Logistic Regression Models of the Safety of Large Trucks

Xiao Qin, Kai Wang, and Chase E. Cutler

Statistics show that crashes involving large trucks are generally more severe than those involving other vehicles because of the size, weight, and speed differential between trucks and other vehicles. Given the critical position of trucking in the process of economic recovery and growth, the improvement of truck safety and the mitigation of any negative impacts on non-truck vehicles are urgent issues. Statistical models have been used universally to identify the contributing factors to crash severities and to estimate injury probabilities. These methodologies, albeit addressing different issues, may provide mixed results and estimates with varying degrees of accuracy. The primary objective of this research was to investigate the effects of key determinants of the severity of crashes involving large trucks and to explore the relationship between the determinants. The secondary objective was to provide insight on statistical applications by evaluating three logistic regression models: multinomial logistic, partial proportional odds (PPO), and mixed logistic (ML) models. The model results showed that the majority of the coefficient estimates were consistent across the models studied. A few exceptions included young drivers and the use of safety constraints; these factors were not statistically significant in the ML model. The goodness of fit and model predictive power indicated that the PPO model produced results that more closely resembled the observations.

Freight transportation plays a vital role in economic development and recovery. Measured by value, 70% of the freight in the United States is transported by trucks and freight trucking plays an inarguably pivotal role in the growth and stimulation of the economy (*1*). In 2002, the U.S. transportation system transported \$36 billion of freight, approximately 53 million tons of freight each day (*1*). By 2008, this figure grew to 58.9 million tons. In the same year, Wisconsin's transportation system moved about \$1 billion in goods each day. By 2025, freight volumes in Wisconsin are projected to increase by 70% (*2*).

Among all the issues related to transport by truck, safety is the largest concern for truck industries as well as transportation agencies. In 2010, approximately 276,000 large trucks were involved in traffic accidents; in these accidents, 3,675 people were killed and 80,000 people were injured (*3*). In addition to bodily injuries and lost lives, the traffic disruptions and delays caused by crashes have a direct and immediate effect on the economy of Wisconsin, the Midwest, and the whole nation. The urgent need to reduce the number of truck crashes and, more importantly, the consequences of truck crashes, underscores the importance of the study.

Crash data are the most informative resource for finding the causes and contributing factors of injury. However, the cause of an injury is extremely complicated, often relating to a sequence of events (precrash, during crash, and postcrash) and a number of agents (driver, vehicle, and environment). Therefore, statistical methodologies have been widely popular in addressing the intriguing relationship between crash severity and other data elements. Most statistical methods are categorized as discrete choice models, which can be either fixed or random parameter models according to the parameter assumptions. Others can be classified as nonordinal models, such as multinomial logit (MNL) and multinomial probit (MNP) models, or as ordered probabilistic models, such as ordered probit (OP) and ordered logistic (OL) models, if an ordinal structure for the dependent variable is considered. Other model variations are available if restrictions such as irrelevant and independent alternatives (IIA), proportional odds, or heterogeneity are violated. Savolainen et al. provided an extensive and detailed review of the application of these methods (*4*). There is no agreement regarding which model works best for the crash severity data, although the consensus is to use the most advantageous method.

The primary objectives of this study were to identify the key contributing factors to the severity of crashes involving large trucks and to explore the relationship between the factors. The secondary objective was to offer additional insights through the comparison of three discrete outcome probabilistic models [i.e., MNL, partial proportional odds (PPO), and mixed logistic (ML)]. It was anticipated that the results would be useful for evaluating the different perspectives of the methodologies and helpful in selecting the most appropriate data-driven safety strategies.

LITERATURE REVIEW

Many researchers have accepted that there is an intrinsic ordinal nature in crash severities [i.e., injury severity can be ranked from high to low as fatal injury (K), incapacitating injury (A), nonincapacitating injury (B), possible injury (C), and property damage only (PDO or O)]. To model injury severity as the ordinal response, researchers most frequently have used discrete choice models, including OP and OL models (*5–9*). However, these traditional ordered probabilistic approaches assume regression parameters to be the same across all levels (i.e., proportional odds). This assumption can be overly restrictive because it is common for one or more of the regression parameters to differ across levels. Peterson and Harrell addressed this issue by relaxing the restrictions; they used the PPO model where some of the coefficients can be the same for all levels, while others can differ (*10*). Several recent studies have adopted this approach and the model

Department of Civil and Environmental Engineering, South Dakota State University, 148 Crothers Engineering Hall, Brookings, SD 57007. Corresponding author: X. Qin, xiao.qin@sdstate.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2392, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 1–10.
DOI: 10.3141/2392-01

results have shown that the PPO models consistently outperformed conventional ordered response models (*11–13*).

When modeling crash severity as an ordinal dependent variable, some restrictions can potentially affect the estimation results (*4*). The primary concern is the manner in which the explanatory variables affect the probabilities of the discrete outcome, i.e., the shift in the cutoff thresholds is constrained to move in the same direction. By contrast, nonordinal probabilistic models, such as MNL and MNP models, allow variables to have opposite effects regardless of the order of the injury severities. Yamamoto et al. argued that nonordinal models may offer unbiased estimates of the parameters, especially in the situation of crash underreporting (*14*). Ye and Lord examined the influence of crash underreporting on the estimation of crash severities and found that neither the OP nor the MNL model was immune to this issue (*15*). They suggested setting fatal crashes as the baseline in the MNL model and ranking the crash severity from K to O in descending order in the OP to minimize the bias (*15*).

Much of the crash severity literature has applied fixed parameter approaches, assuming the same effects of the explanatory variables across crash observations. This is valid under the desirable conditions where the data are complete (i.e., unobserved data heterogeneity is not apparent). Unfortunately, this is usually not the case for crash data because unobserved factors that may affect the consequences of crashes are highly likely to exist. Data heterogeneity suggests the parameters may vary across different observations and disregarding this feature may lead to bias and inefficient statistical inferences (*16*). The ML model overcomes this limitation by allowing the parameters to be random. Milton et al. suggested that volume-related variables, such as average daily traffic per lane, average daily truck traffic, truck percentage, and weather effects, can best be modeled as random parameters, while roadway characteristics can best be modeled as fixed parameters (*17*). Chen and Chen concurred that weather characteristics, such as snowy or slushy surface conditions, and a light traffic indictor appeared to be random coefficients (*18*). ML models offer new perspectives to the crash severity models.

