Injury Severity of Multivehicle Crash in Rainy Weather

Soyoung Jung, Ph.D.[1]; Xiao Qin, Ph.D., P.E.[2]; and David A. Noyce, Ph.D., P.E.[3]

Abstract: As part of the Wisconsin road weather safety initiative, the objective of this study was to microscopically assess the factor effects on the severities of multivehicle-involved crashes on high-speed roadways during rainfall utilizing a sequential logistic regression approach. Research began by considering interstate freeways in Wisconsin. Weather-related factors considered in the research included estimated rainfall intensity, water film depth, temperature, wind speed and direction, and the car-following distance at the time of crash. With each crash observation, weather data were obtained through the three most adjacent weather station locations and the inverse-squared distance method. Nonweather factors such as roadway geometries, traffic conditions, collision manners, vehicle types, and driver and temporal attributes were also considered. Sequential logistic regression was applied to predict multivehicle crash severities in ascending (forward) and descending (backward) orders, respectively. The final model was selected on the basis of a combination of model performance, parameter significance, and prediction accuracies. The backward sequential logistic regression model produced the most desirable results for predicting crash severities in rainy weather in which deficiency of car following, wind speed, the first harmful spot, vehicle types, temporal, and at-fault driver-related actions at the crash moment were found to be statistically significant. These findings can be used to provide quantitative support of road weather safety improvements via weather warning systems, highway infrastructure enhancements, and traffic management.
DOI: 10.1061/(ASCE)TE.1943-5436.0000300.

CE Database subject headings: Highways and roads; Weather conditions; Rainfall; Vehicles; Traffic accidents; Traffic safety.

Author keywords: Road weather safety; Rain effect; Multivehicle crash severities; Sequential logistic regression.

Introduction

Road pavement surfaces become wet by precipitation with 0.01 in./h or more precipitation (Kokkalis and Panagouli 1998). The mean number of wet days is 120 days per year in the United States (U.S. Environmental Protection Agency 2006), which implies that approximately 33% of a year is under the effect of poor-weather conditions. This wet weather is known as one of the significant contributing factors to roadway safety. On the basis of the Bureau of Transportation Statistics (BTS 2006), approximately 12% of all motor vehicle fatal crashes on U.S. highways occurred in the adverse weather conditions from 1999–2006.

Specifically, average annual proportion of fatal crashes in rainy weather, on wet-pavement surface during rainfall, to total fatal crashes was 7.6%, which is the highest rate among all kinds of adverse weather conditions.

Wisconsin crash rate is consistent with the national data. From 1999–2006, Wisconsin fatal crashes that occurred in rainy weather are the most frequent of all fatal crashes that occurred in any other adverse weather conditions (Wisconsin Department of Transportation 2006). When considering severe crashes with injuries and fatalities that occurred on Wisconsin highways, 3,296 injury and fatal crashes on wet pavement during rainfall are reported by the Wisconsin Traffic Crash Facts during the same period. The number of Wisconsin severe crashes with human injuries and fatalities is also greatest in rainy weather conditions than all kinds of inclement weather conditions during the same period. Specifically on Wisconsin interstate highways, multivehicle crashes occurred more frequently than single-vehicle crashes in rainy weather. During this 8-year period, 899 multivehicle crashes occurred on Wisconsin interstate highways in rainy weather, approximately 1.5 times more than the number of single-vehicle crashes (Wisconsin Department of Transportation 2006).

Weather is frequently cited and found to affect the severe crashes in the past studies. The previous studies have been primarily focused on predicting the crash count or frequency by injury severity. As a result, corresponding weather variables were often employed in an aggregate format. For example, Shankar et al. (1995) explored the relationship between weather variables and overall highway crash frequency using a negative binomial model. In their study, maximum and average daily rainfall, monthly rainy days, and rainfall-horizontal curve interaction were significant to either increases or decreases in rear-end, sideswipe, fixed-object, and overturn collisions. Specifically, maximum daily rainfall, monthly rainy days, and rainfall-horizontal curve interaction were found to decrease rear-end collisions whereas average daily rainfall was identified to increase rear-end collisions.

Caliendo et al. (2007) investigated the effect of rain-related factors on the frequency of multilane road crash occurrence by comparing Poisson, negative binomial and negative multinomial regression models. As a result, wet-pavement surfaces caused by rain precipitation were found to be a highly significant variable to increase severe crashes. The association between the number of crashes and weather conditions was also identified in an urban

[1]Research Professor, Korea Advanced Institute of Science and Technology, Dept. of Civil and Environmental Engineering, 291 Daehak-ro, Yuseong-gu, Daejeon 130-743, Republic of Korea (corresponding author). E-mail: jung2@kaist.ac.kr

[2]Assistant Professor, South Dakota Univ., Dept. of Civil and Environmental Engineering, 148 Crothers Engineering Hall, Brookings, SD 57007. E-mail: Xiao.Qin@sdstate.edu

[3]Associate Professor, Dept. of Civil and Environmental Engineering, Univ. of Wisconsin-Madison, 1204 Engineering Hall, 1405 Engineering Dr., Madison, WI 53706. E-mail: noyce@engr.wisc.edu

Note. This manuscript was submitted on February 20, 2010; approved on May 19, 2011; published online on December 15, 2011. Discussion period open until June 1, 2012; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Transportation Engineering***, Vol. 138, No. 1, January 1, 2012. ©ASCE, ISSN 0733-947X/2012/1-50–59/$25.00.

freeway crash analysis in a study by Kopelias et al. (2007). The authors utilized a standard linear regression model to estimate the frequency of crash occurrence by injury severity and then combined all the crash frequency estimates by severity weighting factors. According to their findings, wetpavement was found to significantly increase crash frequency.

