Accident Analysis and Prevention 106 (2017) 305–314

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Cross-classified multilevel models for severity of commercial motor vehicle crashes considering heterogeneity among companies and regions

Ho-Chul Park[a], Dong-Kyu Kim[a,*], Seung-Young Kho[a], Peter Y. Park[b]

[a] Department of Civil and Environmental Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, 08826 Seoul, Republic of Korea
[b] Department of Civil Engineering, Lassonde School of Engineering, York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3

ARTICLE INFO

Keywords:
Commercial motor vehicle safety
Cross-classified multilevel model
Type I statistical error
Fatal and injury severity
Logit model

ABSTRACT

This study analyzes 86,622 commercial motor vehicle (CMV) crashes (large truck, bus and taxi crashes) in South Korea from 2010 to 2014. The analysis recognizes the hierarchical structure of the factors affecting CMV crashes by examining eight factors related to individual crashes and six additional upper level factors organized in two non-nested groups (company level and regional level factors). The study considers four different crash severities (fatal, major, minor, and no injury). The company level factors reflect selected characteristics of 1,875 CMV companies, and the regional level factors reflect selected characteristics of 230 municipalities. The study develops a single-level ordinary ordered logit model, two conventional multilevel ordered logit models, and a cross-classified multilevel ordered logit model (CCMM). As the study develops each of these four models for large trucks, buses and taxis, 12 different statistical models are analyzed. The CCMM outperforms the other models in two important ways: 1) the CCMM avoids the type I statistical errors that tend to occur when analyzing hierarchical data with single-level models; and 2) the CCMM can analyze two non-nested groups simultaneously. Statistically significant factors include taxi company's type of vehicle ownership and municipality's level of transportation infrastructure budget. An improved understanding of CMV related crashes should contribute to the development of safety countermeasures to reduce the number and severity of CMV related crashes.

1. Introduction

1.1. Problem statement

This study provides a data-driven scientific analysis of the factors that contribute to commercial motor vehicle (CMV) safety in South Korea. We used an advanced multilevel model known as a cross-classified multilevel model to examine factors that may be useful in explaining different severity levels in crashes involving CMVs. We also compared the results of our cross-classified multilevel model with the results obtained from a traditional single-level model and a conventional multilevel model.

Transportation engineers have various CMV classifications depending on the specific purpose of an agency and/or study. For instance, the US Department of Transportation (DOT) defines a CMV as "…a motor vehicle or combination of motor vehicles used in commerce to transport passengers or property…" They classify CMVs into four categories: by weight (two categories), by number of passengers that can be transported, and by materials transported (all hazardous material transportation defines the vehicle as a CMV) (USDOT, 2008). The US DOT uses this approach for various purposes including the reporting of crash statistics involving a CMV. Chatterjee and Cohen (2004) classified CMVs into three main groups: commercial passenger vehicles, freight vehicles, and service vehicles. The classification included 12 detailed categories. Commercial passenger vehicles, for example, included school buses, rental cars, taxis, etc. Freight vehicles included emergency vehicles, such as police cars and fire engines, and construction transportation, and service vehicles included public service vehicles. Chatterjee and Cohen used their classification when forecasting future CMV demand as part of an urban transportation planning study.

In this study, we defined a CMV as simply a motor vehicle used to carry goods and/or passengers for a commercial purpose. We used three categories of CMV: 1) large trucks (e.g., heavy vehicles), 2) buses (e.g., inter-city and intra-city public transit buses), and 3) taxis.

Transportation engineers and researchers agree that CMV safety needs further improvement (Miaou et al., 1992; Zegeer et al., 1994; Park et al., 2005; Jovanis et al., 2005). The US Federal Motor Carrier Safety Administration (FMCSA) reported 329,000 CMV crashes across America in one year (2011). These crashes resulted in 116,018

* Corresponding author.
E-mail addresses: cheerul8@snu.ac.kr (H.-C. Park), dongkyukim@snu.ac.kr (D.-K. Kim), sykho@snu.ac.kr (S.-Y. Kho), peter.park@lassonde.yorku.ca (P.Y. Park).

http://dx.doi.org/10.1016/j.aap.2017.06.009
Received 30 December 2016; Received in revised form 21 April 2017; Accepted 13 June 2017
0001-4575/

causalities and an estimated cost of $87 billion. The FMCSA figures include crashes involving large trucks and buses, but not taxis (FMCSA, 2013). In South Korea, crashes involving large trucks and buses averaged 114,465 per year from 2010 to 2014. This is a huge number of crashes for a country with a total population (2014) of about 50.42 million people (about 15% of the population of the United States) and comparable in area to the state of Indiana. During the same five-year period, taxi crashes in South Korea averaged an additional 94,416 crashes per year.

In 2013, crashes involving large trucks, buses and taxis in South Korea resulted in 77,295 causalities and an estimated cost of $3.8 billion. South Korea recognizes that the number of crashes involving CMVs needs to be addressed.

Crashes involving CMVs are known to have particularly severe consequences. This is partly due to the physical characteristics of CMVs including their heavy weight, large size, and maneuvering limitations (e.g., large minimum turning radius), and the potentially hazardous materials carried by large trucks. In addition, CMV drivers are often on the road for great distances and long travel times which may result in a higher risk of fatigue related crashes for CMV drivers, and the commercial nature of CMV journeys may lead CMV drivers to drive in adverse weather conditions that other drivers might choose to avoid.

