Accident Analysis and Prevention 43 (2011) 49–57

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A comprehensive analysis of factors influencing the injury severity of large-truck crashes

Xiaoyu Zhu[a], Sivaramakrishnan Srinivasan[b,*]

[a] Department of Civil and Coastal Engineering, University of Florida, 512 Weil Hall, PO Box 116580, Gainesville, FL 32611, USA
[b] Department of Civil and Coastal Engineering, University of Florida, 513-A Weil Hall, PO Box 116580, Gainesville, FL 32611, USA

ARTICLE INFO

Article history:
Received 13 May 2010
Received in revised form 15 June 2010
Accepted 12 July 2010

Keywords:
Large-truck crash
Injury severity measures
LTCCS
Ordered-probit model

ABSTRACT

Given the importance of trucking to the economic well being of a country and the safety concerns posed by the trucks, a study of large-truck crashes is critical. This paper contributes by undertaking an extensive analysis of the empirical factors affecting injury severity of large-truck crashes. Data from a recent, nationally representative sample of large-truck crashes are examined to determine the factors affecting the overall injury severity of these crashes. The explanatory factors include the characteristics of the crash, vehicle(s), and the driver(s). The injury severity was modeled using two measures. Several similarities and some differences were observed across the two models which underscore the need for improved accuracy in the assessment of injury severity of crashes. The estimated models capture the marginal effects of a variety of explanatory factors simultaneously. In particular, the models indicate the impacts of several driver behavior variables on the severity of the crashes, after controlling for a variety of other factors. For example, driver distraction (truck drivers), alcohol use (car drivers), and emotional factors (car drivers) are found to be associated with higher severity crashes. A further interesting finding is the strong statistical significance of several dummy variables that indicate missing data – these reflect how the nature of the crash itself could affect the completeness of the data. Future efforts should seek to collect such data more comprehensively so that the true effects of these aspects on the crash severity can be determined.

1. Background and objectives

The importance of trucking to freight logistics and its impact on the economic well being of a nation is well acknowledged in the literature. Based on the 2007 Commodity Flow Survey, among all the modes, trucks moved 70.7% of all freight by value, 68.8% by weight, and 39.8% by ton-miles (USDOT/BTS, 2008). The substantial volume of truck traffic, the unique operating characteristics of the trucks and its drivers, and the design/weight-related issues of trucks have all contributed to the large numbers of crashes, injuries, and fatalities. In 2005, over 5000 people died and an additional 114,000 were injured in the 442,000 large-truck (gross vehicle weight rating greater than 10,000 pounds) crashes in the United States. Approximately 12% of all traffic fatalities involved a large-truck crash (NHTSA, 2006). In 2007, 413,000 large trucks were involved in traffic crashes resulting in 4808 fatalities, which count 12% of the total fatality of all crashes (NHTSA, 2008). Large-trucks account for approximately 4% of all the vehicles but are about 8% of vehicles in fatal crashes.

In addition to all the above cross-sectional statistics, time-series trends reported by Lyman and Braver (2003) are also illuminating. Based on aggregate data from 1975 to 1999, these authors find that the involvement of large-trucks in fatal crashes per truck vehicle-mile-traveled has decreased. However, with a corresponding increase in the volume of truck travel, the involvement per unit population has not seen the same declining trend. Hence, the authors argue that there is continued public concern about large-truck crashes.

The above statistics clearly underscore the need for studying large-truck crashes towards improving the safety of the transportation system. Further, the results from such studies will be valuable in transportation policy, improvement of carrier operation, and incident-cost reduction. The objective of this study is to contribute towards that end. Specifically, data from a recent, nationally representative sample of large-truck crashes are examined to determine the factors affecting the overall injury severity of these crashes. The explanatory factors include the characteristics of the crash, vehicle(s), and the driver(s).

The rest of this paper is organized as follows: a synthesis of literature is presented in Section 2. The data sources and the sample-formation procedure are discussed in Section 3. The statistical-modeling methodology is outlined in Section 4. The

* Corresponding author. Tel.: +1 352 392 9537x1456; fax: +1 352 392 3394.
E-mail addresses: shuxy03@gmail.com (X. Zhu), siva@ce.ufl.edu (S. Srinivasan).

0001-4575/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2010.07.007

model results are presented and discussed in Section 5. Finally, Section 6 presents a summary of work accomplished, highlights the key contributions, and identifies the areas of future research.

2. Literature review

A synthesis of literature on large-truck crashes, with particular focus on the analysis of injury severity of such crashes is presented in this section. A large truck is defined as a commercial vehicle weighing more than 10,000 pounds. It is useful to acknowledge that the body of literature on modeling injury severity, in general, is very extensive, but substantively focused on passenger-vehicle crashes.

Work undertaken by Khattak and colleagues (Duncan et al., 1998; Khattak et al., 2003) and Chang and Mannering (1999) are most directly related to our efforts.

Duncan et al. (1998) examined the injury sustained by passenger-car occupants in the case of rear-end collisions between heavy trucks and passenger cars. Ordered-probit models were developed using the Highway Safety Information System (HSIS) data from North Carolina for the years 1993–1995. The results indicate that higher speeds (and speed differentials), darkness, and grade increase the severity of the injury. Females and drunk drivers were estimated to sustain more severe injuries compared to male and non-drunk drivers, respectively. Snowy/icy road conditions and traffic congestion were found to decrease the effect of the injury severity compared to respectively dry and free-flow traffic conditions. Finally, the car being struck in the rear was found to lead to more severe injuries compared to the truck being struck in the rear.

Khattak et al. (2003) used the HSIS data from North Carolina for the years 1996–1998 to examine the injury severity of single large-truck crashes. In particular, the intent was to examine the differences between rollover and non-rollover crashes. Using ordered-probit models, the authors found that rollover leads to more severe injuries in single-truck crashes. Further, dangerous driving behavior such as drug/alcohol use, and speeding, and not wearing seat belts were found to increase the injury severity.