Disaggregate models are an effective alternative approach to quantify factor impact on injury severities because it reveals more detailed and specific information. The outcomes in a disaggregate approach are usually determined by the purpose of the study and the data availability. Weather, particularly rain-derived factors, can affect multivehicle crash severities differently by collision type. Shankar et al. (1996) estimated a nested logit model of crash severity on Washington rural interstate highways. In their study, wet-pavement rear-end crash indicator was found to increase the likelihood of possible injuries, capturing the effect of rear-end crashes occurring in rainy weather. They explained that rainy weather conditions made it difficult for drivers to see the vehicles in front of them, resulting in rear-end crashes. Similarly, a wet and slippery road surface was found to contribute to rear-end crashes significantly at signalized intersections compared to a dry road surface (Yan et al. 2005). In a study by Duncan et al. (1998), the interaction between wet pavement and pavement grade was also found to significantly increase all injury propensities. The authors used an ordered probit model to identify variables significantly influencing the levels of injury in truck-passenger car rear-end involvements on divided roadways.

Special attention has been given to vehicle types for multivehicle crash severity prediction. In the study by Duncan et al. (1998) wet grade was found to increase injury severity of passenger car occupants in truck-passenger car rear-end crashes. The study provided a possible explanation that injury severity in the rear-end crashes was influenced by following factors: attributes of the passenger car driver, the speed and braking differential between the truck and the car, the unique visibility limitations of the truck, and differences in the size and other performance characteristics between trucks and passenger cars. Haque et al. (2009) utilized a binary logit model to differentiate between at-fault and not-at-fault cases to identify the motorcyclist-related factors in their study. In the study, motorcyclists were more likely to be victims than at-fault in multivehicle crashes, and wet-road surface was found to increase the likelihood of at-fault crashes at nonintersections.

Rain-related effects on multivehicle crash severities have been identified along with roadway characteristics. Khorashadi et al. (2005) divided driver injury severity in large truck-involved crashes into four levels in their study: no injury, complain of pain, visible injury, and severe/fatal injury. Then, they explored the differences between the driver injuries using multinomial logit analysis. In their study, rainfall was found to increase the likelihood of complaint of pain injury only in urban area. The severities of head-on crashes that occurred on Connecticut two-lane roads were also predicted by utilizing ordered probit model in a study by Deng et al. (2006). In their study, it was found that a wet-roadway surface and narrow road segments were significantly related to more severe crashes.

Driver attributes such as age or gender also play important roles in injury severity associated with weather conditions. Hill and Boyle (2006) investigated the fatality and incapacitating injuries to occupants of passenger vehicles using a logistic regression model in their study. Their study showed that crashes in adverse weather conditions with rain, snow, or fog increased the risk of severe injuries to females that were 55 and older. According to their study, the result may reflect an inability to cope with poor-weather conditions, in part attributed to a lack of exposure resulting from avoiding such conditions when possible.

Though numerous studies have been conducted in hopes of identifying the contributing factors to crash severities in which wet roadway and rainy weather were considered (Shankar et al. 1995; Dissanayake and Lu 2002; Golob and Recker 2003; Donnell and Mason 2004; Yau 2004; Savolainen and Tarko 2005; Abdel-Aty and Pemmanaboina 2006; Qin et al. 2006; Yau et al. 2006; Caliendo et al. 2007; Kopelias et al. 2007; Khan et al. 2008). Nevertheless, the specific impacts of wet roads during rainfall were not specifically considered. This study identifies a variety of significant predictors that contribute to more serious multivehicle crashes using crashes occurred on wet-pavement surface during rainfall only.

Data Collection

The study area consisted of approximately 75 mi of southeastern Wisconsin highway segments including I-43, I-94, I-43/94, and I-43/894. The study area is shown in Fig. 1.

In the study area, average annual daily traffic and vehicle miles traveled were higher than any other interstate highway segments between 2004 and 2008. Generally, sufficient and microscopic data collection should be conducted to obtain reliable modeling results. Correspondingly, all of the following databases were commonly available in the study area to collect sufficient amount of detailed data, which can lead the reliable modeling of crash severity.

Multivehicle crashes occurring in rainy weather were obtained from Wisconsin Department of Transportation (WisDOT) crash database. In addition to controlling rainy weather condition, the crash data used in this study were filtered through several criteria to ensure data homogeneity: wet pavement, multivehicle involved, interstate highways divided by barriers, no construction zones, no hit-and-run crashes, and no pedestrian involved in a crash. Consequently, 535 crashes were retrieved in the study area from 2004–2008. Crash data set included variables indicating severity, roadway geometries, driver demographics, collision types, vehicle types, pavement conditions, and temporal and weather information.

In this study, fatal (Type K), incapacitating injury (Type A), and nonincapacitating injury (Type B) crashes were combined as the highest level of crash severity to obtain a meaningful sample size (Federal Highway Administration 2005). Possible injury (Type C) crashes were separated as the second highest level of crash severity because they were invisible injuries with drivers' pain complaints. Property damage only (PDO) crashes made up the lowest level of crash severity. Crash frequency and severity codes are provided in Table 1.

State Trunk Network (STN) highway log from WisDOT contains roadway geometric attributes, including the number and lane width for travel lane and shoulder as well as pavement surface. Using the STN highway milepost information, the geometric attributes were linked to crash data set.

