. Similarly, Jiang et al. [11] revealed that, in China, the crash volume is reduced by about 40% after installing the electronic police facilities at the intersection, and it also has a deterrent effect on the surrounding intersections [12]. The above conclusions support the utilization of electronic enforcement facilities. Nevertheless, electronic enforcement can also cause some side effects, especially for certain types of traffic crashes. For example, Persaud et al. [13] found that after installing a red-light-running camera, although the right-angle crashes reduced by 26%, the number of rear-end crashes increased by 18%.

The second research hot issue on traffic violation is to investigate the quantitative relationship with traffic crashes. However, due to privacy concerns, the actual traffic violation record data is hard to obtain from traffic management authorities. Conventional studies [14] are often carried out using questionnaires. Related studies [15, 16], meta-analysis [17], and a review article [18] all show a statistically significant positive correlation between traffic violations and traffic crashes. Yoh et al. [19] illustrated that the type of specific illegal behavior is related to the severity of the crash. Wang et al. [7] conducted a descriptive analysis using the traffic violation statistical data in China from 1995 to 2005 and inferred that five kinds of illegal behaviors such as speeding and drunk driving caused 40% of fatal crashes.

Summarizing the above studies, it is not difficult to find that (a) electronic police enforcement can significantly inhibit traffic violations and (b) there is a clear positive relation between traffic violations and crashes. However, (a) much of previous research has been based on self-report data and, therefore, may be affected by social desirability bias. (b) As far as we can find, the relevant researches are concerned about the relationship between a certain type of traffic violation (such as red-light running) and the crash. In fact, the accidents that happen at an intersection in one year are bound to be related to different kinds of traffic violations. Hence, there is a need for further work using a different data source and investigating the relationship between multiple kinds of traffic violations and crashes. In particular, there is a lack of research on traffic violations at intersections, wherein around 30% of traffic crashes happen. Therefore, this paper will conduct further study on traffic violations based on empirical data.

### 3. Data

#### 3.1. Research Site

A total of 31 four-legged signalized intersections in Yinchuan and Suqian, China, are selected as the research sites. Google Earth is utilized to explore the spatial data for assisting in the selection of signalized intersections. The selection criteria of research sites are as follows:

1. The intersections should be located in urban areas of these two cities.
2. The research site should be 90° intersections. Skewed intersections tend to exhibit higher crash rates than 90° intersections [20], and the safety mechanisms for skewed and 90° intersections are also different [21]. Hence, only 90° intersections will be included in this study.
3. The selected intersections should have a high traffic volume in order to maximize the crash sample size and reduce randomness. The annual average daily traffic (AADT) is required to exceed 10,000 PCU.
4. The selected intersection data should be complete. Because of the damage of detectors, especially the traffic-flow detectors, it is not easy to find intersections with complete one-year records of traffic.
volumes, violations, and crashes. Additionally, there are currently no GIS statistical crash data available, so it is necessary to manually extract the crash record. Figure 1 shows the installation and monitoring screen capture of the electronic police.

3.2. Descriptive Statistics. The 2017 annual traffic volume, violation, and crash data for the above intersections are used within this study. Among them, traffic volume and violation data are automatically obtained from the electronic police system, which can also automatically record the traffic volume every 5 minutes. Since the crash analysis systems have not been utilized in these two cities, the traffic crash data were manually extracted from paper archives. The total number of crashes for a signalized intersection is calculated by both at-intersection and influenced-by-intersection crashes [22]. The traffic crashes influenced by intersection are those that occurred in road segments close to an intersection. In this study, a traffic crash that occurred within 150 m from the center of the intersection is classified as the influenced-by-intersection crash [23].

The descriptive statistics of traffic crashes and traffic violations in these two cities are shown in Table 1. The statistics result shows that there were 3,696,659 traffic violations and the ratio of injury-or-fatal crashes, property-loss crashes, and traffic violations was 1:1.32:27 181.31. As shown in Figure 2, compared with Heinrich’s Triangle, it can be found that the ratio of injury-or-fatal crashes to non-injury-or-fatal crashes in the traffic field is larger; that is, the severity of traffic crashes is often higher than that of factory accidents. Additionally, the ratio of traffic crashes and traffic violations in the traffic field is smaller, which indicates that drivers often lead to crashes after multiple illegal operations compared to factory production.