Chang and Mannering (1999) modeled the vehicle occupancy and the most severe injury sustained by an occupant of the vehicle using data from the state of Washington. Unlike, the previous efforts discussed, Chang and Mannering adopt an un-ordered discrete-choice structure to model injury severity. The authors segmented the data into truck-involved and non-truck-involved crashes and demonstrated the statistical and empirical validity of such segmentation. For example, the results indicate that higher speeds are strongly associated with more severe crashes when trucks are involved (the effect was insignificant in the case of non-truck crashes). Similarly, the effects of turning movements (right turn and left turn) of the vehicles on the crash severity were also found to be different. Consistent with expectations, the results also indicated that multi-occupant vehicles in truck-involved crashes result in significantly severe injuries. Overall, these authors argue that counter-measures aimed at reducing the severity of truck-involved crashes could be different from those aimed at reducing the severity of non-truck crashes.

In contrast to the previous three studies which have examined the level of injury severity, other researchers have focused on fatal crashes involving large-trucks. Braver et al. (1996) examined the effect of roadway geometry, weather, and other factors on the incidence of fatal large truck–car crashes. Defiance of traffic-control devises, curves, slippery and roadway conditions were some of the conditions found to be associated with fatal crashes. Campbell (1991) examined the impact of driver age on the involvement in fatal crashes. Based on nationally representative data for the years 1980–1984, younger drivers (age <27 years) were found to be over-involved in fatal crashes. Further, the relative risk of very young drivers (less than 21 years of age) was found to be about six times the overall risk for all drivers. Golob et al. (1987) examined the severity (both injury severity and incident duration) of truck-involved freeway accidents. About 9000 crashes from the years 1983–1984 were obtained from TASAS (Traffic Accident Surveillance and Analysis System) database maintained by the California Department of Transportation. All data were from the Los Angeles area. Based on the number of fatalities per accident, the "hit-object" type crashes were found to be most dangerous (0.025 fatalities per accident). "Rear end" and "other type" (other than hit-object, sideswipe, broad-side, and overturn) of crashes were also very dangerous (0.021 fatalities per accident).

It is useful to mention here that past studies have also examined other aspects of large-truck safety (other than injury severity). For example, research undertaken by Blower et al. (1993) examined the factors affecting the crash propensities (or the risk of being involved in a crash) and show that truck crash-rate is significantly affected by truck configuration, location (rural or urban), traffic density, and time of day. Hallmark et al. (2009) focused on the incidence of a specific type of crash – the lane-departure crashes. Using logistic-regressions, the authors identify that such crashes were more likely to happen when driver is fatigued, upset, distracted, or unfamiliar with the roadway. More generally, driver fatigue has been recognized as an important factor affecting truck crashes. Based on a survey conducted in New Zealand, Gander et al. (2006) identified 7.6% of crashes were identified as fatigue-related. The duration of the most recent sleep period was considered as a measurement of fatigue. Although the primary intent of our paper is on injury severity (conditional on a crash) and not on the risk of a crash happening, insights from studies discussed above are useful and appropriate explanatory variables (such as fatigue) will be included in our models.

Overall, the literature on the injury severity of large-truck crashes appears to be reasonably limited. Past studies have focused on specific types of crashes (such as rollover or rear end) or on specific injury severity levels (such as fatal crashes). In this context, the intent of this paper is to present a comprehensive analysis of the injury severity of large-truck crashes. The injury severity of a crash is defined as the highest level of severity among all those injured in the crash. The data to be used in this study come from the Large Truck Crash Causation Study (LTCCS – discussed in more detail in Section 3). The database assembled by this study augments the conventional crash-data obtained from police reports in several ways. For instance, additional data related to "human factors" such as the fatigue, illness, and distraction of the drivers was collected. Historical records on the safety of the drivers, vehicles, and carriers (past violations and citations) involved in the crashes were also obtained and added to the crash data obtained from the police accident reports. Thus, the database available for this study would enable the development of a rich empirical model that simultaneously captures the marginal effects of a variety of factors on injury severity.

3. Data

This study uses data from the LTCCS released in 2006. These data represent a sample of large-truck crashes that occurred between April 2001 and December 2003. Data on approximately a thousand crashes were collected from 24 sites in 17 states. Each crash in the LTCCS sample involves at least one large truck and resulted in injury to at least one person. These data were collected by the Federal Motor Carrier Safety Administration (FMCSA) and the National Highway Traffic Safety Administration (NHTSA) of the U.S. Department of Transportation (DOT). USDOT/FMCSA (2006) and Hedlund and Blower (2006) provide an overview of the data. The full data and related documentation are available for download from the LTCCS website at: http://ai.fmcsa.dot.gov/ltccs/default.asp.

A unique aspect of the LTCCS database is that the information from the traditional police crash reports was augmented with site investigations, reviews of hospital records and coroners' reports, and interviews of motor carriers, drivers, passengers, and crash witnesses. The interviews are of particular interest as these provide several behavioral (i.e., human factors) descriptors. For example, drivers were questioned about their inattention and their "distraction" behavior was determined based on several actions such as "talking to other occupants, dialing a phone, adjusting the sound system/air-conditioning, trying to retrieve an object, looking at a previous crash, and approaching traffic or a street-address". Descriptions of all such human-factor variables used in this paper are presented in Table 1. The reader is referred to USDOT (2006a, b) for further details. All data collection was performed at each crash site by a two-person team comprising a trained researcher and a state truck inspector.

As a consequence of the extensive data collection, the raw database has over a thousand data fields and is organized into 58 different files. Extensive data processing was undertaken which involved cleaning, consistency checks, and data reduction. It is useful to point out that there were several variables (particularly those describing driver and vehicle characteristics) of potential interest that had missing values for a significant fraction of the cases. Further, it was also found that these values were "systematically missing" for many of the variables. For example, the value of certain variables could be more likely to be missing for crashes with greater severity. Alternatively, the value of other variables could be more likely to be missing for crashes with lesser severity. Consequently, simply removing the cases with missing values would skew the sample in addition to reducing the sample size. Simplified imputation methods such as replacing the missing values with sample averages would not be appropriate as well. Therefore, we chose to retain all cases, but indicator variables were created to explicitly identify cases with missing values for each of the variables of interest. These indicator variables were also included in the model specifications and some of these turned out to be statistically significant (discussed further in Section 5).