Traffic detector data in southeast Wisconsin are collected every 30 s and archived in a data portal application called V-SPOC. The data can be aggregated up to 5, 10, and 15 min. In this study, average vehicle volume, speed, and occupancy data in 5-min intervals were obtained for 1 h prior to each crash. The associated standard deviation within 1 h prior to the crash was also collected to account for the difference in density between crashes and detector locations in the study area.

The objective of this study is to examine the impacts of various factors on injury severity of multivehicle crashes during rainfall. Comparing to dry weather conditions, rain-derived factors are additionally considered in the examination. Accordingly, one of the

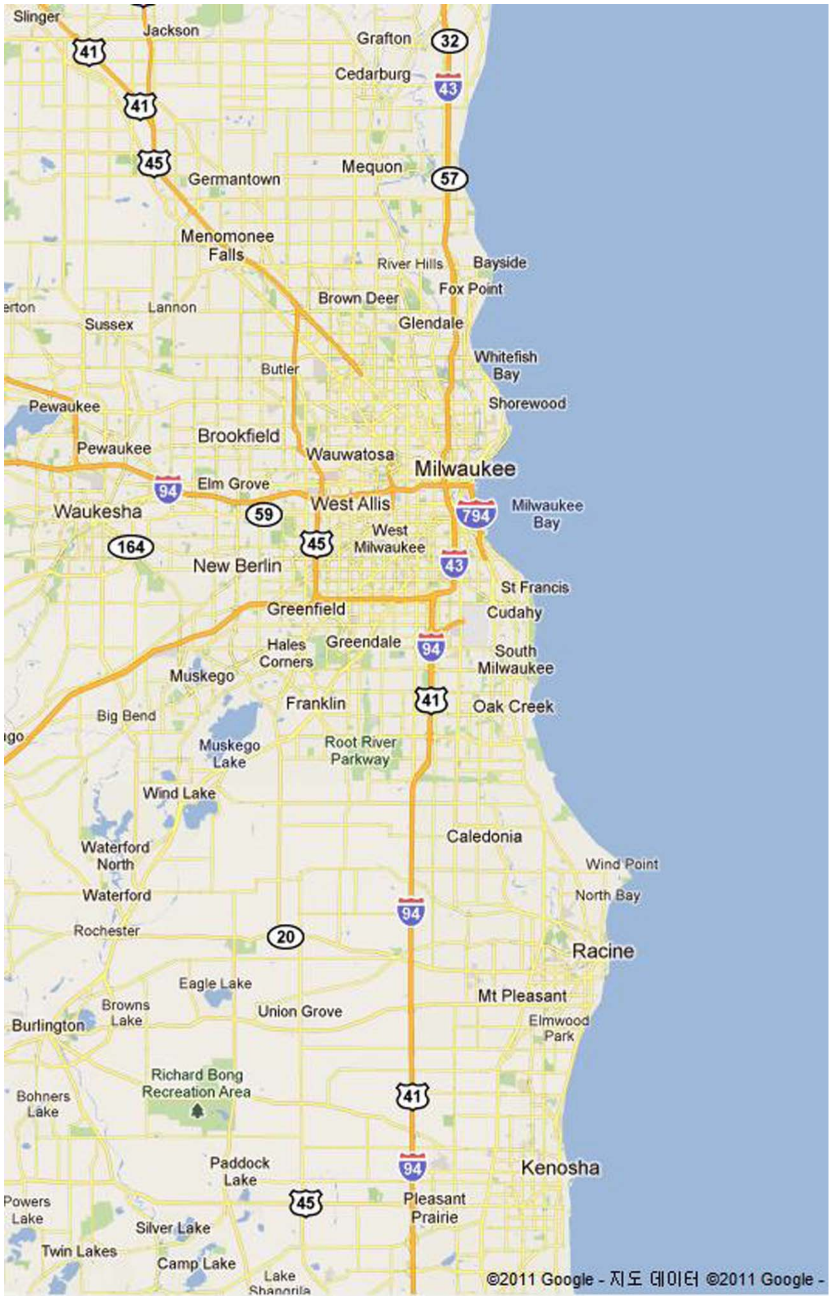

Fig. 1. Study area for injury severity of multivehicle crash in Wisconsin 2004–2008 (image courtesy Google Maps, © 2011 Google)

Table 1. Multivehicle Crash Frequency in Rainy Weather

Injury severity	Number of crashes	Sequential logistic regression: Forward format Stage 1	Forward format Stage 2	Backward format Stage 1	Backward format Stage 2
Fatal/incapacitating/nonincapacitating injury	46	1	1	1	—
Possible injury	147	1	0	0	1
PDO	342	0	—	0	0

most important tasks was to collect microscopic data of rainy weather at the time of crash. Given the limited weather data sources with minute-base measurements, a website operated by Weather Underground, combined with local road weather information system, was used. From there, 6 airport weather stations and 10 private weather stations were identified as the most reliable and real-time weather data source for this study.

Weather-Related Parameters

Data collected from selected weather stations included temperature, wind speed/direction, rainfall precipitation, and duration. Real-time weather data at specific crash locations were rarely observed because in most cases the weather stations are not located sufficiently near to the crash location. In this study, the real-time weather data at the crash location were approximated by interpolating between the nearest three weather stations. Average distance between each crash location and one-weather station closest to the crash location was found to be 2.6 mi. The data interpolation was applied to rainfall intensity and wind speed because they show special and temporal variation. Water film depth and deficiency of car-following distance (DCD) were estimated by the interpolated hourly rainfall precipitation, traffic, and road geometry data.

