ELSEVIER

Accident Analysis and Prevention 37 (2005) 417–425

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Exploring the overall and specific crash severity levels at signalized intersections

Mohamed Abdel-Aty*, Joanne Keller

Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450, USA

Received 19 September 2004; received in revised form 24 November 2004; accepted 29 November 2004

Abstract

Many studies have shown that intersections are among the most dangerous locations of a roadway network. Therefore, there is a need to understand the factors that contribute to injuries at such locations. This paper addresses the different factors that affect crash injury severity at signalized intersections. It also looks into the quality and completeness of the crash data and the effect that incomplete data has on the final results. Data from multiple sources have been cross-checked to ensure the completeness of all crashes including minor crashes that are usually unreported or not coded into crash databases. The ordered probit modeling technique has been adopted in this study to account for the fact that injury levels are naturally ordered variables. The tree-based regression methodology has also been adopted in this study to explore the factors that affect each severity level. The probit model results showed that a combination of crash-specific information and intersection characteristics result in the highest prediction rate of injury level. More specifically, having a divided minor roadway or a higher speed limit on the minor roadway decreased the level of injury while crashes involving a pedestrian/bicyclist and left turn crashes had the highest probability of a more severe crash. Several regression tree models showed a difference in the significant factors that affect the different severity types. Completing the data with minor non injury crashes improved the modeling results and depicted differences when modeling the no injury crashes.

Keywords: Signalized intersections; Crash severity; Ordered probit models; Tree-based regression; Traffic safety data

1. Introduction

1.1. Background

Intersections are a common place for crashes, which may be due to the fact that there are several conflicting movements as well as a myriad of different intersection design characteristics. Intersections also tend to experience severe crashes due to the fact that some of the injurious crashes such as angle and left turn collisions commonly occur at intersection. During 1999, there were 243,409 crashes recorded in the Florida Crash Database. Of these, 98,756 crashes (about 40%) occurred at or were influenced by a signalized intersection. About 9.6 crashes occur at signalized intersections per year compared to 2 per year where stop or yield signs control traffic. The factors affecting injury levels of crashes occurring at signalized intersections are not well understood. Therefore, there is a need to identify the effects that certain geometric and crash-specific aspects have on the injury level of crashes occurring at signalized intersection.

Furthermore, when a crash occurs and the local police department is notified, the responding officer will determine whether to fill out a long- or short-form crash report. For instance, if a crash involves an injury or a felony (e.g., hit and run), the crash must be filed on a long-form. If a crash involved only property damage (a minor crash with no injuries), usually it is up to the officer to report it on a long- or a short-form. Crash forms are then forwarded to the respective counties. From here, only the crashes reported on long-forms are forwarded onto the Florida Department of Transportation (FDOT) and the Department of Highway Safety and Motor

* Corresponding author. Tel.: +1 407 823 5657; fax: +1 407 823 3315.
E-mail address: mabdel@mail.ucf.edu (M. Abdel-Aty).

0001-4575/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2004.11.002

Vehicles (DHSMV), which maintain electronic records based on only crashes reported on long-forms. By focusing on state-maintained databases, most non injury crashes are neglected. Ignoring these non injury crashes will bias the distribution of crash injury severity levels. Therefore, for this study, two databases were considered. The first database consisted of crashes from the state agencies (reported only on long-forms) and was considered the restricted dataset. The second database consisted of all crashes (reported on both long- and short-forms) and was completed by obtaining all crashes reported on short-forms (we here refer to it as the complete dataset). Furthermore, multiple databases were cross-checked to ensure that the crashes reported on long-forms are as complete as possible (it is worth noting that even the complete data set does not include non reported crashes).

This study explores the hypothesis that crash injury levels are affected by both crash-specific and intersection-specific variables. Furthermore, the authors investigate the significant differences in the important crash-related factors between models based solely on crashes reported on long-forms and models based on crashes reported on both long- and short-forms (i.e., models based on restricted and complete datasets). Additionally, several databases were cross-checked to ensure the completeness of our data. The authors anticipate that these results will provide a significant contribution to the area of safety at signalized intersections as well as consider the possible consequences of the common practice of analyzing restricted datasets. Separate tree regression models for crashes of every severity level were also estimated to identify the significant factors that affect each. Injury levels are categorized as follows: no injury (property damage only), possible injury (no visible signs of injury), non-incapacitating injury (any visible injuries, e.g., bruises or limping), incapacitating injury (any visible signs of injury and the person is carried from the scene) and fatal injury (an injury sustained in a motor vehicle crash that results in death within 90 days).

2. Relevant studies

Many previous studies have used the ordered probit modeling methodology to study injury severity at different roadway locations. For example, Abdel-Aty (2003) applied the ordered probit models to predict crash injury severity on roadway sections, signalized intersections and toll plazas. Jianming and Kockelman (2004) used the ordered probit technique to predict injury severity based on factors including traffic, roadway and occupant characteristics and weather conditions at the time of a crash and type of vehicle. Kweon and Kockelman (2003) showed that wearing seatbelts decreases the risk of injury in crashes on highways.

Khattak and Targa (2004), Khattak et al. (2002, 2003) used ordered probit models to predict the injury level for crashes occurring at construction zones and involving trucks, to predict injury severity for single-vehicle truck rollovers, and to determine vehicle, roadway, driver, crash, and environmental characteristics that influence the severity level of older drivers involved in crashes, respectively.

Duncan et al. (1998) determined the specific variables that influenced the injury levels in two-vehicle (truck and car) rear-end crashes on divided roadways. Zajac and Ivan (2003) determined the effect of roadway and area features on the severity of pedestrian crashes in rural areas. Renski et al. (1999) revealed the effect that increases in the speed limit on interstate highways has on injury severity. Again, the ordered probit modeling methodology was the chosen technique in all of these studies. O'Donnell and Connor (1996) created two ordered probit models to predict the injury levels for crashes in Australia. Increases in both the age of injured person and the speed of vehicle caused a greater injury level.

It could be seen from this section that many studies have adopted the ordered probit modeling methodology to account for the natural ordering of the severity levels. In this study, not only the ordered probit will be adopted to depict the factors that affect the overall severity level, but the regression tree approach will also be attempted to show the importance of the significant factors for each severity level.

3. Data collection

Data collection began in early 2003 when several counties across the midsection of the State of Florida were contacted for cooperation. Four agencies that comprise a majority of Central Florida were identified and contacted: Brevard County, Seminole County, City of Orlando and Hillsborough County. Each jurisdiction provided drawings for several hundred intersections and each drawing was then individually examined and identified by the authors. Information obtained from each drawing was the number of through lanes on each roadway, the number of left turn lanes and whether they were exclusive, the presence of medians on each approach, whether any of the right turns were channelized and the speed limits. Each county also provided a database of crashes reported on both long- and short-forms for 3 recent years. In the meanwhile, crashes reported on long-forms were also downloaded from FDOT and DHSMV databases and cross-referenced against the crashes reported on long-forms provided by the counties. This process served as a check to ensure that each county's database was accurate. It was found that no database by itself was complete and each was missing crashes that another database included. Finally, crashes reported on long-forms from county, FDOT and DHSMV databases were combined with crashes reported on short-forms from the counties' databases to ensure that the dataset for this analysis was complete as much as possible. The master database created for this analysis includes 33,592 crashes from 832 intersections. Table 1 summarizes the data that was collected. Only data from years 2000 and 2001 would be used in model estimation (those were the only consistent years among the four jurisdictions). Table 1 shows that all counties have more crashes reported on short-forms