The final "reduced" dataset assembled for this analysis was organized into the three major files: crash-level data (including highest level of injury severity in the crash, crash type, and environment variables), vehicle-level data (characteristics of the trucks and cars such as age, body-type, cargo, occupancy levels, and deficiencies), and driver-level data (characteristics of the truck and car drivers including demographics, fatigue, health, and behavior). The impacts of all these different variables were examined during the statistical-modeling procedure. There are 953 crashes in this estimation sample.

As already discussed, the data were collected from both police accident reports (including case narratives) and additional sources such as site investigation, interviews, and review of medical records. Correspondingly, there are two measures of injury severity available for each crash: (1) determined from the police accident reports (referred to as PAR) and (2) determined by the LTCCS researchers (referred to as RES for "researcher determined").

A cross-tabulation of the highest injury-severity level of the 953 crashes determined from each of PAR and RES is presented in Table 2. Note the four levels of injury severity for PAR in contrast to only three levels for the RES. This is because the crashes in the LTCCS were sampled such that there is at least one injured person (level B) on the RES-scale.

The cross-tabulation indicates that fatal crashes are recorded almost identically by both measures (approximately 22% each). However, discrepancies are observed in the case of non-fatal crashes. About 12% of all crashes (119 crashes) that were classified

Table 1
Definitions of human-factor variables.

Variables	Definitions (page numbers from the LTCCS codebook in parenthesis)
Aggression	Documents whether or not the driver exhibited aggressive behavior in terms of exceeding the speed limit by a minimum of 5 mph and the vehicle's speed has a bearing on subsequent crash events, or traveling in close proximity to a vehicle forward of his/her position, or weaving in and out of traffic to pass slower vehicles, or violating a displayed red signal phase or stop sign, or engaging in these activities in a repeating fashion, of repeatedly honking the vehicle's horn at surrounding traffic to gain a time/space advantage, or repeatedly flashing the vehicle's lights in an attempt to have traffic forward move either to the right or left, or using his/her vehicle to physically obstruct the path of another vehicle, or other aggressive driving behavior. (pp. 90–93)
Distraction	Documents whether or not the driver is distracted by internal factors (e.g., other occupants in the vehicle, or as a result of either dialing or hanging up a phone during the pre-crash phase, or attempting to adjust the sound system controls of heat vent or air-conditioning controls during the pre-crash phase, or trying to retrieve an object from either the floor or seat while driving, etc.) or exterior factors (e.g., driver removes his/her focus from the driving task to look at a previous crash, or approaching traffic, or task to search for a street address, or a person exterior to this vehicle or a building). (pp. 107–111)
Emotional behavior	Established if the driver was upset, in a hurry (reflected in pre-crash driving behavior, e.g., speeding, sudden starts/stops, weaving in and out of traffic, etc.) prior to the crash or there were other emotional factors relevant to this driver's pre-crash behavior) (pp. 73,74, 269)
Fatigue	Assessed based on an evaluation of the driver's current and preceding sleep schedules, current and preceding work schedules, and a variety of other fatigue-related factors including recreation and non-work activities. (pp. 73, 268)
Illness	Establishes whether or not driver experienced a heart attack, or epileptic seizure, or blackout that can be traced to a medically diagnosed diabetic condition, or severe cold/flu which influence driving performance or any other illness or physical symptoms. (pp. 96–98)
Speeding	Documents whether the there are speed-related or gap distance-related factors coded for this driver travel speed of this driver; in addition, a driver is said to be speeding if his/her speed is at least 10 km/h. more than the roadway's posted speed limit (pp. 84–86, 271)
Work pressure	Establishes whether or not the driver was under pressure from his/her employer as a result of learning a new position in his/her primary work place, or being under time-related pressures associated with production/shipping deadlines, or to accept loads with little or no advance notice, or driver was experiencing any pressure on the job as it relates to his/her work schedule, or with regard to additional production or sale requirements, or driver had recently been forced to accept a demotion and/or pay decrease, or the driver experienced self-induced work pressure (e.g., a truck driver continuing to driver even though he knows he is over his allowed driving hours), or other work-related pressure. (pp. 75–77)
Vision Problem	Establishes whether or not the driver wears corrective lenses to compensate for a near-sighted condition, or far-sighted condition, or diagnosed as having astigmatism, or other vision problems. (pp. 98-100)

Note: Each of the above variables was ultimately recoded into three levels (1) driver exhibited the behavior, (2) driver did not exhibit the behavior, (3) it is known whether or not the driver exhibited such behavior.

Table 2
Cross-tabulation of police-determined and researcher–determined injury severity.

		RES						
		Possible injury	Non-incapacitating injury	Incapacitating injury	Killed	Total	Percentage	Weighted percentage
PAR	Possible injury	0	119	15	0	134	14.06	14.39
	Non-incapacitating Injury	0	239	61	2	302	31.69	31.32
	Incapacitating injury	0	103	192	10	305	32.00	45.90
	Killed	0	1	5	206	212	22.25	8.40
	Total	0	462	273	218			
	Percentage		48.48	28.65	22.88			
	Weighted percentage		55.00	36.52	8.48			

as level C by PAR were classified as level B by RES indicating under-estimation of the injury severity by the police reports (consistent with the findings of others such as Tsui et al., 2009). At the same time, 103 crashes classified as level A by the PAR were classified as level B by the RES indicating an over-estimation by the police reports. Given these discrepancies, it is useful to compare models estimated using each of these severity measures. Such an effort is undertaken in this study.

It is also important to note that the sampling of the crashes was not purely random. Therefore, weights have been calculated to scale the sample to be nationally representative (see the RATWEIGHT variable described in user's manual by USDOT, 2006a, b). Table 2 also presents the weighted percentages of the injury-severity levels according to the two measures (see the last row and last column). The results indicate that crashes with higher levels of severity (particularly fatal crashes) have been over-sampled.