Most of the literature that is pertinent to crash severities has included a vehicle type variable in which truck is listed as one of the values. Some studies have explicitly estimated the injury severities for crashes involving light trucks or large trucks (*7, 18–20*). Kockelman and Kweon found that both light-duty trucks and heavy-duty trucks seem to be better at protecting their drivers in all crash types (*7*). The results become more apparent in a two-vehicle crash involving a heavy-duty truck and in which more severe injury is sustained by the driver of the other vehicle (*7*). Chen and Chen identified the critical risk factors such as driver, vehicle, temporal, roadway, environmental, and accident characteristics for truck crashes involving single and multiple vehicles (*18*). Chang and Mannering studied the occupancy (the most severe injury sustained by the occupant and the number of occupants) and injury severity relationship; they used truck-involved crashes and identified some risk factors unique to large trucks (*19*). Zhu and Srinivasan used data from the Large Truck Crash Causation Study and found that driver behaviors, such as driver distraction (truck drivers), alcohol use (car drivers), and emotional factors (car drivers), were statistically significant for higher-severity crashes (*20*). Milton et al. argued that trucks can slow down the travel speed in the traffic stream, which may decrease crash injury severity, but that a truck's size and weight may increase the severity of a crash if the colliding partner is of lighter weight (*17*). In another study, Wang and Kockelman used the National Automotive Sampling System's Crashworthiness Data System to analyze the effects of vehicle weight and type; they found that these values can simultaneously affect the injury outcome, creating a complicated picture (*21*). The factors that affect the propensity of truck crash severities may not be identifiable if all vehicle types were considered. Because of the influence that trucks pose in traffic compared with all vehicles, studying only truck-related crashes would provide greater insight in determining the factors and their effects. In this study, three logistic models (MNL, PPO, and ML) were developed to analyze the large-truck crash severity data.

METHODOLOGIES

This section presents three logistic models and their respective assumptions and mathematical equations. The three models are the MNL, the PPO, and the ML.

MNL Model

The MNL model is a discrete choice model that considers a response variable with three or more levels without accounting for order between levels. This model is known for its IIA assumption, meaning adding or deleting an alternative will not change the ratio between the probabilities of any pair of existing alternatives. The general framework used to define the relationship between injury severities and contributing factors can be expressed by a linear function $\boldsymbol{U}$ (i.e., a utility function) that determines the preference or possible value of attaining the outcome i ($i = 1, 2, \ldots, I$) for observation n as

$$\boldsymbol{U}_{in} = \boldsymbol{\beta}_i \boldsymbol{X}_{in} + \boldsymbol{\varepsilon}_{in} \tag{1}$$

where

$\boldsymbol{X}_{in}$ = vector of independent variables for nth observation with ith outcome,
$\boldsymbol{\beta}_i$ = vector of corresponding unknown coefficients, and
ε_{in} = disturbance term that accounts for random noise.

If ε is assumed to be distributed logistically across observations, it composes a MNL formulation as in Equation 2:

$$P_n(i) = \frac{\exp[\boldsymbol{\beta}_i \boldsymbol{X}_{in}]}{\sum_{\forall I} \exp[\boldsymbol{\beta}_I \boldsymbol{X}_{In}]} \tag{2}$$

where $P_n(i)$ is the probability of nth observation falling into the ith outcome.

The estimated coefficients are usually presented as a log odds ratio between the probability of a given category and the reference one, resulting in $(I - 1)$ estimates for each independent variable if the response variable has I levels. The odds ratio is defined as the ratio between the probabilities of two specific categories and specifies the propensity of an individual falling into one category compared with the other. If level I is the reference, the model can be rewritten as

$$\mathbf{log}\left[\frac{\boldsymbol{P}_n(\boldsymbol{i})}{\boldsymbol{P}_n(\boldsymbol{I})}\right] = \boldsymbol{\alpha}_i \mathbf{X}_{in} \tag{3}$$

where $\boldsymbol{\alpha}_i$ is a vector of estimable parameters that represents the log odds ratio between the probabilities of two alternatives. With reorganization of Equation 1, the probability of each level can be expressed as

$$P_n(i) = \frac{\exp[\boldsymbol{\alpha}_i \boldsymbol{X}_{in}]}{1 + \sum_{i=1}^{I-1} \exp[\boldsymbol{\alpha}_i \boldsymbol{X}_{in}]} \tag{4}$$

PPO Model

A latent variable $\boldsymbol{U}$ introduced into the model as a linear function for each observation is specified as

$$U = \boldsymbol{\alpha X} + \varepsilon \tag{5}$$

where

- $\boldsymbol{X}$ = vector of independent variables determining the discrete ordering for each observation,
- $\boldsymbol{\alpha}$ = vector of estimable parameters, and
- ε = error term.

The observed response variable, y, for each observation is defined as (*22*)

$$\begin{aligned} y &= 1 && \text{if } U \le \mu_1 \\ y &= 2 && \text{if } \mu_1 < U \le \mu_2 \\ &\dots \\ y &= I && \text{if } U \ge \mu_{I-1} \end{aligned} \tag{6}$$

where the μs are estimable parameters (referred to as thresholds) that define y and I is the highest-integer ordered response. If the error term ε is assumed to be logistically distributed across observations, an OL model can be derived. The probability of each category can be expressed as

$$P(y > i|\boldsymbol{X}) = \frac{\exp(\boldsymbol{\alpha X} - \mu_i)}{1 + \exp(\boldsymbol{\alpha X} - \mu_i)} \tag{7}$$

An important restriction associated with an OL model is that the estimated coefficients between each pair of outcome groups are the same. This restriction is known as the proportional odds assumption or the parallel regression assumption (*23*). The use of the OL may be inappropriate if this assumption is violated. Instead, the PPO model formulates part of the coefficients as identical values while varying the other values across the different levels of the response variable in Equation 8:

$$P(y > i|\boldsymbol{X}) = \frac{\exp(\boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 - \mu_i)}{1 + \exp(\boldsymbol{X}_1\boldsymbol{\alpha}_1 + \boldsymbol{X}_2\boldsymbol{\alpha}_2 - \mu_i)} \tag{8}$$

where the coefficients of $\boldsymbol{\alpha}_1$ are maintained across the injury severities while the coefficients of $\boldsymbol{\alpha}_2$ are changing across the injury severities. Whether the coefficient is the same or different depends on the proportion assumption.