Traffic crashes, in this study, are classified into two types: injury-or-fatal crashes and property-loss crashes. A total of 315 crashes were recorded during the study period, including 135 injury-or-fatal crashes and 180 property-loss crashes, as shown in Figure 3. At the same time, as mentioned before, the electronic police installed at a signalized intersection in China can record multiple types of traffic violations. There are seven types of traffic violations data collected in this paper: designated approach lane violation (1208), driving commercial vehicle during internship (1234), wrong-way entry (1301), against traffic signs (1344), against traffic marks (1345), speeding (0–20% over the speed limit) (1352), and against signals (1625).

4. Methodology

4.1. White’s Test. Since the research sites of this study are distributed in two different cities, it is of considerable importance to explore whether the data are nested. If the data is nested, the variance of the errors in a regression model is not constant. One can also disaggregate the data of intersections in different cities into a general linear regression model, but this will violate the homoscedasticity hypothesis. The test, which is an estimator for heteroscedasticity-consistent standard errors, was proposed by White [24].

White’s test is used to test for heteroscedastic (“differently dispersed”) errors in regression analysis. The null hypothesis for White’s test is that the variances for the errors are equal; that is,

\[ \text{original Hypothesis: } H_0 = \delta_1^2 = \delta_2^2. \]  

And the alternate hypothesis (the one we are testing) is that the variances are not equal:

\[ \text{alternate Hypothesis: } H_1 = \delta_1^2 \neq \delta_2^2. \]  

STATA 12 software package is employed in this study, and the result of White’s test is shown in Table 2.

It can be seen that the \( P \) value is 0.42 > 0.1, which means rejecting the alternate hypothesis \( H_1 \) and accepting the original hypothesis \( H_0 \). That is, data in this research is homoscedastic.

4.2. Multiple Linear Regression Model. Since the data in this paper satisfies the homoscedasticity hypothesis, we consider establishing a multiple linear regression model in which the number of traffic crashes is the dependent variable and the numbers of traffic violations for each violation type are the independent variables. In addition, we also need to test whether the relationship between crashes and violations is heterogeneous between cities, so the city is also introduced into the model as a dummy variable.

\[
\text{Model 1: } Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_i X_i + \mu, \\
\text{Model 2: } Y = \beta_0 + \beta_1 X_1 \cdots + \beta_i X_i + \beta_{i+1} \text{city} + \beta_{i+2} \text{city} \cdot X_1 \cdots + \beta_{2i+1} \text{city} \cdot X_i + \mu',
\]  

where \( Y \) is the number of traffic crashes in the intersection; \( i \) varies from 1 to 7 and corresponds to one of the violation types in this research (see Table 3); \( X_i \) is the number of the \( i \)th traffic violation; \( \text{city} \) is a dummy 0-1 variable; \( \alpha_i \) is the regression coefficient of model 1; \( \beta_i \) is the regression coefficient of model 2; \( \mu \) is the residual of model 1; and \( \mu' \) is the residual of model 2.

5. Results

5.1. Model Results. The parameter estimation results for the model 1 and model 2 are illustrated in Table 3. It can be seen that the \( P \) values of the variables associated with city are greater than 0.1, which indicates that the city variable has
insufficient ability to interpret the dependent variables. As a result, the model 1 is chosen in the end. In model 1, the value of R squared is 0.7814, and the adjusted R squared is 0.7149, which shows that this model has a good fit. We found that there are a total of 4 independent variables (driving commercial vehicle during internship, wrong-way entry, speeding (0–20% over speed limit), and against signals) that are statistically significant enough to cause considerable traffic crashes.

5.2. Discussion. Among the four types of traffic violations, driving commercial vehicle during internship (1234) has the highest traffic crash risk. Its regression coefficient is 0.0422, which indicates that traffic crashes are expected to increase by 0.0422 times for each increase of driving commercial vehicle during internship violations. In China, with the rapid development of public transportation, the drivers holding A1 driving license are in short supply; however, according to the law in China, buses must be

Table 1: Descriptive statistics of variables.