4. Methodology

Injury severity is generally recorded in an ordinal scale. Most commonly, a five-level scale in an increasing order of injury severity is used: property-damage only (O), possible injury (C), non-incapacitating injury (B), Incapacitating Injury (A), and killed/fatal (K). An ordered-response discrete-choice model (either probit or logit) is appropriate for the analysis of such data. In fact, this has been a popular approach for modeling injury severity in general (For example, see Kockelman and Kweon, 2002; O'Donnell and Connor, 1996). Un-ordered choice models, such as multinomial logit/probit and nested logit while being more flexible (because of increased number of parameters) do not recognize the natural ordering (increasing severity) of the alternatives (injury levels). Advanced/flexible model structures that also recognize ordering of alternatives such as the partial proportional odds model (see Wang and Abdel-Aty, 2008) and the ordered generalized extreme value model (Small, 1987) are available. However, these models are more difficult to interpret than the simpler ordered-probit model. Further, the ordered-probit is also more attractive as it is parsimonious in the number of parameters, an issue of practical relevance for this effort as the sample size is small relative to the number of explanatory factors available. In consideration of these issues, the ordered-probit was chosen for this analysis. A comparative analysis of alternate model structures, subject to the availability of adequate data, is identified as an area of future research.

For any crash n, the observed, ordinal, injury-severity level (I_n) is related to an unobserved (latent), continuous injury propensity (U_n) in the following way:

$$I_n = \begin{cases} O & \text{if} \quad -\infty \le U_n \le \psi_1 \\ C & \text{if} \quad \psi_1 \le U_n \le \psi_2 \\ B & \text{if} \quad \psi_2 \le U_n \le \psi_3 \\ A & \text{if} \quad \psi_3 \le U_n \le \psi_4 \\ K & \text{if} \quad \psi_4 \le U_n \le \infty \end{cases}$$

In the above expression, O, C, B, A, and K are the levels of injury severity and ψ_1, ψ_2, ψ_3, and ψ_4 are thresholds on the injury propensity (to be estimated from the data). The thresholds demarcate the different ordinal levels of the injury severity. The continuous injury propensity (U_n) is then related to a vector of explanatory variables (X_n) corresponding to the crash via the following linear-in-parameters specification:

$$U_n = X_n\beta + \varepsilon_n \quad \forall n$$

In the above equation, β is a vector of model parameters or coefficients on the explanatory variables. These capture the marginal effect of the corresponding factor on the injury severity of the crash. In our specification, a positive value of a coefficient indicates that the corresponding explanatory factor is associated with more severe crashes.

Finally, ε_n is the error term capturing the effect of unobserved factors (factors other than those included in the vector X_n) on the injury-severity propensity. For model identification purposes, this error term is taken to independently and identically standard normal distributed across the crashes. This leads to the ordered-probit model.

The model parameters, i.e., the coefficients on the explanatory variables and the threshold parameters were estimated using the weighted maximum-likelihood approach. The weights (ratio of the sample to the population shares) are included to correct for the over-sampling of higher severity crashes in sample. The sandwich estimator is used to calculate robust standard errors for the parameter estimates. The likelihood function and the gradients were coded in the GAUSS programming language.

5. Results

As already discussed, the data contain two measures of injury severity for each crash (PAR and RES). However, in the LTCCS, every crash sampled had at least one injured person (level B) on the RES-scale. Therefore, there are only three levels for injury severity according to the RES-scale. To facilitate comparisons, we aggregated the four-category PAR injury scale to three categories (possible injury and non-incapacitating injury were combined into a single group). Therefore, only three levels (B, A, and K) are relevant for this study.

Ordered-probit models were developed using each of the PAR and RES injury-severity measures. The rho-square values of these models are, respectively, 0.1780 and 0.1827 indicating that the model based on RES injury severity has the better fit. These models were estimated using the weighted maximum-likelihood method as indicated previously. Unweighted models were also estimated, however, the weighted models outperformed the unweighted models in terms of the goodness-of-fit measures.

The estimation results are presented in Tables 3–5. Each table presents both models (PAR and RES) and focuses on one set of explanatory factors: (1) crash-level variables (Table 3), (2) truck-

Table 3
Empirical model results: effects of crash-level variables.

Variables	Sample share (%)	PAR		RES	
		Parameters	*t*-stat	Parameters	*t*-stat
Crash type					
Truck only, rollover	12.9	−0.0862	−0.479	**−0.3913**	**−1.973**
Truck only, object	7.5	**0.4332**	**1.816**	0.3626	1.533
Truck only, other (multi-truck)	10.5	–	–	–	–
Truck–car, angle	10.1	–	–	–	–
Truck–car, sideswipe	21.0	**0.6176**	**3.694**	0.0954	0.533
Truck–car, rear end	20.7	−0.0159	−0.099	**−0.3892**	**−2.206**
Truck–car, head-on	4.0	**1.7030**	**6.270**	**1.1661**	**3.835**
Truck–car, multi-impact	3.3	–	–	–	–
Truck–car, other	10.2	**0.4254**	**2.187**	−0.1445	−0.662
Fire					
Fire exist	5.0	**0.5970**	**2.842**	**0.4402**	**1.771**
Crash location					
Interstate, no junction	32.4	**−0.4674**	**−3.095**	**−0.4399**	**−2.673**
Interstate, ramp (interchange)	8.2	**−0.8268**	**−3.302**	**−1.1510**	**−5.462**
Interstate, crossover, rail grade crossing (interchange)	4.7	**−1.0190**	**−4.006**	**−0.9622**	**−3.991**
Interstate, other	1.3	**−1.9760**	**−5.893**	**−1.8714**	**−5.231**
Highway, no junction	16.5	–	–	–	–
Highway, intersection (not interchange)	6.8	**0.3623**	**2.300**	**−0.4074**	**−2.002**
Highway, other	4.5	**−0.5213**	**−1.652**	**−0.5567**	**−2.299**
Other	25.6	–	–	–	–
Roadway design characteristics					
Number of lanes (mean = 3.5, standard deviation = 1.27)		**−0.1279**	**−2.889**	−0.0404	−0.868
Speed limit[a] (mean = 5.6, standard deviation = 1.85)		**0.1606**	**4.957**	**0.0814**	**2.370**
Road-surface conditions					
Wet	14.5	−0.2330	−1.425	**−0.3883**	**−2.322**
Snow and ice	2.3	–	–	–	–
Dry	83.2	–	–	–	–
Time of day and lighting					
Daylight, peak (7:00 a.m.–10:30 a.m.)	23.1	–	–	–	–
Daylight, off peak (10:30 a.m.–7:30 p.m.)	52.0	–	–	–	–
Dark, not lighted (7:30 p.m.–6:00 a.m.)	11.6	–	–	–	–
Dark, lighted (7:30 p.m.–6:00 a.m.)	13.2	**0.4183**	**2.365**	**0.6114**	**3.809**
Day of the week					
Weekday	90.0	−0.2767	−1.332	**−0.4710**	**−2.186**

Bold indicates significance at $p \leq 0.05$.
[a] The coefficients capture the effect of an increase in speed limit by 10 km/h.

level variables (Table 4), and (3) car-level variables (Table 5). The effects that are significant at 95% or higher confidence (t value >1.9) are shown in *bold* font. The parameters that were significant in one model were also retained in the other. Parameters insignificant in both models were removed from the specification. The threshold parameters of are 0.6224 and 2.6196 for the PAR model and 0.3950 and 2.0918 for the RES model.