ML Model

When unobserved factors affect injury severities and cause the parameters to vary across observations, the ML model defines $\boldsymbol{\beta}_i$ as a vector of estimable parameters for discrete choice i with a probabilistic distribution. The outcome probabilities are defined as (*16*)

$$P_n(i) = \int \frac{\exp(\boldsymbol{\beta}_i \boldsymbol{X}_{in})}{\sum_{\forall I} \exp(\boldsymbol{\beta}_I \boldsymbol{X}_{In})} q(\boldsymbol{\beta}_i|\boldsymbol{\varphi}) d\boldsymbol{\beta}_i \tag{9}$$

where $q(\boldsymbol{\beta}_i|\boldsymbol{\varphi})$ is a density function of $\boldsymbol{\beta}_i$ and $\boldsymbol{\varphi}$ is a vector of parameters that describe the density function; all other terms are as previously defined. The density function can be a normal, lognormal, or uniform distribution, among others.

DATA COLLECTION AND ANALYSIS

Between 2004 and 2009, there were 15,393 traffic accidents involving at least one large truck in Wisconsin, accounting for 2.1% of all crashes (*24*). Large trucks include single unity trucks (two axles and three axles), trucks or tractors (double, semitrailer, or triple), buses, school buses, and other unknown heavy trucks (*25*). Large-truck crashes were identified via a large-truck flag assigned by the Wisconsin Department of Motor Vehicles and were retrieved from the online Wisconsin crash database through the WisTransportal System (*24*). Among the large-truck crashes, 7,652 (49.71%) were PDOs; 3,076 (19.98%) were injury Type C; 2,843 (18.47%) were injury Type B; 1,410 (9.16%) were injury Type A; and 412 (2.68%) were fatal injuries. To obtain sufficient observations in each category, crash injury severities were aggregated into three categories: PDO; B + C, which combined Type B and Type C injuries, and K + A, which combined Type A and fatal injuries. After combining, 5,919 crashes were either Type B or Type C injury crashes and 1,822 were either Type K or Type A injury crashes.

Crash data elements were classified into four categories: human factors, highway and traffic conditions, accident characteristics, and environmental factors. The driving record of the driver who sustained the most severe injuries in the collision was collected and the factors were used in the analysis. Human factors included driver behavior and characteristics. Driver behavior can be an officer's opinion of the possible contributing circumstances of a driver (e.g., exceeding the speed limit, failing to yield right of way) or an officer's observation (e.g., driving under the influence of drugs or alcohol, the use of safety constraints). Driver characteristics, such as age and gender, were noted for the vehicle whose occupant(s) withstood the most-severe injuries. Highway and traffic conditions included the highway geometric characteristics and traffic control types and were of primary interest. Here as well, the opinions of an officer regarding any possible contributing circumstances relating to a highway were considered. Accident characteristics consisted of objects struck and the manner of collision. A small number of crashes (1.7% of total crashes) involved a truck and vulnerable road users, such as pedestrians, bicyclists, or motorcyclists, and they were excluded from the data set because of the small sample size. Environmental factors included both weather and pavement conditions, although, intuitively, the roadway pavement conditions were the direct result of weather conditions. Both the weather and pavement conditions were included because the Pearson product–moment coefficient showed a weak correlation between the two factors, indicating varying effects of weather phenomena. Table 1 describes the selected variables, their frequency, and their percentage in the crash data.

TABLE 1 Description of Selected Variables

Variable	Type	Description	Frequency	Percentage
Human Factors				
Young	Dummy	The age of the most severely injured driver is younger than 25.	3,065	19.91
Old	Dummy	The age of the most severely injured driver is older than 55.	3,027	19.66
Female	Binary	Indicates the most severely injured driver is a female	3,986	25.89
Alcohol	Binary	1 indicates an alcohol flag.	499	3.24
Drug	Binary	1 indicates a drug flag.	95	0.62
Safety constraints	Binary	1 indicates the most severely injured driver used safety equipment.	15,379	99.91
Speed	Dummy	Speed-related factors (exceeding speed limit, too fast for conditions)	5,588	36.30
Rule violation	Dummy	Violation of the traffic rules (failed to yield the right-of-way, disregarded the traffic controls, improper turning)	4,260	27.67
Reckless behavior	Dummy	Reckless driving behavior (improper overtake, unsafe braking, following too close)	5,568	36.17
Highway and Traffic Conditions				
Roadhor	Binary	The road terrain at the point of impact is horizontal.	2,016	13.10
Roadvert	Binary	The road terrain at the point of impact is vertical.	2,811	18.26
Debris	Dummy	Presence of debris before the accident	213	1.38
Visibility	Dummy	Visibility obscured	364	2.36
Traffic control	Categorical	The type of traffic control at intersection		
	Signal	Traffic signal	2,009	13.05
	Two way	Two-way stop	1,486	9.65
	Four way	Four-way traffic stop	305	1.98
	Yield–none	Yield controlled or no traffic control	135	0.88
Accident Characteristics				
Guardrail	Dummy	The truck hit the guardrail of the highway.	221	1.44
Median barrier	Dummy	The truck hit the median barrier of the highway.	230	1.49
Bridge	Dummy	Accident caused by the circumstance of a bridge (parapet, pier, rail)	119	0.77
Ditch	Dummy	The truck ran into a ditch.	464	3.01
Tree	Dummy	The truck hit a tree.	122	0.79
Pole	Dummy	The truck hit a pole on the road (traffic sign or utility pole).	301	1.96
Jackknife	Dummy	The truck jackknifed.	256	1.66
Overturn	Dummy	The truck overturned.	849	5.52
Mnrcoll	Categorical	Manner of collision		
	Angle	Angle	3,994	25.95
	Head	Head on	502	3.26
	Rear	Rear end	3,963	25.75
	SSS	Sideswipe–same direction	718	4.66
	SSO	Sideswipe–opposite direction	2,724	17.70
	None	No collision with other vehicles	3,492	22.68
Trk–trk	Dummy	Truck with truck	2,899	18.83
Trk–pc	Dummy	Truck with passenger cars	8,689	56.45
Environmental Factors				
Weather–fog	Dummy	The weather was foggy.	221	1.44
Weather–rain	Dummy	The weather was raining.	1,034	6.72
Weather–sleet	Dummy	The weather was sleeting or hailing.	215	1.40
Weather–snow	Dummy	The weather was snowing.	2,323	15.09
Weather–wind	Dummy	The weather was windy.	201	1.31
Dark	Dummy	Nighttime without street lights	2,840	18.45
Light	Dummy	Nighttime with street lights	1,091	7.09
Ice	Dummy	Icy road surface	1,147	7.45
Snow	Dummy	Snow or slush road surface	2,410	15.66
Wet	Dummy	Wet road surface	2,037	13.23