<table>
<thead>
<tr>
<th>Description</th>
<th>Mean</th>
<th>S.D.</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of traffic crashes</td>
<td>10.15</td>
<td>6.88</td>
<td>24</td>
<td>2</td>
</tr>
<tr>
<td>The number of property-loss crashes</td>
<td>5.80</td>
<td>5.38</td>
<td>17</td>
<td>0</td>
</tr>
<tr>
<td>The number of injury-or-fatal crashes</td>
<td>4.38</td>
<td>1.68</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Designated approach lane violation (1208)</td>
<td>20725.06</td>
<td>9318.77</td>
<td>54853</td>
<td>8448</td>
</tr>
<tr>
<td>Driving commercial vehicle during internship (1234)</td>
<td>92.13</td>
<td>42.81</td>
<td>159</td>
<td>27</td>
</tr>
<tr>
<td>Wrong-way entry (1301)</td>
<td>4306.06</td>
<td>2841.62</td>
<td>12190</td>
<td>817</td>
</tr>
<tr>
<td>Against signs (1344)</td>
<td>23894.03</td>
<td>9654.51</td>
<td>48758</td>
<td>10078</td>
</tr>
<tr>
<td>Against marks (1345)</td>
<td>27982.35</td>
<td>10686.43</td>
<td>51806</td>
<td>11890</td>
</tr>
<tr>
<td>Speeding (0~20% over speed limit) (1352)</td>
<td>33696.52</td>
<td>15594.46</td>
<td>97516</td>
<td>14268</td>
</tr>
<tr>
<td>Against signals (1625)</td>
<td>8550.90</td>
<td>3171.46</td>
<td>18284</td>
<td>2973</td>
</tr>
</tbody>
</table>

Some of the variable descriptions are followed by brackets, which is the traffic violation code, based on the “People’s Republic of China Public Safety Industry Standard: Codes for Classification of Traffic Violation Actions (GA/T 16.31-2017).”

Heinrich’s triangle of factory accidents

A similar triangle of traffic crashes

Figure 2: The comparison of Heinrich’s triangle between factory accidents and traffic crashes.
driven by an A1-licence driver, which is the highest level of Chinese driving license system. The A1 driving license cannot be obtained directly in China but must be upgraded from a B-level driving license. At the same time, after passing the A1-license exam, there is a one-year internship period during which driving commercial vehicles such as buses are not allowed. Driving commercial vehicles requires higher driving skills; hence, driving commercial vehicles during internship illegally will lead to higher traffic crash risk.

Table 2: The result of White’s test.

<table>
<thead>
<tr>
<th>Source</th>
<th>Chi-square</th>
<th>Degree of freedom</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heteroscedasticity</td>
<td>31.00</td>
<td>30</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 3: The result of White’s test.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regression coefficient</td>
<td>Std. error</td>
<td>t value</td>
</tr>
<tr>
<td>Designated approach lane violation (1208), X1</td>
<td>0.000147</td>
<td>0.000255</td>
<td>0.57</td>
</tr>
<tr>
<td>Driving commercial vehicle during internship (1234), X2</td>
<td>0.042167</td>
<td>0.017219</td>
<td>2.45</td>
</tr>
<tr>
<td>Wrong-way entry (1301), X3</td>
<td>0.001702</td>
<td>0.000508</td>
<td>3.35</td>
</tr>
<tr>
<td>Against signs (1344), X4</td>
<td>0.000156</td>
<td>0.000156</td>
<td>1</td>
</tr>
<tr>
<td>Against marks (1345), X5</td>
<td>$-7.9E-05$</td>
<td>0.000151</td>
<td>$-0.52$</td>
</tr>
<tr>
<td>Speeding (0~20% over speed limit) (1352), X6</td>
<td>$-0.00027$</td>
<td>0.000148</td>
<td>$-1.82$</td>
</tr>
<tr>
<td>Against signals (1625), X7</td>
<td>0.001286</td>
<td>0.000499</td>
<td>2.58</td>
</tr>
<tr>
<td>City</td>
<td>0.5633981</td>
<td>3.105556</td>
<td>0.18</td>
</tr>
<tr>
<td>City-X1</td>
<td>$-2.70E-06$</td>
<td>0.0003241</td>
<td>$-0.01$</td>
</tr>
<tr>
<td>City-X2</td>
<td>$-0.0152558$</td>
<td>0.0209794</td>
<td>$-0.73$</td>
</tr>
<tr>
<td>City-X3</td>
<td>$-0.0006035$</td>
<td>0.0006589</td>
<td>$-0.92$</td>
</tr>
<tr>
<td>City-X4</td>
<td>0.0001937</td>
<td>0.0001923</td>
<td>1.01</td>
</tr>
<tr>
<td>City-X5</td>
<td>$-0.0001082$</td>
<td>0.0001865</td>
<td>$-0.58$</td>
</tr>
<tr>
<td>City-X6</td>
<td>$-0.0000623$</td>
<td>0.0002778</td>
<td>$-0.22$</td>
</tr>
<tr>
<td>City-X7</td>
<td>0.0003037</td>
<td>0.0007569</td>
<td>0.4</td>
</tr>
<tr>
<td>Constant</td>
<td>$-7.48794$</td>
<td>0.53862</td>
<td>$-2.95$</td>
</tr>
<tr>
<td>R squared</td>
<td>0.7814</td>
<td>0.7149</td>
<td>0.7814</td>
</tr>
</tbody>
</table>

Note: *Significance level of 0.1.
Wrong-way entry (1301) violation is the second highest violation of traffic crash risk. In model 1, the coefficient is 0.0017, which means that, for every 10,000 such traffic violations, it is expected that there will be 17 traffic crashes. Wrong-way entry generally leads to a frontal collision that can cause serious casualties. Wrong-way entry behavior at intersections in China is usually due to the large size of intersections and the lack of guiding marks inside. In particular, if there is no median between the entrance and exit lane in one leg of the large-sized intersection, it is easier for drivers to drive into the entrance lane after turning left and eventually lead to frontal collision with the approaching/queuing vehicles.