Many studies have examined injury severity in CMV crashes, especially in crashes involving heavy vehicles such as large trucks and buses. Much of the research was based on a choice modeling approach (an ordered-response discrete-choice model (either probit or logit)) to establish risk factors that contribute to injury severity in CMV crashes (Chang and Mannering, 1999; Khorashadi et al., 2005; Lemp et al., 2011; Zhu and Srinivasan, 2011; Islam et al., 2014). Chang and Mannering (1999), for example, developed a set of nested logit models for truck-involved crashes and non-truck-involved crashes to identify risk factors unique to trucks. Khorashadi et al. (2005) and Islam et al. (2014) applied logit analysis and found significant differences in the risk factors for CMV crashes in rural and urban areas. Many other statistical approaches have also been used to examine factors that affect injury severity in CMV crashes. These approaches included multiple regression (Elvik, 2002) and logistic regression (Häkkänen and Summala, 2001; Boufous and Williamson, 2009).

It is clear from the literature that an understanding of the factors associated with CMV crashes, particularly fatal and injury crashes, is an important issue. We hope that our analysis based on cross-classified multilevel model will be a useful addition to existing studies and will provide a better understanding of the nature of CMV crashes. A better understanding will lead to the introduction of the safety countermeasures most likely to reduce the severity of CMV crashes.

1.2. CMV crashes: potential contributing factors

Crash data that include both individual and group level factors are called hierarchical crash data and need special care in the screening of the crash factors that potentially affect the severity of crashes. However, most research into CMV crashes has focused on the characteristics of the driver, vehicle and roadway environment without clear consideration of the hierarchical structure of the crash data.

Driver related factors include age, gender, drug and alcohol use, speeding, fatigue, misjudgment, carelessness, etc. Vehicle related factors include vehicle type, vehicle age and failure in mechanical components. Roadway environment related factors include adverse weather conditions such as rain and snow, and road alignment issues such as curvature and gradient (Cantor et al., 2010; Murray and Lantz Keppler, 2005; Chang and Chien, 2013; Blower et al., 2009). In our study, driver, vehicle and roadway environment factors are known as individual level factors or level 1 factors.

Some studies have also used an additional set of CMV crash factors referring to the nature of the companies that own and operate the CMV involved in the crash. These company level CMV crash factors are designed to include the working conditions of CMV drivers, for example, the effect of the company's occupational health and safety management practices on the drivers. Li and Itoh (2013) analyzed road crashes and working conditions information for 18 trucking companies and considered factors that could potentially contribute to crashes. The factors included emotional stability, safety attitudes, delivery area or range, driver workload, and driving experience. Li and Itoh suggested that safety-oriented work schedules and safety-related attitudes and behaviors were necessary to improve CMV safety. In our study, company factors are known as group level factors or level 2 factors.

Group level crash factors can also include the demographic and socio-economic characteristics of the area in which the CMV crash occurs. Khorashadi et al. (2005), for instance, found significant differences between urban and rural crashes involving large trucks. Huang and Abdel-Aty's (2010) group level crash factors included road density, spatial features, regional factors such as traffic regulations, and socio-economic features specific to the region. As it appears reasonable to expect that group level crash factors such as company specific issues, traffic regulations and other region specific characteristics might affect the severity of crashes associated with a specific company or region, this study gives special attention to company and region as group level factors.

1.3. Methodological challenge

Analyzing hierarchically structured crash data with traditional single-level models that assume that the residuals from the model are independent across subjects may result in standard errors that are underestimated and confidence intervals that are too narrow (Kreft and De Leeuw, 1998; Dupont et al., 2013). Statistically insignificant input factors may then be included in the model. This may lead to the incorrect rejection of a true null hypothesis, i.e., a typical example of a type I error in statistical hypothesis testing. Simply said, a type I error refers to detecting an effect that is not present (a false positive).

To overcome this problem, transportation engineers have turned to multilevel modeling using various hierarchical data structures. One of the most common approaches to multilevel modeling uses the geographical location of a crash (e.g., corridor, region and county) as the different levels of the hierarchical data analysis (Guo et al., 2010; Xie et al., 2013; Dupont et al., 2013). Another typical approach defines the purpose of the vehicle or its occupants as the lower level of the data and the road characteristics (e.g., intersection or segment) or regional characteristics (e.g., municipality or region) as the higher level of the data (Jones and Jørgensen, 2003; Lenguerrnad et al., 2006; Helai et al., 2008; Dupont et al., 2010; Quddus, 2015; Kim et al., 2007). All these studies used a conventional multilevel model.

Conventional multilevel models, which include two-level or three-level models, have limitations with respect to the analysis of cross-classified hierarchical data structure. To deal with cross-classified hierarchical data structures, cross-classified multilevel models (CCMMs) have been used in various fields. In education, Simonite and Brown (2003) applied a CCMM to identify the effects of students' background and learning environments on the students' performance. Dunn et al. (2015) used CCMMs to disentangle school and neighborhood effects on adolescent smoking behavior, and pointed out that a conventional multilevel model could lead to overestimation of the effect of the non-nested upper level factors. In medicine, Muntaner et al. (2006) used a CCMM to study the effect of county, organizational, and workplace level factors on depression disorders among nursing assistants. CCMMs have also been used in studies involving various sets of non-nested data including housing prices (Uyar and Brown, 2007), crime (Johnson, 2012), and veterinary medicine (Aunsmo et al., 2009). However, few studies have used CCMMs in traffic safety.

Fig. 1 contrasts conventional and cross-classified hierarchical structures for crash data. Fig. 1(a) shows a typical conventional multilevel model designed to handle a single hierarchical data structure. In