ANALYSIS OF RESULTS AND DISCUSSION

Tables 2 to 4 show the respective model results for MNL, PPO, and ML. The STATA commands mlogit, gologit2, and mixlogit were used to estimate the coefficients for the MNL, PPO, and ML models, respectively (*26–28*). The interpretation of each model varies because the methodology is different. In the MNL model (Table 2), each coefficient defines the log odds ratio between the probabilities for a specific injury type and no injuries. In the PPO model (Table 3), the cumulative probability was used to derive the coefficients, which estimate the log odds ratio between the possibilities of injury severities greater than level *i* and the sum of the possibilities of injury severities less than or equal to level *i*. For example, the second column of headers quantifies the log odds ratio between the probabilities of all injury levels and PDO. The presentation of coefficients for fixed parameters in the ML model (Table 4) is the same as for the MNL model except when the coefficient is a random variable, in which case the standard deviation of the coefficient is displayed.

Model Results

In the MNL model, the coefficient estimates are explained as the comparison between injury level *i* and the base level PDO. As shown in Table 2, a driver usually sustained more severe injuries if alcohol or drug use was involved. If the driver had been drinking, the probabilities of Level B or C and Level K or A were 1.21 ($e^{0.191}$) times and 2.34 ($e^{0.850}$) times, respectively, higher than that of PDO. Similarly, if a driver was under the influence of drugs, his or her chance of getting injured increased drastically, with probabilities of Level B or C and Level K or A being 4.57 times and 16.12 times, respectively, that of PDO. Other factors relating to unsafe driving behavior, such as exceeding the speed limit, violating the traffic rules, and driving recklessly, all suggest an increased probability of serious injuries. In contrast, good driving behavior, such as the use of safety constraints, can have a positive impact. The probability of no injuries was five times higher than the probability of being injured if safety constraints were used. There were also several environmental

TABLE 2 Coefficient Estimates for MNL Model

	B + C				K + A			
Variable	Coeff.	SE	*z*	$P > \|z\|$	Coeff.	SE	*z*	$P > \|z\|$
Intercept	–0.870	0.320	–2.740	.010	–2.910	0.440	–6.550	.000
Human Factors								
Young	0.143	0.050	2.940	.003	0.183	0.080	2.430	.015
Old	–0.017	0.050	–0.360	.718	0.450	0.070	6.560	.000
Female	0.879	0.040	20.510	.000	0.619	0.060	9.570	.000
Alcohol	0.191	0.110	1.740	.083	0.850	0.130	6.360	.000
Drugs	1.520	0.340	4.490	.000	2.781	0.340	8.090	.000
Safety constraints	–1.482	0.280	–5.260	.000	–1.907	0.360	–5.300	.000
Speed	0.422	0.050	9.210	.000	0.667	0.070	9.530	.000
Rule violation	0.328	0.050	6.580	.000	1.080	0.070	14.930	.000
Reckless behavior	0.205	0.040	4.900	.000	0.452	0.060	7.190	.000
Highway and Traffic Conditions[a]								
Debris	–0.509	0.170	–3.000	.003	–0.231	0.260	–0.900	.368
Visibility	0.091	0.130	0.720	.469	0.565	0.160	3.530	.000
Signal	0.841	0.142	5.920	.000	0.734	0.260	2.820	.005
Two way	0.873	0.148	5.910	.000	1.423	0.260	5.500	.000
Yield–none	0.627	0.140	4.610	.000	0.992	0.250	3.960	.000
Accident Characteristics								
Total units	0.461	0.027	17.200	.000	0.630	0.030	21.870	.000
Jackknife	–0.931	0.210	–4.380	.000	–1.231	0.460	–2.650	.008
Overturn	0.737	0.080	8.720	.000	0.774	0.140	5.600	.000
Ditch	0.412	0.110	3.770	.000	0.236	0.200	1.200	.232
Median barrier	0.161	0.160	1.000	.317	0.587	0.220	2.720	.006
Environmental Factors								
Weather–snow	–0.295	0.080	–3.740	.000	0.135	0.130	1.030	.304
Snow	–0.397	0.080	–5.140	.000	–1.087	0.140	–7.880	.000
Ice	–0.478	0.090	–5.570	.000	–0.415	0.130	–3.130	.002
Wet	0.032	0.050	0.590	.556	–0.330	0.090	–3.700	.000
Dark	0.137	0.050	2.750	.006	0.439	0.070	6.030	.000

NOTE: Akaike information criteria (AIC) = 27,291.03; likelihood ratio χ^2 (df = 48) = 2,545.97; probability > χ^2 = 0; log likelihood (LL) = –13,620.515; coeff. = coefficient; SE = standard error.
[a]Four way is base level for all models.