Because of the deficiency of actual weather data at the crash location, one closest weather station data, particularly rainfall intensity and wind speed were compared with the interpolated data created from three weather stations. As a result, weather data between interpolated and one-weather station measurements were similar within the range of 2.6-mi distance between the crash location and the closest weather station. However, the difference in weather data between interpolated weather stations and one-weather station tended to increase when each crash occurred far from the nearest weather station. This tendency was found to be consistent in wind speed data measurements. Results imply that weather data interpolation with several weather stations may be effective to approximate actual weather condition at a crash location as the distance between the crash location and the closest weather station increases.

Rainfall intensity is defined as rainfall precipitation divided by the duration, which reflects visibility on highways in rainy conditions. The average measurement interval of rainfall precipitation was 15 min. Compared to the weather data used in the previous studies, 15-min interval was a more precise measure of the real-time rainfall intensity at the crash moment.

The water film between the tire and pavement surface leads to a lower coefficient of friction. The depth of water film can be calculated through the following empirical equations (Russam and Ross 1968):

$$D = 0.046(WS'I)^{1/2}/S^{1/5} \tag{1}$$

$$S = (S_l^2 + S_c^2)^{1/2} \tag{2}$$

where D = water film depth (mm/h); I = rainfall intensity (mm/h); $S' = S/S_c$; S_l = longitudinal slope (%); S_c = slope of pavement cross section (%); and W = width of pavement (m).

In this study, there was no visibility data directly measured. Alternatively, deficiency of DCD was considered as a surrogate measure to the risk of having a multiple-vehicle crash. The DCD is defined as the difference between the minimum safe stopping distance (SSD) for a following vehicle and the actual distance between the lead and following vehicles. DCD is calculated by the following formula:

$$\mathrm{DCD} = \mathrm{SSD} - \mathrm{AVG} \tag{3}$$

SSD = minimum safe stopping distance; AVG = average vehicle gap. The AVG is calculated by subtracting average vehicle length from distance headway which is an inverse of density obtained from traffic detector data (Roess et al. 2004). The SSD is the minimum distance required to stop a vehicle and is determined by the following formula (AASHTO 2004):

$$\mathrm{SSD} = 1.47Vt + 1.075V^2/a \tag{4}$$

where V= vehicle speed (mi/h); t = brake reaction time (s); and a = deceleration rate (ft/s^2).

Kokkalis and Panagouli (1998) studied the coefficient of wet-pavement friction in great detail and developed the relationship among friction force, vehicle speed, and water film depth. Using pavement surface material information from Wisconsin STN highway log, a deceleration rate was applied to Eq. (4). In addition, 2.5 exceeding the 90th percentile of reaction time for all drivers was used for brake reaction time to encompass the capabilities of most drivers (AASHTO 2004).

Strictly speaking, the vehicle speed in Eq. (4) should be individual vehicular speeds, so the gap is between each pair of vehicles. Considering the varying responding times of the police officers, average 5-min traffic speeds for an hour prior to a crash occurrence were used to offset the deviation and that was the most reliable resource available to this study as the prevailing real-time traffic conditions at the crash moment.

As a result of data collection from several data sources, explanatory variables and the associated category coding used are shown in Table 2. Driver-related variables such as alcohol/drug use, safety belt use, etc. came from the WisDOT crash data set.

To estimate unknown weather data, a study regarding the comparison of several interpolation methods concluded that the inverse-squared distance method was stable and appropriate for the localized field with short-spatial correlation length scale and large variability (Patrick and Stephenson 1990). The minimum number of weather stations to apply the inverse-squared distance interpolation is three (Press et al. 2007). Therefore, in this study, the inverse-squared distance interpolation was utilized to estimate localized weather data for the crash.

Methodology

To model discrete outcome data, several modeling techniques such as traditional ordered probability, multinomial and nested logit models can be considered, but the application to the data set varies from one to another because of their limitations. Crash severities are inherently ordered multiple discrete outcomes. The multinomial and nested logit models do not account for the ordering of crash severities (Abdel-Aty 2003; Milton et al. 2008; Wang and Abdel-Aty 2008). The traditional ordered probability approaches also impose a critical restriction that regression parameters have to be the same for different response outcomes, so called proportional odds assumption. Supportively, the corresponding parameter restriction imposed by the traditional ordered probability approach were also addressed in past studies (Eluru and Bhat 2007; Wang and Abdel-Aty 2008).

Alternatively, a generalized version of the ordered logit model was used to relax the proportional odds assumption (Eluru et al. 2008). However, the generalized ordered response model with separate parameter coefficients across the ordered response levels is anticonservative and recommended only to conclude that the proportional odds assumption is valid (Peterson and Harrell 1990). Even though the generalized ordered logit model allows a separate coefficient for each predictor, the set of significant predictors is invariant over all the crash severity comparisons.