In addition to the parameter estimates and their t statistics, the first major column of Tables 3–5 present the (unweighted) sample shares of the variables. For example, the "12.9%" beside "Truck-only, roll over" in Table 3 indicates that 12.9% of the crashes in the sample are truck-rollover crashes. In the case of continuous variables, the sample mean and standard deviation are presented. For example, in Table 3, the average number of lanes per segment in the sample is 3.5, with a standard deviation of 1.27.

The models are discussed in detail in the rest of this section. As already indicated in the section on methodology, a positive coefficient on a variable indicates that the corresponding factor is associated with more severe injuries. Further, unless explicitly stated, the interpretations presented can be taken to apply to both models.

5.1. Crash-level variables

Crash-level explanatory variables found to impact the injury severity are presented in Table 3.

There are nine types of crashes in the sample based on (1) whether or not cars are involved in the crash, and (2) the direction of the impact (angle, sideswipe, rear end, head-on, etc.). In the case of the truck-only crashes (the first three categories under crash type; about 31% of the crashes in the sample fall in this category), rollover crashes are less severe while crashes of trucks with fixed objects are more severe. In crashes involving trucks and cars (the last six categories under crash type), the severity of the injury depends upon the nature of impact. Head-on crashes are found to be most severe (most severe of all nine crash types). This result is intuitively reasonable considering the large difference in the size/weight of the two vehicles and the relative velocity of impact considering the opposing directions of movement. Rear-end crashes are shown to be least severe (the coefficients for this type are zero or negative).

Crashes that result in fire also lead to more severe injuries. This effect is estimated to be statistically significant only in the PAR model.

The severity of large-truck crashes also varies by crash location. In this analysis, crash location can be one of eight categories depending on the facility type (interstate versus other highway) and the location (segment, intersection, interchange, etc.). Crashes in interstate "other" segments (non interchange/junction/ramp locations) are least severe (most negative coefficient). Crashes in non-interstate highways are more severe than crashes in interstate highways (coefficients corresponding to interstates are more negative), and in particular, the highway intersections are associated

Table 4
Empirical model results: effects of truck-level variables.

Variables	Sample share (%)	PAR		RES	
		Parameters	*t*-stat	Parameters	*t*-stat
Truck driver demographics					
Age					
<35 years old	32.8	0.0323	0.253	−**0.2438**	−**1.786**
<45 years old	65.6	−**0.4334**	−**3.347**	−**0.2435**	−**1.785**
Ethnicity					
Caucasian	60.1	–	–	–	–
Black	13.7	**0.4027**	**2.956**	0.2069	1.417
Hispanic	14.7	–	–	–	–
Other/missing	11.5	–	–	–	–
Height (mean = 176.6, standard deviation = 7.85)		**0.0046**	**2.420**	**0.0043**	**2.668**
Vision					
Wear eyeglasses	24.2	−**0.3448**	−**2.708**	−**0.5030**	−**3.685**
Wear contact lenses	2.2	–	–	–	–
No eyewear	58.6	–	–	–	–
Eyewear unknown	15.0	–	–	–	–
Work pressure					
Has work pressure	9.7	–	–	–	–
No work pressure	75.2	–	–	–	–
Work pressure unknown	15.1	0.1228	0.466	−**0.5586**	−**2.288**
Illness					
Illness	2.6	–	–	–	–
No illness	85.0	–	–	–	–
Illness unknown	12.4	0.5105	1.971	1.0337	4.604
Distraction					
Has distraction behavior	11.9	0.0759	0.490	**0.3042**	**1.956**
No distraction behavior	75.8	–	–	–	–
Distraction behavior unknown	12.3	**0.5303**	**2.024**	**0.8522**	**3.924**
Familiarity with vehicle					
Driven this vehicle <10 times in the past 6 months	9.4	**0.5276**	**3.397**	0.2590	1.479
Driven this vehicle >10 times in the past 6 months	71.0	–	–	–	–
Truck driver knew vehicle unknown	19.6	–	–	–	–
Truck type					
Single-unit straight truck	27.9	−**0.5098**	−**3.552**	−**0.2847**	−**1.792**
Truck–tractor pulling one trailer <10,000 kg	24.8	−**0.3387**	−**2.604**	−**0.3176**	−**2.096**
Truck–tractor pulling one trailer 10,000–20,000 kg	15.8	−**0.4902**	−**2.787**	−**0.3426**	−**1.958**
Truck–tractor pulling one trailer >20,000 kg	12.6	–	–	–	–
Truck–tractor pulling one trailer weight missing	9.3	–	–	–	–
Other	9.6	–	–	–	–
Truck deficient					
Brake deficient	29.2	–	–	–	–
Truck deficiency unknown	2.4	−0.5909	−1.575	−**1.1430**	−**2.832**

Bold indicates significance at $p \leq 0.05$.

with high severity crashes (the coefficient has the highest magnitude among all locations in the PAR model and second highest magnitude in the RES model).

Roadway design characteristics that found to impact the injury severity are number of lanes and the speed limit. Crashes on roadways with more lanes are found to be less severe (RES model). This is probably because of better separation of the trucks from other types of vehicles in multi-lane highways (most cases in the sample have two, three, or four lanes). Crashes in higher speed facilities are estimated to be more severe. We also examined the effect of horizontal (straight versus curved) and vertical (flat versus uphill/downhill) alignment of the roadway. However, these are found to be statistically insignificant.