TABLE 3 Coefficient Estimates for PPO Model

	$\frac{p(\text{K + A and B + C})}{p(\text{PDO})}$				$\frac{p(\text{K + A})}{p(\text{B + C and PDO})}$			
Variable	Coeff.	SE	z	$P > \lvert z \rvert$	Coeff.	SE	z	$P > \lvert z \rvert$
Intercept	–0.930	0.257	–3.620	.000	–2.712	0.259	–10.470	.000
Human Factors								
Young	0.123	0.043	2.840	.004	0.123	0.043	2.840	.004
Old	0.093	0.044	2.110	.035	0.457	0.062	7.390	.000
Female	0.816	0.041	19.920	.000	0.147	0.059	2.490	.013
Alcohol	0.384	0.100	3.840	.000	0.730	0.117	6.240	.000
Drugs	1.873	0.204	9.160	.000	1.873	0.204	9.160	.000
Safety constraints	–1.385	0.216	–6.400	.000	–1.385	0.216	–6.400	.000
Speed	0.483	0.041	11.890	.000	0.483	0.041	11.890	.000
Rule violation	0.485	0.046	10.520	.000	0.870	0.062	14.030	.000
Reckless behavior	0.296	0.037	8.010	.000	0.296	0.037	8.010	.000
Highway and Traffic Conditions[a]								
Debris	–0.317	0.151	–2.100	.036	–0.317	0.151	–2.100	.036
Visibility	0.203	0.116	1.750	.080	0.537	0.142	3.770	.000
Signal	0.816	0.138	5.910	.000	0.338	0.156	2.160	.031
Two way	1.041	0.140	7.450	.000	1.041	0.140	7.450	.000
Yield–none	0.810	0.138	5.880	.000	0.810	0.138	5.880	.000
Accident Characteristics								
Total units	0.476	0.025	19.070	.000	0.281	0.020	14.310	.000
Jackknife	–1.005	0.197	–5.090	.000	–1.005	0.197	–5.090	.000
Overturn	0.739	0.080	9.220	.000	0.236	0.129	1.830	.067
Ditch	0.369	0.104	3.570	.000	–0.108	0.190	–0.570	.569
Median barrier	0.410	0.135	3.040	.002	0.410	0.135	3.040	.002
Environmental Factors								
Weather–snow	–0.179	0.074	–2.430	.015	0.395	0.118	3.350	.001
Snow	–0.549	0.073	–7.500	.000	–0.978	0.124	–7.870	.000
Ice	–0.380	0.076	–4.990	.000	–0.380	0.076	–4.990	.000
Wet	–0.040	0.052	–0.770	.442	–0.363	0.083	–4.380	.000
Dark	0.210	0.046	4.530	.000	0.356	0.067	5.290	.000

NOTE: The variables in italic represent that their coefficients should vary across levels. AIC = 27,219.05; likelihood ratio χ^2 (38) = 2,647.94; probability > χ^2 = 0; LL = –13,569.526.
[a]Four way is base level for all models.

factors, such as snow and snowy surfaces, that were associated with decreased injury severities.

The PPO model treats crash severity as an ordinal response variable but allows the coefficients to vary across levels if the proportional odds assumption is violated. Table 3 shows a high chance of severe injury if the driver was under the influence of alcohol. The probability of Level K + A and B + C injuries is 1.47 ($e^{0.384}$) times that of PDO, while the possibility of Level K + A injuries is expected to be 2.08 ($e^{0.730}$) times higher than that of Level B + C and PDO. Similarly, drugs appear to consistently and substantially increase the possibility of injuries of all levels. In contrast, the use of safety constraints can consistently and effectively decrease the possibility of all levels of injuries. The traffic control strategy variables imply that four-way-stop–controlled intersections have the lowest truck crash severities and two-way-stop–controlled intersections have the highest crash severities. Signalized intersections tend to have a smaller possibility of Level K or A injuries than Level B or C injuries.

The ML model is able to account for data heterogeneity by treating coefficients as random variables. The selection of random variables followed the general procedures documented by Moore et al. (*29*). The first step was to select all the parameters as random parameters; the second step was to reduce one random parameter at a time until no further reduction of the random variables could be made. According to the model outputs in Table 4, the parameter for speed-related factors for Injury Level B or C is normally distributed with mean 0.456 and standard deviation 0.71, indicating 73.96% [$P(Z > -0.642) = 73.96\%$] of the distribution is greater than 0 and 26.04% of the distribution is less than 0 (*18*). This phenomenon indicates that 73.96% of the crashes occurred as a result of excessive speed, which led to a higher probability of Type B or C injuries than that of PDOs, while 26.04% of the crashes with a speed-related cause had a lower possibility of Type B or C injuries. Similarly, the environmental parameter of snowy surface for Type B or C truck crash severity was normally distributed with mean –0.967 and standard deviation 2.665. On a snowy pavement,

TABLE 4 Coefficient Estimates for the ML Model

	B + C				K + A			
Variable	Coeff.	SE	*z*	*P* > |*z*|	Coeff.	SE	*z*	*P* > |*z*|
Intercept	−1.962	0.081	−24.140	.000	−4.400	0.122	−35.950	.000
Human Factors								
Old	0.027	0.050	0.540	.594	0.502	0.073	6.870	.000
Female	0.971	0.048	20.229	.000	0.354	0.189	1.870	.061
Standard deviation	NA	NA	NA	NA	0.999	0.316	3.160	.002
Alcohol	0.231	0.116	1.991	.047	0.344	0.431	0.800	.425
Standard deviation	NA	NA	NA	NA	2.004	0.788	2.540	.011
Drugs	1.663	0.363	4.581	.000	3.180	0.390	8.150	.000
Speed	0.456	0.523	0.872	.000	0.774	0.078	9.950	.000
Standard deviation	0.710	0.300	2.367	.020	NA	NA	NA	NA
Rule violation	0.268	0.053	5.057	.000	1.052	0.081	12.940	.000
Reckless behavior	0.189	0.044	4.295	.000	0.491	0.070	7.000	.000
Highway and Traffic Conditions[a]								
Debris	−0.600	0.180	−3.333	.001	−0.297	0.281	−1.060	.290
Visibility	0.146	0.140	1.043	.297	0.676	0.183	3.680	.000
Signal	0.866	0.143	6.056	.000	0.956	0.263	3.635	.000
Two way	0.887	0.150	5.913	.000	1.299	0.261	4.977	.000
Yield–none	0.642	0.142	4.521	.001	1.137	0.240	4.738	.000
Accident Characteristics								
Total units	0.539	0.033	16.500	.000	0.781	0.038	20.400	.000
Jackknife	−1.074	0.262	−4.100	.000	−1.049	0.535	−1.960	.050
Overturn	0.877	0.093	9.450	.000	1.001	0.149	6.700	.000
Ditch	0.539	0.117	4.610	.000	0.443	0.212	2.080	.037
Median barrier	0.235	0.179	1.320	.188	0.592	0.248	2.380	.017
Environmental Factors								
Weather–snow	−0.475	0.111	−4.290	.000	0.264	0.178	1.490	.137
Snow	−0.967	0.179	−5.400	.000	−11.076	3.925	−2.820	.005
Standard deviation	2.665	0.410	6.490	.000	8.727	2.791	3.130	.002
Ice	−0.422	0.094	−4.520	.000	−0.561	0.154	−3.640	.000
Wet	0.030	0.056	0.550	.585	−0.390	0.095	−4.100	.000
Dark	0.143	0.055	2.590	.009	0.615	0.083	7.410	.000