In our study, there are two important issues: (1) a variety of predictors particularly related to rainy weather across different injury severity levels and (2) prediction accuracy for crashes with apparently visible injuries. On the basis of the modeling issues, it is appropriate to choose sequential logistic regression because it not only accounts for the inherent order of crash severities but also allows different sets of regression parameters to be independently considered in the model specification. To achieve that, the sequential logistic regression describes severities via a series of standard logistic regression in a coherent manner. On the basis of the S-shaped cumulative density function for the logistic regression, the probability of a certain outcome in the standard logistic regression is found with the following formula (Kleinbaum and Klein 2002):

$$P(Y)/(1 - P(Y)) = \mathrm{EXP}(\alpha + \beta X) \tag{5}$$

Table 2. Explanatory Variables Used in Crash Severity Prediction Model

Variable	Min.	Max.	Mean	Category coding
At-fault driver's sex	—	—	—	Female=1 versus Male=0
Alcohol or drug	—	—	—	Under alcohol/drug effect=1 versus Sobriety=0
Safety belt	—	—	—	Use of safety belt=1 versus Nonused=0
At-fault driver's action	—	—	—	Going straight=1 versus Others=0
				Lane change/merging/overtaking=1 versus Others=0
				Negotiating curve=1 versus Others=0
				Slowing or stopped=1 versus Others=0
Curve direction	—	—	—	Curve to the right=1 versus Others=0
				Curve to the left=1 versus Others=0
Injury transport	—	—	—	Injured people transported to hospital=1 versus Others=0
Terrain	—	—	—	Horizontal curve=1 versus Others=0
				Vertical curve=1 versus Others=0
				Horizontal/vertical curve=1 versus Others=0
				Tangent/flat=1 versus Others=0
First harmful spot	—	—	—	Ramp/gore=1 versus Others=0
				Shoulder/outside shoulder=1 versus Others=0
				Median=1 versus Others=0
				On roadway=1 versus Others=0
Pavement surface	—	—	—	Asphaltic cement plant mix/rigid base=1 versus Others=0
Lighting condition	—	—	—	Daylight=1 versus Others=0
				Dusk/dawn/dark=1 versus Others=0
				Night but street light=1 versus Others=0
Crash type	—	—	—	Median related=1 versus Others=0
				Noncollision=1 versus Others=0
				Fixed object=1 versus Others=0
First harmful collision	—	—	—	Sideswipe=1 versus Others=0
				Rear end=2 versus Others=0
				Others=3 versus Others=0
At-fault driver's vehicle	—	—	—	Car=1 versus Others=0
				Truck(straight)/truck-tractor=1 versus Others=0
				Motorcycle=1 versus Others=0
Time of day	—	—	—	Peakhour (6-8 a.m. and 3-5 p.m.)=1 versus Offpeak=0
Day of week	—	—	—	Tuesday/Thursday=1 versus Others=0
				Monday/Friday=1 versus Others=0
				Saturday/Sunday= 1 versus Others=0
Quarter of year	—	—	—	December to February=1 versus Others=0
				March to May=1 versus Others=0
				June to August=1 versus Others=0
				September to November=1 versus Others=0
Wind direction	—	—	—	North=1 versus Others=0
				East=2 versus Others=0
				South=3 versus Others=0
				West=4 versus Others=0
At-fault driver's age	16	87	35	—
Number of vehicles	2	5	2	—
Number of lanes	1	4	3	—
Lane width (ft)	12	18	12	—
Shoulder width(ft)	0/0[a]	13/16	7/11	—
Speed limit (mi/h)	35	65	55	—
Average 5-min VOL	5	172	94	—
Average 5-min SPD	1	91	48	—
Average 5-min OCC (%)	0.45	49.11	13.00	—
S.D[5]. of VOL	0.94	69.46	9.78	—
S.D. of SPD	0.32	63.29	5.88	—
S.D. of OCC	0.10	18.23	3.02	—

Table 2. (*Continued.*)

Variable	Min.	Max.	Mean	Category coding
DCD (ft)	0	3825	153	$< 50 = 1$ versus Others=0
				(50, 225) =1 versus Others=0
				$> 225 = 1$ versus Others=0
Wind speed (km/h)	0.0	43.9	9.0	$< 2.6 = 1$ versus Others=0
				(2.6, 13.5) =1 versus Others=0
				$> 13.5 = 1$ versus Others=0
Temperature (ºC)	0.1	29.4	11.0	$< 5.0 = 1$ versus Others=0
				(5.0, 17.0) =1 versus Others=0
				$> 17.0 = 1$ versus Others=0
Water film (mm/h)	0.00	0.45	0.06	$< 0.02 = 1$ versus Others=0
				(0.02, 0.10) =1 versus Others=0
				> 0.10 =1 versus Others=0
RI (mm/15 min)	0.00	8.96	0.26	$< 0.05 = 1$ versus Others=0
				(0.05, 0.30) =1 versus Others=0
				$> 0.30 = 1$ versus Others=0

[a]Left shoulder width/right shoulder width.

where $P(Y)$ = probability of response outcome; Y = response variable; α = intercept parameter; β = vector of parameter estimate; and X = vector of explanatory variable.

An interpretation of the logistic regression model uses the odds and the odds ratio of an event. The odds of an event is a ratio of the probability that the event will occur divided by the probability that it will not. The odds ratio is a ratio of the predicted odds for a 1-unit change in X_i with other variables in the model held constant.

In this study, a series of standard logistic regression concept is applied at two stages to fit the sequential logistic regression model. At the second stage, a subsample is used after removing observations of a certain crash severity used in the previous stages (Maddala 1983). To explore whether there is an impact in the development of the sequential structure, forward and backward formats are conducted in the following way:

Forward format:

- Stage one: PDO versus others
- Stage two: Possible injuries versus fatal/incapacitating/nonincapacitating injuries

Backward format:

- Stage one: Fatal/incapacitating/nonincapacitating injuries versus others
- Stage two: Possible injuries versus PDO

Combining the fact that the sum of crash severity probabilities over the two stages of each format is one with Eq. (5), the probabilities of crash severity levels can be written as follows.