On examining the effect of road-surface conditions, crashes happening on wet roadways are found to be less severe. This is probably because drivers are inherently more cautious during such rainy conditions and such behavior has been suggested by past research as well. We also explored the effect of weather in lieu of road-surface condition and these were found to be insignificant.

On examining the temporal characteristics of the crash, we find that crashes happening during “dark but lighted” conditions (7:30 p.m. to 5:30 a.m.) lead to most severe crashes (relative to daylight and dark periods). Weekday crashes are estimated to be less severe than weekend crashes.

5.2. Truck-level variables

Truck-level explanatory variables found to impact the injury severity are presented in Table 4. A substantial fraction of the crashes (abut 86%) involved only one truck, and hence the characteristics of this vehicle and its driver were included in the model. For the remaining cases, we used data from one of the large-trucks involved in the crash. This “representative” truck was chosen based on the extent to which the truck contributed causally to the crash (determined subjectively from various data elements), the injury-severity level of the driver (the truck that had more severely injured drivers were chosen), and the overall completeness of the data.

Within the class of truck-level variables, the characteristics of the truck driver are first discussed followed by the characteristics of the vehicle.

Age, ethnicity, and height are truck-driver characteristics found to be statistically significant. Older truck drivers (greater than 45 years of age) are found to be involved in more severe crashes, perhaps the longer reaction times of older drivers is an influential aspect. In the PAR model, African-American drivers are found to be involved in more severe crashes compared to drivers of other ethnicities, although this effect is not strongly significant in the RES model. Taller drivers are estimated to be involved in more severe crashes. An explanation for this interesting finding is not quite apparent. It is also useful to mention that almost all the truck drivers in the sample are men and, therefore, it is not appropriate to examine gender differences.

The model indicates that those who wear eyeglasses are involved in less severe crashes. Given that about 65% of the truck drivers in the sample (and hence involved in the crash) do have some sort of a vision problem, this effect appears reasonable. That is, those drivers for whom the vision deficiency is corrected with the eyeglass are likely to be involved in less severe crashes.

Driver fatigue is an important factor of interest in the context of truck safety. In this study, variables related to fatigue (12.9% fatigued and 15.4% missing data) and work pressure turned out to be generally statistically insignificant predictors of injury severity, after controlling for a wide-array of other factors. It is quite possible that the effects of fatigue are partially captured by certain other statistically significant variables such as time of day and crash type. Given the small sample size (relative to the number of explanatory factors of interest), we could not explore second-order effects by creating interaction variables of fatigue and crash characteristics. This is identified as an important future avenue for research where larger samples are available.

Related to the fatigue effect is the impact of illness. Interestingly, the crashes in which illness variable could not be ascertained were more severe compared to crashes for which illness variable could be ascertained. In the cases for which illness could be ascertained, there appears to be no statistical difference in injury severity between the crashes in which illness effects are present and those in which these effects are not present.

Distracted drivers are found to be involved in more severe crashes (based on the RES model). At the same time, the drivers for whom the distraction level could not be ascertained were involved in ever more severe crashes. Perhaps it was the extreme severity of the crash that prevented the ascertainment of the distraction level of the driver. The effect of aggressive behaviors (6.3% of drivers exhibited aggressive behavior) were also examined, but found to be statistically insignificant.

Table 5
Empirical model results: effects of car-level variables.

Variables	Sample share (%)	PAR		RES	
		Parameters	*t*-stat	Parameters	*t*-stat
Seat belt use					
Used both shoulder and lap belt	56.7	–	–	–	–
User other type of belt	24.9	–	–	–	–
Seatbelt not used	13.7	–	–	–	–
Seatbelt unknown	4.8	**0.8590**	**2.430**	**0.7586**	**2.255**
Vision					
Has vision problem	45.7	0.1990	1.642	**0.5308**	**3.893**
No vision problem	18.7	–	–	–	–
Vision problem unknown	35.7	–	–	–	–
Alcohol					
Alcohol involvement for this vehicle/driver	9.8	0.3582	1.436	**0.5162**	**2.095**
Emotional behavior					
Emotional behavior present	10.0	**0.8200**	**3.847**	0.2801	1.108
No Emotional behavior	48.7	–	–	–	–
Emotional behavior unknown	41.4	**0.9387**	**3.006**	**0.5986**	**1.921**
Fatigue					
Fatigued	10.0	**−0.4001**	**−1.995**	−0.0080	−0.035
Not fatigue	46.6	–	–	–	–
Fatigue unknown	43.5	**0.7989**	**3.125**	**0.7634**	**2.577**
Familiarity with vehicle					
Driven this vehicle <10 times in the past 6 months	3.8	–	–	–	–
Driven this vehicle >10 times in the past 6 months	53.8	–	–	–	–
Unknown	42.5	**−0.5492**	**−2.563**	**−0.5790**	**−2.680**
Familiarity with roadway					
First time driving on this roadway	2.6	–	–	–	–
Drives on this roadway rarely or once per month	13.2	–	–	–	–
Drives on this roadway several times per month to weekly	14.5	**−0.4004**	**−2.088**	−0.0318	−0.166
Drives on this roadway daily	33.0	–	–	–	–
Unknown	37.0	**−0.5465**	**−2.090**	0.3021	1.139
Speed					
Over speeding	5.9	–	–	–	–
Not over speeding	29.8	–	–	–	–
Unknown	64.4	**−0.3041**	**−2.394**	−0.1626	−1.183
Car type					
Car	59.0	–	–	–	–
Compact and large utility vehicle	12.1	–	–	–	–
Van	9.3	**0.4776**	**2.423**	**0.6378**	**2.979**
Compact and large pickup truck	16.1	–	–	–	–
Other	3.5	–	–	–	–

Bold indicates significance at $p \leq 0.05$.

The PAR model indicates the familiarity of the truck driver with the vehicle affects the severity of the crash. Specifically, drivers who have driven a truck fewer times are likely to be involved in more severe crashes. The familiarity of the driver with the roadway was, however, found to be statistically insignificant.

The use of seat belts by truck drivers is found to be insignificant in predicting the injury severity of large-truck crashes, after controlling for other factors. It is possible that the use of seat belts is indicative of other safe driving patterns. Therefore, it would be useful to examine interaction effects of seat belt use and driver behavior variables subject to additional data availability. It is also useful to note that over 80% of the truck drivers in our analysis sample wore seat belts. However, in a recent study, Kim and Yamashita (2007), found that truck drivers were least likely (60%), among all commercial motor vehicle drivers, to report wearing seat belts always.