NOTE: Italic coefficients are random variables. NA = not available. AIC = 27,537.94; likelihood ratio $\chi^2(5) = 122.02$; probability > $\chi^2 = 0$; LL = −13,717.969.
[a]Four way is base level for all models.

35.94% of the truck crashes had an increased possibility of Type B or C injuries and 64.06% of the truck crashes had a decreased possibility of Type B or C injuries. For K or A injuries, the parameter was normally distributed with mean −11.076 and standard deviation 8.727. This implies that the majority of truck crashes (89.8%) on snowy pavement may have a decreased possibility of Type A or K injuries. This observation coincided with the findings of Chen and Chen concerning the influence of snow or slush on the road surface (*18*). It is plausible that people often drive more slowly and cautiously on snowy roads but the slick conditions still have a tendency to cause accidents.

The results of the three logistic models provide a unique perspective on how various factors affect crash severity. Some factors have consistent impacts across all levels while others have varied impacts. A direct comparison of the parameters between the models is difficult because each methodology is different. For comparison, the coefficient estimates need to be converted to marginal effects.

Marginal Effects for Logistic Models

The marginal effects calculate how the injury severity probabilities change with a small (unit) change in an explanatory variable. For continuous variables, the marginal effects can be explained as the difference in the probability at each level following a one-unit change in the independent variables; for dummy variables, the marginal effects can be calculated as the changes in the probabilities for each level caused by a change in the value of the dummy variable from 0 to 1. The marginal effects of the three models are listed in Table 5.

Human Factors

The human factors that are statistically significant at the 5% level and have consistent effects on crash severities in all models are drugs,

TABLE 5 Marginal Effects of MNL, PPO, and ML Models

Variable	PDO MNL	PDO PPO	PDO ML	B + C MNL	B + C PPO	B + C ML	K + A MNL	K + A PPO	K + A ML
Human Factors									
Young	–0.039	–0.031	NS	0.026	0.019	NS	0.013	0.012	NS
Old	–0.024	–0.023	–0.021	NS	–0.022	NS	0.050	0.045	0.050
Female	–0.202	–0.198	–0.189	0.186	0.184	0.172	0.016	0.014	NS
Alcohol	–0.089	–0.094	–0.086	NS	0.009	0.003	0.094	0.085	NS
Drugs	–0.367	–0.359	–0.336	0.055	0.043	0.051	0.312	0.316	0.285
Safety constraints	0.322	0.294	NS	–0.192	–0.089	NS	–0.130	–0.205	NS
Speed	–0.116	–0.120	–0.100	0.068	0.074	0.055	0.048	0.046	0.045
Rule violation	–0.144	–0.120	–0.092	0.035	0.028	0.010	0.109	0.092	0.082
Reckless behavior	–0.059	–0.074	–0.049	0.025	0.046	0.018	0.034	0.028	0.031
Highway and Traffic Conditions[a]									
Debris	0.109	0.079	0.108	–0.107	–0.054	–0.112	NS	–0.025	NS
Visibility	–0.052	–0.050	–0.051	–0.012	–0.008	NS	0.064	0.058	0.030
Signal	–0.236	–0.240	–0.172	0.169	0.111	0.106	0.067	0.129	0.066
Two way	–0.212	–0.194	–0.176	0.186	0.161	0.123	0.026	0.033	0.053
Yield–none	–0.192	–0.193	–0.176	0.158	0.100	0.094	0.034	0.093	0.082
Accident Characteristics									
Total units	–0.125	–0.119	–0.225	0.085	0.094	0.187	0.040	0.025	0.038
Jackknife	0.233	0.236	0.212	–0.172	–0.175	–0.168	–0.061	–0.061	–0.044
Overturn	–0.177	–0.176	–0.176	0.135	0.153	0.144	0.042	NS	0.032
Ditch	–0.091	–0.091	–0.097	0.086	0.100	0.096	NS	NS	0.001
Median barrier	–0.063	–0.100	–0.056	NS	0.058	NS	0.060	0.042	0.051
Environmental Factors									
Weather–snow	0.051	0.045	0.063	–0.076	–0.084	–0.100	NS	0.039	NS
Snow	0.125	0.136	0.109	–0.059	–0.069	–0.035	–0.066	–0.067	–0.074
Ice	0.113	0.095	0.104	–0.096	–0.065	–0.071	–0.017	–0.030	–0.033
Wet	0.007	0.010	0.009	NS	NS	NS	–0.030	–0.029	–0.031
Dark	–0.049	–0.052	–0.043	0.006	0.018	–0.002	0.043	0.034	0.045

NOTE: NS = variable is not statistically significant at 5% level of significance.
[a]Four way is base level for all models.

speed, rule violation, and reckless behavior. Among these variables, drugs have a substantial effect on increasing the probability of Type B or C injuries and a major effect on increasing the probability of Type K or A injuries. Speed and traffic rule violations have an evident impact on probabilities of Type B or C injuries. For Injury Type K or A, the augmented effect of traffic rule violation is more apparent than speed. Reckless behavior also has a noticeable effect on increasing the probability of all levels of injury. The use of safety constraints decreases the probability of Injury Type B or C, but the result is more noticeable for Injury Type K or A. However, the effect of use of safety constraints cannot be identified in the ML model. The ML model suggests that alcohol should be considered as a random variable with zero mean; but in both the MNL and PPO models, alcohol is statistically significant in increasing the severity of injury to Level K or A.