Similarly, another dangerous traffic violation is going against signals. The regression coefficient of this variable is 0.0013, which shows that each against-signal violation will result in 0.0013 crashes. The against-signal behavior recorded by the electronic police in China is actually red-light running. Numerous other studies [25–27] have also shown a significant positive correlation between red-light running and traffic crashes.

Some researches show positive correlations between speeding and traffic crashes. However, in this study, an interesting conclusion is obtained that the regression coefficient is negative. This should be related to the unreasonable speed management of Chinese urban roads. According to the code of urban road design in China, the design speed at intersections should be 0.5~0.7 times that of the adjacent road segment, as shown in Figure 4, and the traffic management departments often use this value as the speed limit at intersection. At the same time, Chinese drivers are taught to slow down when driving through intersections, even if they are facing the green light [28]. The speed limit at intersections thus leads to overloading of the driving task [29] and a decrease of speed smoothness [30]. Therefore, it can be considered that, under the premise of unreasonable speed limit, a proper increase in the speed limit at intersections may lead to a reduction in the number of crashes.

Additionally, we find that there is no statistically significant correlation between against signs/marks and traffic crashes. The so-called against signs/marks in these two cities essentially refer to an uncivilized driving behavior called “cutting in lane,” that is, approaching the intersection from the adjacent lane, crossing the lane-separation mark, and forcing through the way between the queuing vehicles. However, since this situation is common in some Chinese cities, drivers are likely to have the concept of defensive driving [31]; as a result, such traffic violation will not significantly lead to traffic crashes.

In addition, here we use the multiple linear regression model to describe the relationship between the AADT and the number of traffic crashes, as shown in Figure 5. It can be seen that there are significant positive linear correlations between the number of property-loss/injury-or-fatal crashes and the AADT of intersections. An understandable explanation may be that a high intersection of AADT means a higher probability of traffic congestion, which leads to the increase of illegal lane changes. Therefore, the crash rate increases. However, it is worth noting that the relationship between them is the exponential model in the highway safety manual [32].

When we use the crash rate instead of the crash number as the dependent variable in the model, it is interesting to find that although the crash rate of property-loss crash is still positively correlated with the AADT, however, the relationship between the AADT and the crash rate of injury-or-fatal crash rate is negative, as shown in Figure 6. This reveals...
that, in this research, intersections with smaller traffic volumes often have higher traffic crash severity. This may be due to the fact that the smaller traffic volume can significantly increase the speed of the vehicle, thereby increasing the severity of the crashes [30].

6. Conclusions and Recommendations

According to our dataset, there is a significant linear relationship between traffic crashes and violations. At the same time, the city-related variables are not significant in our research. This suggests that the above result may be statistically stable between different cities.

The model shows that four kinds of traffic violations can significantly lead to traffic crashes, namely, driving commercial vehicle during internship, wrong-way entry, speeding, and traffic-light violation. Based on the above conclusions, the traffic management authorities can be recommended to conduct more targeted enforcement and reasonable speed management measures. Additionally, some countermeasures can also be taken. For example, at intersections with high frequency of wrong-way entry violation, guide marks can be drawn and a median can be set between the approach and exit lane to reduce the crashes caused by this traffic violation.

At the same time, we found that the total number of traffic crashes increased with higher AADT at the intersection. However, the injury-or-fatal rate of the crash decreased with the increase of the AADT. This means that intersections with smaller traffic volumes have higher traffic crash severity. Therefore, this suggests that, in order to improve the overall safety of the road network, it is necessary to invest management resources not only at the intersection of large traffic volume, but also at the intersections of small traffic volume, such as intersections at rural or suburban areas.

Since the traffic crash and violation data is not available to the public in China, acquisition of research data is difficult; hence, the sample size in this study consists of only two cities. In the future, more data will be collected from different cities to verify the relationship.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors express their sincere gratitude National Natural Science Foundation of China (61773293) and Chinese Government Scholarship (201806260148) for the support.

References