The discussion thus far focused on the characteristics of the truck driver. The following discussions focus on the characteristics of the truck.

Truck type is found to affect the injury severity of the crashes. Crashes involving trucks hauling a trailer with heavy cargo (>20,000 kg) result in more severe injuries relative to trucks hauling less heavy cargo and single-unit straight trucks. Single-unit straight trucks are also estimated to be associated with least severe injuries. Quite interestingly, the tractor-trailer-type trucks whose cargo-weight could not be ascertained and "other" types of trucks are also associated with higher severity crashes. Other (disaggregate) classification schemes for truck type were also explored, but the efforts did not yield statistically superior models.

Other truck characteristics examined include vehicle deficiencies. Almost one-third of the trucks had some brake deficiency, however, this was found not to affect the severity of the crash. In the few cases in which truck deficiency was unknown, the crashes were found to be statistically less severe. The impact of the age of the truck was also examined, but was found to be statistically insignificant.

5.3. *Car-level variables*

Car-level explanatory variables found to impact the injury severity are presented in Table 5. A substantial fraction (about 74%) of the crashes in which cars were involved, have only one car. For the remaining cases, with two or more cars involved in the crash, we used data from one of the cars involved in the crash. This "representative" car was chosen based on the extent to which the vehicle contributed causally to the crash (subjectively determined from various data elements), the injury-severity level of the driver (the car that had more severely injured drivers were chosen), and the overall completeness of the data. As already described in the case of Table 3, for binary or "dummy" variables, the sample share of that variable is presented in the second column.

Within the class of car-level variables, the characteristics of the car driver are first discussed followed by the characteristics of the vehicle.

Quite interestingly none of the socio-economic characteristics of the car driver (age, gender, ethnicity, and height) turned out to be significant predictors of crash severity (average age is 39; 61% of the car drivers in the sample were men; 54.7% white, 12.0% black, 9.4% Hispanic, 4.9% other and 9.14% missing race; mean height 171.7 cm with a SD of 9.40).

The use of seat belts is found to be statistically insignificant; however, those persons whose seat belt usage could not be determined were involved in more severe crashes. The effects of air-bags were also examined and these effects were found to be statistically insignificant as well (air-bags were available in 44% of the cars, not available in 26% and was not ascertained in the remaining cases).

The RES model indicates that car drivers with vision problems (45% of all car drivers in the sample) are involved in more severe crashes. The same model also indicates that car drivers under the influence of alcohol are involved in more severe crashes. Both these effects were insignificant in the PAR model.

The PAR model indicates that car drivers with emotional factors (such as being upset or clinically depressed) are involved in more severe crashes. At the same time, the drivers for whom this could not be ascertained were involved in even more severe crashes (according to both models). Both models also indicate that the cases in which the fatigue of the car driver could not be ascertained were also associated with more severe crashes. In the case of fatigue and emotional state, the data was missing for approximately 30% of the car drivers. As already discussed, the "dummy" variable identifiers for these cases have strong positive coefficients indicating more severe injuries. Thus, it would be of value for future data-collection efforts to try and obtain more complete data on these variables so that their effect on crash severity may be comprehensively ascertained. Other behavioral factors explored include aggression and distractions and both these were insignificant.

The familiarity of the car driver with the vehicle was found to not have a statistically significant effect. The familiarity with the roadway was, however, found to impact crash severity. The reader will note that opposite effects were observed in the case of truck drivers (familiarity with the vehicle was significant and familiarity with the roadway is not). According to the PAR model, crashes on unfamiliar roadways (drives on this road rarely or for the first time) are more severe than crashes on roadways more familiar to the drivers (drives on this road several times a month) – a result that is intuitively reasonable. At the same time, the PAR model also indicates that crashes on roadways which are used by the drivers on a daily basis are also more severe. These roads are likely to be commute corridors and therefore it is possible that familiarity effects are offset by time pressures imposed by the work-related travel.

The cases in which familiarity with either the vehicle or the roadway could not be ascertained have lower injury severity. The same is also true for cases in which the speed of the car driver was not ascertained. Perhaps, in light of the low severity of these crashes, it was deemed unnecessary to collect these data elements.

Among the characteristics of the vehicle, the vehicle type was the statistically significant factor. Vans are estimated to be involved in more severe crashes compared to other types of cars (such as sedans, pick-up trucks, and SUVs). In general, the expectation is that larger passenger vehicles will be able to better sustain impacts with a truck and, hence, result in lower injury severity. The estimated result is not completely consistent with this effect. At the same time, vans are likely to have more passengers, especially children, and this increased vehicle occupancy makes it more likely for a more severe injury to occur. We also included a variable to capture the effect of the presence of other occupants in the car but this variable was not found to be statistically significant. The age of the vehicle was another factor found to be statistically insignificant.

6. Summary and conclusions

Given the importance of trucking to the economic well being of a country and the safety concerns posed by the trucks (because of their large size, weight, maneuverability, etc.), a study of large-truck crashes is critical. This paper contributes towards that end by undertaking an extensive analysis of the empirical factors affecting injury severity of large-truck crashes. Data from the Large Truck Crash Causation Study were used. The injury severity was modeled using both police-reported and researcher-determined scales. Several similarities and some differences were observed across the two models. This underscores the need for improved accuracy in the assessment of injury severity of crashes.

The models estimated capture the marginal effects of a variety of explanatory factors simultaneously. In general, the results are intuitively reasonable and have substantive implications. For example, among truck–car crashes, head-on collisions and collisions at intersections are estimated to be most serious while crashes on multi-lane highways are less severe. Consequently, effective separation of opposing traffic movements is important for improving safety in truck corridors. Further, intersections are particularly important. Weekend days and night times are also particularly critical. Increased familiarity of truck drivers with their vehicle decreases the severity of the crash. Therefore, better education/training of the drivers about the vehicle and on safe driving practices can help improve safety. Car drivers unfamiliar with the roadways are associated with more severe crashes. This suggests that particular attention must be paid to truck-corridors near major tourist destinations as these areas have a higher share of car drivers in unfamiliar environments. At the same time, attention must also be paid to truck corridors along commute routes.