Highway and Traffic Conditions

Four-way–stop controlled intersections have the lowest truck crash severity, possibly because of the low posted speed limit and vehicle travel speed at these intersections. The effects of other intersection traffic controls vary across the models. The consensus is that their positive effect on injury level K or A is lower than that on injury level B or C. In other words, compared with four-way-stop intersections, the other traffic control types may increase the possibility of Type B or C injuries but only slightly increase the likelihood of Type K or A injuries. Visibility appears to increase the possibility of Type K or A injuries but decrease the possibility of Type B or C injuries. The presence of debris had a negative effect on injury severities, an indication that many drivers may slow down after observing debris and avoid accidents.

Accident Characteristics

If a truck did not strike a fixed object, it may have overturned or jackknifed. The crash severity outcome of an overturned vehicle is much more severe than that of a jackknifed one. Most jackknifed truck crashes did not cause any injuries; this result was not the same for overturned trucks. Although large trucks are sturdier than passenger cars, injuries were still prevalent if a truck struck a ditch or a median barrier. Although the size difference between vehicles affects the injury severity effects of collision, the variables Trk–Trk and Trk–pc (Table 1) were not statistically significant at the 5% level. It is plausible that both variables may confound with other factors, such as accident types and human factors, which had stronger correlation with injury severity.

Environmental Factors

With the exception of the ML model, the possibility of B or C level injuries on darkened roadways and areas without street lights increased by less than 2%, while the possibility of Level K or A injuries increased by more than 3%. All the coefficients of adverse surface conditions, such as snow, ice, and wet, seem to be associated with lower injury severities in truck crashes, possibly because of the reduced speed most drivers use on slippery pavement.

Model Comparisons

The previous sections discussed the variables in detail to ensure that the directions of the indicated effects and values were logical. Overall, the results were consistent between models. Some variations provided new insights about the data as well as their effects. Because all three logistic models have been popular in modeling crash severities, their performance can be evaluated by statistical metrics such as goodness of fit, prediction accuracy, and the number of statistically significant variables. The results of these comparisons are in Table 6.

The Akaike information criterion (AIC) is a measure of the relative goodness of fit of a statistical model. The general formula is $\text{AIC} = 2k - 2\ln(L)$, where k is the number of parameters in the statistical model and L is the maximized value of the likelihood function for the estimated model. The preferred model is the one with the minimum AIC value. According to the AIC values in Table 6, the PPO models fit the data slightly better than the other two models. The overall prediction accuracy of the models was compared between the observations and the estimation results. The proportions of injury severities are arranged in rows by model; values in parentheses measure the difference between the predicted and observed proportions. On the basis of this standard, the ML model performed the best because the predicted probabilities for all three levels of injury severities are the closest to the observations. Both the MNL model and the PPO model underestimated K + A injuries and overestimated B + C injuries. Note that the prediction accuracy is an aggregated value that may not represent the accuracy at the individual level. For instance, the ML model was unable to identify alcohol, drugs, and safety constraints as statistically significant variables for the K or A injury levels. On the basis of all the information available, the PPO model is recommended because it achieved the lowest AIC value, had the most variables significant at the 5% level, and satisfied logical effects on the crash severity.

CONCLUSIONS

Crash severity issues have been extensively studied in the past decades and numerous statistical methodologies have been utilized to define the relationship between injury severity and its determinants. Different models sometimes return mixed results, making it difficult for decision makers to choose a reliable model. This study employed three of the most representative logistic models: MNL, PPO, and ML. The models were used to investigate the contributing factors of crash severity for an exclusive large-truck crash data set. On the basis of the results, the PPO model was the recommended choice because it demonstrated the best predictive power, had the greatest number of variables that were significant at the 5% level, and satisfied logical effects on crash severity.

The study shows that in the human factor category, drivers who were under the influence of drugs, speeding (exceeding the speed limit or driving too fast for the road conditions), committing rule violations (failing to yield the right-of-way, disregarding traffic controls, or improper turning), or driving recklessly (improper overtake, unsafe braking, following too close) illustrated consistent effects on crash severity in all model results at the 5% level of significance. The use of safety constraints and alcohol had the expected effects on crash severity but were not statistically significant at the 5% level in the ML model.

None of the highway alignment characteristics was statistically significant in any of the models. However, all three models suggested that four-way-stop–controlled intersections had the lowest injury severities; the effects of other intersection traffic control types varied from one model to another. Among all the roadside objects struck by large trucks, ditches and median barriers consistently caused increased injury at all levels. All the coefficients of surface conditions, such as snow, ice, and wet, seemed to be associated with lower injury severities, possibly because of reduced speeds. It was anticipated that the study results that were validated by the different models would shed light on the causes of large-truck crash severities. The study also offered additional insight on the application of statistical methodologies in safety analysis.

ACKNOWLEDGMENTS

The authors acknowledge the National Center for Freight and Infrastructure Research and Education for financial support for this study and the University of Wisconsin–Madison Traffic Operations and Safety Laboratory for provision of the data.

TABLE 6 Estimated Results and Performance of Models

Model	PDO (%)	B + C (%)	K + A (%)	AIC	Number of Statistically Significant Variables
Observed	49.71	38.45	11.84	na	na
MNL	48.70 (−1.01)	40.85 (2.40)	10.45 (−1.39)	27,291.03	17
PPO	48.26 (−1.45)	41.87 (3.42)	9.87 (−1.97)	27,219.05	22
ML	50.12 (0.41)	38.13 (−0.32)	11.75 (−0.09)	27,573.94	16

NOTE: na = not applicable.