Forward format:

$$\text{Stage 1: } (1-P)/P1 = \text{EXP}(\alpha_1 + \beta X1) = h1 \tag{6}$$

$$\text{Stage 2: } P3/P2 = \text{EXP}(\alpha_2 + \beta X2) = h2 \tag{7}$$

$$P1 = 1/(1+h1) \tag{8}$$

$$P2 = h1/(1+h1)(1+h2) \tag{9}$$

$$P3 = h1h2/(1+h1)(1+h2) \tag{10}$$

Backward format:

$$\text{Stage 1: } P3/(1-P3) = \text{EXP}(\alpha_1 + \beta X1) = I1 \tag{11}$$

$$\text{Stage 2: } P2/P1 = \text{EXP}(\alpha_2 + \beta X2) = I2 \tag{12}$$

$$P1 = 1/(1+I1)(1+I2) \tag{13}$$

$$P2 = I2/(1+I1)(1+I2) \tag{14}$$

$$P3 = I1/(1+I1) \tag{15}$$

where $P1$ = probability of PDO; $P2$ = probability of possible injury; and $P3$ = probability of fatal/incapacitating/nonincapacitating injury.

Model Performance Measures

Likelihood ratio (LR) test and parameter estimate significance are typical measures of model performance and these measures were synthetically considered to select an optimal model in this study.

The LR test reveals whether or not a global null hypothesis for a specific model should be rejected. In other words, an estimated model containing at least one nonzero parameter coefficient is better than a constant only model when p-value of LR test is less than a conventional criterion. For this study, all of the two stage models are combined to estimate some injury severity levels through each format of the sequential logistic regression. Correspondingly, the parameter estimate significance at both stages is one of the most important criteria to measure a model performance.

In addition, prediction accuracy classification standard logistic regression model classifies an observation as an event if the estimated probability of the observation is greater than or equal to a given cutpoint. Otherwise, it is classified as a nonevent. In the statistical term, sensitivity measures the proportion of actual events that are also predicted to be such. Similarly, specificity measures the proportion of actual nonevents that are also predicted to be such. The overall predictive power of a model depends on the proportion

Table 3. Event Proportion in Study Area

Year	Forward format Stage 1	Forward format Stage 2	Backward format Stage 1	Backward format Stage 2
1999	0.35	0.20	0.07	0.30
2000	0.35	0.20	0.07	0.30
2001	0.37	0.19	0.07	0.32
2002	0.32	0.20	0.07	0.28
2003	0.30	0.22	0.06	0.25
2004	0.35	0.19	0.07	0.31
2005	0.34	0.28	0.10	0.27
2006	0.32	0.19	0.06	0.28
2007	0.35	0.24	0.08	0.29
2008	0.38	0.27	0.10	0.30
Minimum	0.32	0.19	0.06	0.25
Maximum	0.38	0.28	0.10	0.32
Average	0.34	0.22	0.08	0.29
Standard deviation	0.02	0.03	0.01	0.02
Coefficient of variation	0.07	0.16	0.19	0.07

of correctly predicted observations (i.e., the sum of sensitivity and specificity). In addition, there are two rates for incorrectly classified observations: false positive rate and false negative. The false positive rate is the ratio of the number of nonevents incorrectly classified as events to the total events whereas the false negative rate is a ratio of the number of events incorrectly classified as nonevents to the sum of total nonevents.

Even though the predictive power of a model can be measured for all severity levels, in traffic safety studies, sensitivity and false negative rate are normally emphasized for high-crash severities because of their enormous economic loss. Hence, in this study, a model that produces high sensitivity and low false negative rate at the classification stages for fatal/incapacitating/nonincapacitating injuries will be considered as a good one.

The prediction accuracy in a model can be different and is determined by a probability cutpoint. More severe crashes (events) do not frequently occur compared with less severe crashes (nonevents). From this perspective, the probability cutpoint may be determined by practical consideration. Because desirable prediction models should fit the field data well, probability cutpoints used in this study were determined by overall trends of actual event proportions in the field data. Because of insufficient samples in the high-injury levels, a bootstrap sampling method called jackknife procedure was applied. Each time, one observation was withheld from the data set used for building the model. The restricted model was then compared to the model using the full data set. The process repeated until all the observations were tested. As a result of the jackknife procedure, high- and similar prediction accuracies between an estimated model and the model in the validation process

Table 4. Forward Format of Multiple Sequential Logistic Regression

Stage 1						
	Chi-square		D.F.	Probability > Chi-square		
LR test	40.5099		6	< 0.0001		
Analysis of maximum likelihood estimates	Parameter		Estimate	Standard error	Odds ratio	Probability > Chi-square
	Intercept 1		1.8666	0.7135	—	0.0089
	Safety belt		−1.6353	0.6835	0.195	0.0167
	Median-related crash		0.9213	0.4105	2.513	0.0248
	DRV 2		−1.0483	0.3133	0.351	0.0008
	SDV		−0.0595	0.0224	0.942	0.0079
	DRV 4		−1.1587	0.3734	0.314	0.0019
	SDV*DRV 4		0.0654	0.0315	1.068	0.0376
Prediction accuracy ($P_{\text{cut-off}} = 0.34$)		Overall %	sensitivity %	Specificity %	FP %	FN %
	Estimated model	59	59	59	54	29
	Cross-validation	58	59	58	55	30
Stage 2						
	Chi-square		D.F.	Probability > Chi-square		
LR test	19.6940		3	0.0002		
Analysis of maximum likelihood estimates	Parameter		Estimate	Standard error	Odds ratio	Probability > Chi-square
	Intercept		−0.9280	0.2959	—	0.0150
	DCD 1		0.9487	0.3882	2.582	0.0145
	OCC		−0.0506	0.0244	0.951	0.0378
	Curve to the left		1.4631	0.9264	4.319	0.1043
Prediction accuracy ($P_{\text{cut-off}} = 0.22$)		Overall %	Sensitivity %	Specificity %	FP %	FN %
	Estimated model	71	54	76	58	16
	Cross-validation	70	54	77	58	16

suggest the estimated model is not overly influenced by the data used.