The empirical results also indicate the impacts of several driver behavior variables on the severity of the crashes, after controlling for a variety of other factors. For example, driver distraction (truck drivers), alcohol use (car drivers), and emotional factors (car drivers) are found to be associated with higher severity crashes. Other factors such as truck-driver fatigue, aggression, and seat belt usage turned out to be statistically insignificant.

A particularly interesting finding is the strong statistical significance of the dummy variables that indicate missing data. In some cases (such as fatigue, distraction, and illness), the coefficients are positive. This indicates that the injury is more severe in the crashes for which these values were missing/unknown. It is possible that some of these data could not be collected because of the severe nature of the crash. In other cases (such as car speed and driver familiarity), the coefficients are negative. Perhaps, it was deemed unnecessary to collect these data elements because of the low severity of these crashes. Overall, it is reasonable to expect that the nature of the crash could affect the completeness of the data. Future efforts should seek to collect all crash-related data comprehensively so that the true marginal effects of each factor on the crash severity can be accurately determined. Methodologies for addressing this challenging issue are an area of future research.

Overall, this study is an important effort in trying to understand the complex relationship between injury severity and a vast number of inter-dependent explanatory factors. However, the effort was limited by the relatively small sample (the sample size is small relative to the number of explanatory variables). Therefore, there are clearly areas for further empirical enhancements. For example, with the availability of a larger dataset, one could explore second-order effects by creating interaction variables of driver behavior and crash characteristics. In addition, the focus of this paper was on the overall (highest) level of severity of the crash. It is known that a vast majority of the fatalities that resulted from crashes involving large trucks are occupants of other vehicles. Therefore, future work should also seek to examine the injury severity to each occupant involved in the crash instead of just the overall crash severity. Finally, the use of flexible model structures is also identified as a methodological extension to this work.

References

Blower, D., Campbell, K.L., Green, P.E., 1993. Accident rates for heavy truck-tractors in Michigan. Accident Analysis and Prevention 25 (3), 307–321.

Braver, E.R., Preusser, D.F., Williams, A.F., Weinstein, H.B., 1996. Major types of fatal crashes between large trucks and cars. In: Proceedings of the 40th Annual Conference of the Association for the Advancement of Automotive Medicine. Association for the Advancement of Automotive Medicine, Des Plaines, IL, pp. 153–169.

Campbell, K.L., 1991. Fatal accident involvement rates by driver age for large trucks. Accident Analysis and Prevention 23 (4), 287–295.

Chang, L.Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accidents. Accident Analysis and Prevention 31, 579–592.

Duncan, C.S., Khattak, A.J., Council, F.M., 1998. Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transportation Research Record 1635, 63–71.

Gander, P.H., Marshall, N.S., James, I., Quesne, L.L., 2006. Investigating driver fatigue in truck crashes: trial of a systematic methodology. Transportation Research: Part F 9, 65–76.

Golob, T.F., Recker, W.W., Leonard, J.D., 1987. An analysis of the severity and incident duration of truck-involved freeway accidents. Accident Analysis and Prevention 19 (5), 375–395.

Hallmark, S.L., Hsu, Y.Y., Maze, T., McDonald, T., Fitzsimmons, E., 2009. Investigating factors contributing to large truck lane departure crashes using the federal motor carrier safety administration's Large Truck Crash Causation Study (LTCCS) database. Final report prepared for the U.S. Department of Transportation Vople National Transportation Systems Center http://www.intrans.iastate.edu/reports/lg_truck_lane_departure.pdf.

Hedlund, J., Blower, D., 2006. Large Truck Crash Causation Study (LTCCS) Analysis Series: Using LTCCS Data for Analyses of Crash Risk, FMCSA-RI-05-037. http://www.fmcsa.dot.gov/facts-research/research-technology/analysis/ltccs.htm.

Khattak, A.J., Schneider, R.J., Targa, F., 2003. Risk factors in large truck rollovers and injury severity: analysis of single-vehicle collisions. In: Presented at 2003 Annual Transportation Research Board Meeting, on TRB CD-ROM.

Kim, K., Yamashita, E.Y., 2007. Attitudes of commercial motor vehicle drivers towards safety belts. Accident Analysis and Prevention 39 (6), 1097–1106.

Kockelman, K.M., Kweon, Y.J., 2002. Driver injury severity: an application of ordered probit models. Accident Analysis and Prevention 34 (3), 313–321.

Lyman, S., Braver, E., 2003. Occupant deaths in large truck crashes in the United States: 25 years of experience. Accident Analysis and Prevention 35 (5), 731–739.

NHTSA, 2006. Traffic Safety Facts 2005: Large Trucks, DOT HS 810 619. http://www-nrd.nhtsa.dot.gov/Pubs/810619.PDF.

NHTSA, 2008. Traffic Safety Facts 2007: Large Trucks, DOT HS 810 805. http://www-nrd.nhtsa.dot.gov/Pubs/810989.PDF.

O'Donnell, C.J., Connor, D.H., 1996. Prediction the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis and Prevention 28 (6), 739–753.

Small, K.A., 1987. A discrete choice model for ordered alternatives. Econometrica 55 (2), 409–424.

Tsui, K.L., So, F.L., Sze, N.N., Wong, S.C., Leung, T.F., 2009. Misclassification of injury severity among road casualties in police reports. Accident Analysis and Prevention 41 (1), 84–89.

USDOT/FMCSA, 2006. Report to the Congress on the Large Truck Crash Causation Study. http://www.fmcsa.dot.gov/facts-research/research-technology/report/ltccs-2006.htm.

USDOT, 2006. Large Truck Crash Causation Study Analytical User's Manual. http://ai.fmcsa.dot.gov/ltccs/data/documents/LTCCS_Manual_Public.pdf, pp. 510–512.

USDOT, 2006. Large Truck Crash Causation Study Codebook. http://ai.fmcsa.dot.gov/ltccs/data/documents/LTCCS_Codebook.pdf.

USDOT/BTS, 2008. 2007 Commodity Flow Survey. http://www.bts.gov/publications/commodity_flow_survey/preliminary_tables_december_2008/index.html.

Wang, X., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models. Accident Analysis and Prevention 40, 1674–1682.