REFERENCES

1. Bureau of Transportation Statistics, U.S. Department of Transportation. Freight in America. 2006. http://www.bts.gov/publications/freight_in_america/html/executive_summary.html. Accessed July 5, 2012.
2. Adams, T., K. Bekkem, and V. Bier. *Operational Resiliency of Beloit-Hudson Interstate Highway Corridor.* CFIRE, Wisconsin Department of Transportation 0092-09-10, CFIRE 01-09. Center for Freight and Infrastructure Research and Education, Wisconsin Department of Transportation, Madison, 2010. http://www.wistrans.org/cfire/documents/CFIRE_01-09_Final_Report.pdf. Accessed July 5, 2012.
3. NHTSA, U.S. Department of Transportation. Traffic Safety Facts, 2010 Data—Large Trucks. http://www.nrd.nhtsa.dot.gov/Pubs/811628.pdf. Accessed June 7, 2012.
4. Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention,* Vol. 43, No. 5, 2011, pp. 1666–1676.
5. Zajac, S. S., and J. N. Ivan. Factors Influencing Injury Severity of Motor Vehicle-Crossing Pedestrian Crashes in Rural Connecticut. *Accident Analysis and Prevention,* Vol. 35, No. 3, 2003, pp. 369–379.
6. Khattak, A. J., P. Kantor, and F. M. Council. Role of Adverse Weather in Key Crash Types on Limited-Access Roadways: Implications for Advanced Weather Systems. In *Transportation Research Record 1621,* TRB, National Research Council, Washington, D.C., 1998, pp. 10–19.
7. Kockelman, K. M., and Y. J. Kweon. Driver Injury Severity: An Application of Ordered Probit Models. *Accident Analysis and Prevention,* Vol. 34, No. 3, 2002, pp. 313–321.
8. Abdel-Aty, M., and J. Keller. Exploring the Overall and Specific Crash Severity Levels at Signalized Intersections. *Accident Analysis and Prevention,* Vol. 37, No. 3, 2005, pp. 417–425.
9. O'Donnell, C. J., and D. H. Connor. Predicting the Severity of Motor Vehicle Accident Injuries Using Models of Ordered Multiple Choice. *Accident Analysis and Prevention,* Vol. 28, No. 6, 1996, pp. 739–753.
10. Peterson, B., and F. E. Harrell. Partial Proportional Odds Models for Ordinal Response Variables. *Applied Statistics,* Vol. 39, No. 2, 1990, pp. 205–217.
11. Wang, X., and M. Abdel-Aty. Analysis of Left-Turn Crash Injury Severity by Conflicting Pattern Using Partial Proportional Odds Models. *Accident Analysis and Prevention,* Vol. 40, No. 5, 2008, pp. 1674–1682.
12. Wang, Z., H. Chen, and J. J. Lu. Exploring Impacts of Factors Contributing to Injury Severity at Freeway Diverge Areas. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2102,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 43–52.
13. Quddus, M. A., C. Wang, and S. G. Ison. Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. *Journal of Transportation Engineering,* Vol. 136, No. 5, pp. 424–435.
14. Yamamoto, T., J. Hashijib, and V. N. Shankar. Underreporting in Traffic Accident Data, Bias in Parameters and the Structure of Injury Severity Models. *Accident Analysis and Prevention,* Vol. 40, No. 4, 2008, pp. 1320–1329.
15. Ye, F., and D. Lord. Investigation of Effects of Underreporting Crash Data on Three Commonly Used Traffic Crash Severity Models: Multinomial Logit, Ordered Probit, and Mixed Logit. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2241,* Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 51–58.
16. Train, K. *Discrete Choice Methods with Simulation.* Cambridge University Press, Cambridge, United Kingdom, 2003.
17. Milton, J. C., V. N. Shankar, and F. L. Mannering. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis. *Accident Analysis and Prevention,* Vol. 40, No. 1, 2008, pp. 260–266.
18. Chen, F., and S. Chen. Injury Severities of Truck Drivers in Single- and Multi-Vehicle Accidents on Rural Highways. *Accident Analysis and Prevention,* Vol. 43, No. 5, 2011, pp. 1677–1688.
19. Chang, L. Y., and F. L. Mannering. Analysis of Injury Severity and Vehicle Occupancy in Truck and Non-Truck-Involved Accidents. *Accident Analysis and Prevention,* Vol. 31, No. 5, 1999, pp. 579–592.
20. Zhu, X., and S. Srinivasan. A Comprehensive Analysis of Factors Influencing the Injury Severity of Large-Truck Crashes. *Accident Analysis and Prevention,* Vol. 43, No. 1, 2011, pp. 49–57.
21. Wang, X., and K. M. Kockelman. Use of Heteroscedastic Ordered Logit Model to Study Severity of Occupant Injury: Distinguishing Effects of Vehicle Weight and Type. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1908,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 195–204.
22. Washington, S., M. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis,* 2nd ed. Chapman and Hall/CRC, Boca Raton, Fla., 2011.
23. Long, J. S. *Regression Models for Categorical and Limited Dependent Variables.* Sage Publications, Thousand Oaks, Calif., 1996.
24. Wisconsin Traffic Operations and Safety Laboratory. WisTransportal. http://transportal.cee.wisc.edu/. Accessed Sept. 3, 2011.
25. *Law Enforcement Officer's Instruction Manual for Completing the Wisconsin Motor Vehicle Accident Report Form (MV 4000).* Division of Motor Vehicles, Wisconsin Department of Transportation, Madison, 1998.
26. STATA command for mlogit. http://www.stata.com/help.cgi?mlogit. Accessed Oct. 5, 2011.
27. Williams, R. Generalized Ordered Logit/Partial Proportional Odds Models for Ordinal Dependent Variables. *Stata Journal,* Vol. 6, No. 1, 2006, pp. 58–82.
28. Arne, R. H. Fitting Mixed Logit Models by Using Maximum Simulated Likelihood. *Stata Journal,* Vol. 7, No. 3, 2007, pp. 388–401.
29. Moore, D. N., W. Schneider, P. T. Savolainen, and M. Farzaneh. Mixed Logit Analysis of Bicycle Injury Severity Resulting from Motor Vehicle Crashes at Intersection and Non-Intersection Locations. *Accident Analysis and Prevention,* Vol. 43, No. 3, 2011, pp. 621–630.

The Statistical Methods Committee peer-reviewed this paper.