Results and Discussion

Exploratory data analysis (EDA) was conducted to effectively reduce the number of variables of interest because of large volume of data collected in the study. First of all, a logistic regression was performed to choose an individual predictor related to the crash severity. Weather parameters were tested either as a continuous or a discrete variable. Using all of predictors chosen by the single predictor selection, PROC LOGISTIC in SAS 9.1 (SAS Institute 1995) was used to estimate sequential logistic regression models with a significance level of 0.10 in which backward elimination with two-way interaction was conducted for each multiple model. The reason for the use of backward elimination is that the full variable set is calculated in the backward elimination procedure, being able to handle multicollinearity (Chatterjee et al. 2000). Because the predicted crash severity is a probability, it has to be classified into either the most severe injuries, possible injuries, or PDOs by some thresholds called cutpoints. The event probability cutpoints can be determined by the actual proportions of severe crashes (event). In this study, the 10-year average event proportions in Table 3 were used to establish the event probability cutpoints.

Comparing the multiple logistic regression models by goodness of fit, parameter significance, prediction accuracies, and cross-validation, the more appropriate logistic regression model for crash severity estimation was selected between forward and backward sequential logistic regression models.

Table 4 shows the results of the forward format of a multiple-sequential logistic regression. As shown from the LR test, the global null hypothesis is rejected, indicating the estimated models are better than constant only model. All of the explanatory variables are statistically significant.

At Stage 1, the likelihood of possible injury crashes decreased when at-fault driver wore safety belt, which is consistent with general expectation. Meanwhile, median-related crash type was likely to increase crash severity. Interestingly, large standard deviation of vehicle volume (SDV), changing lanes/merging into traffic/overtaking (DRV 2), and slowing/stopping (DRV 4) by at-fault driver were found to decrease the likelihood of possible injury crashes. Conversely, interaction of large standard deviation of vehicle volume and at-fault driver's slowing/stopping was likely to increase the possible injury crashes in rainy weather.

At Stage 2, deficiency of DCD less than 50 ft (DCD 1) and existence of curve to the left were likely to increase the most severe crashes including fatal/incapacitating/nonincapacitating injuries. Conversely, the likelihood of the most severe crashes decreases as vehicle occupancy increases.

Table 5 shows the results of the backward format of a multiple-sequential logistic regression. Small P-values in the LR test at both stages reveal that the selected models with all of the significant explanatory variables are better than the global null models.

At Stage 1, the odds ratio of wind speed effect on the highest crash severity was found to be slightly less than 1, indicating 1 unit

Table 5. Backward Format of Multiple Sequential Logistic Regression Model

Stage 1						
	Chi-square		D.F.	Pr obability > Chi-square		
LR test	33.5406		4	< 0.0001		
Analysis of maximum likelihood estimates	Parameter		Estimate	Standard error	Odds ratio	Pr obability > Chi-square
	Intercept 1		−2.9251	0.3299	—	< 0.0001
	DCD 1		0.9358	0.3321	2.549	0.0048
	DRV 1		1.1142	0.3543	3.047	0.0017
	DRV 3		2.0090	0.5569	7.456	0.0003
	Wind speed		−0.0544	0.0246	0.947	0.0272
Prediction accuracy ($P_{\text{cut-off}} = 0.08$)		Overall %	Sensitivity %	Specificity %	FP %	FN %
	Estimated model	67	72	66	83	4
	Cross-validation	67	72	66	84	4
Stage 2						
	Chi-square		D.F.	Pr obability > Chi-square		
LR test	26.2685		4	< 0.0001		
Analysis of maximum likelihood estimates	Parameter		Estimate	Standard error	Odds ratio	Pr obability > Chi-square
	Intercept		−0.3052	0.1747	—	0.0806
	DRV 2		−0.8120	0.3399	0.444	0.0169
	Median-related crash		1.3261	0.4231	3.766	0.0017
	Passenger car		−0.6111	0.2071	0.543	0.0032
	Monday/Friday		−0.4691	0.2168	0.626	0.0305
Prediction accuracy ($P_{\text{cut-off}} = 0.29$)		Overall %	Sensitivity %	specificity %	FP %	FN %
	Estimated model	62	50	67	61	24
	Cross-validation	62	50	67	61	24

of wind speed increment is likely to decrease the fatal/incapacitating/nonincapacitating injury crashes. It is plausible that high wind speed may reduce drivers' visibility or may accelerate the water evaporation as well so that pavement can dry fast. Conversely, DCD deficiency less than 50 ft (DCD 1) was found to be likely to increase the most severe crashes significantly, which is consistent with the result at the second stage of the forward format. Compared to other significant factors, the impacts of driver-related factors on the highest crash severity were found to be stronger. Negotiating curve by at-fault driver was found to much more increase the likelihood of the most severe crashes than straight driving on travel lane even though both drivers' actions were likely to increase the most severe crashes.

At Stage 2, the odds ratio for at-fault driver's action two (DRV 2) was less than 1. In other words, the likelihood of possible injury crashes decreases when at-fault driver changes lanes, merges, and overtakes in the rainfall. Comparatively small size of at-fault driver's vehicle, such as passenger car and Monday/Friday, were also found to decrease the likelihood of possible injury crashes. Conversely, median-related crash type was likely to increase the possible injury crashes, which is consistent result with the result of forward format.

Comparing two formats of the sequential logistic regression, the backward sequential logistic regression model is more effective especially in predicting the highest level of crash severity in terms of higher sensitivity and lower false negative rate. The prediction accuracies in cross-validation step are more similar to the estimated accuracies in the backward format than the ones in the forward format. In addition, weather-related factor such as wind speed and deficiency of car-following influenced by rainfall precipitation are explicitly and significantly identified in the backward format. Therefore, the backward sequential logistic regression is considered as more desirable for this study.

Conclusions

In previous studies, the rainy weather-related factors lacked the accuracy and sophistication to reflect the impact of real-time pavement surface conditions and visibility on crash severities. For instance, wet or dry pavement surface, average annual rainfall precipitation, and even hourly rainfall are not sufficient to capture the real-time rainy weather conditions prior to or during the crash occurrence. Using high-resolution data, this study assessed rainfall effects on the severities of multivehicle crashes on Wisconsin interstate highways. To comprehensively characterize weather conditions and their effects on crash occurrences, this study used several novel variables at the time of crash, in particular, 15-min rainfall intensity, water film depth, and deficiency of DCD. In addition, estimated or measured weather factors were interpolated between three weather stations by inverse-squared interpolation method for each crash location.

An appropriate modeling technique, sequential logistic regression, was employed to predict crash severities because of its flexibility to estimate variant predictor effects and effectiveness in different stages. The backward format of sequential logistic regression model outperformed the forward format in predicting crash severity levels, especially visible injuries including fatal/incapacitating/nonincapacitating injuries. The weather determinants identified in the backward format include deficiency of DCD and wind speed. In many circumstances, especially when weather parameters affect driver behavior through the DCD, visibility, road geometries, and maneuver actions may lead to various crash consequences. Especially under the rainfall, the road geometric characteristics may affect the drivers' maneuver actions which cause severe crash occurrences on high-speed roadway. Because of the low-road-surface friction under the rainfall, drivers may be in trouble in safe cornering on horizontal curves, which can lead to more severe crash occurrences on the high-speed roadway.

For a more comprehensive safety management strategy, the resultant findings in this study can also be combined with a crash frequency model to provide quantitative support on improving road weather safety via weather warning systems, highway facility improvements, and traffic operations enhancements.

Acknowledgments

This work was supported by the National Research Foundation of Korea. Grant funded by the Korean Government (NRF-2009-413-D00001).

Notation

The following symbols are used in this paper:

a = deceleration rate in feet per squared seconds;
D = water film depth in millimeters per hour;
DCD = deficiency of car-following distance;
DCD 1 = deficiency of car-following distance that is less than 50 ft;
DRV 1 = going straight on travel lane by at-fault driver prior to the crash moment;
DRV 2 = changing lanes, merging into traffic, and overtaking by at-fault driver prior to the crash moment;
DRV 3 = negotiating curve by at-fault driver prior to the crash moment;
DRV 4 = slowing or stopping by at-fault driver prior to the crash moment;
I = rainfall intensity in millimeters, per hour;
OCC = average vehicle occupancy measured by 5-min intervals for 1 hour prior to the crash moment in percent;
$P1$ = probability of PDO;
$P2$ = probability of possible injury;
$P3$ = probability of fatal/incapacitating/nonincapacitating injury;
PDO = PDO crash;
$P(Y)$ = probability of response outcome;
RI = 15-min rainfall intensity interpolated between three weather stations in millimeters per 15 min;
S_c = slope of pavement cross-section in percent;
S_l = longitudinal slope in percent;
S.D. = standard deviation;
SDV = standard deviation of average 5-min vehicle volume at the crash locations;
SPD = average vehicle speed measured by 5-min intervals for 1 hour prior to the crash moment in miles per hour;
SSD = stopping sight distance in feet;
t = brake reaction time in seconds;
V = vehicle speed measured by 5-min intervals in miles per hour;
VOL = average vehicle volume measured by 5-min intervals for 1 hour prior to the crash moment in vehicles per 5 min;
W = width of pavement in meters;
X = vector of explanatory variable;
Y = response variable;
α = intercept parameter; and
β = vector of parameter estimate.

References

Abdel-Aty, M. (2003). "Analysis of driver injury severity levels at multiple locations using ordered probit models." *J. Saf. Res.*, 34(5), 597–603.

Abdel-Aty, M., and Pemmanaboina, R. (2006). "Calibrating a real-time traffic crash prediction model using archived weather and ITS traffic data." *IEEE Trans. Intell. Transp. Syst.*, 7(2), 167–174.

AASHTO. (2004). *A policy on geometric design of highways and streets*, U.S. Dept. of Transportation, Washington, DC.

Bureau of Transportation Statistics (BTS). (2006). "National transportation statistics." 〈http://www.bts.gov/publications/national_transportation_statistics/〉 (Nov. 1, 2009).

Caliendo, C., Guida, M., and Parisi, A. (2007). "A crash prediction model for multilane roads." *Accid. Anal. Prev.*, 39(4), 657–670.

Chatterjee, S., Hadi, A., and Price, B. (2000). *Regression analysis by example*, Wiley, New York.

Deng, Z., Ivan, J., and Garder, P. (2006). "Analysis of factors affecting the severity of head-on crashes two-lane rural highways in Connecticut." *Transportation Research Record 1953*, Transportation Research Board, Washington, DC, 137–146.

Dissanayake, S., and Lu, J. (2002). "Analysis of severity of young driver crashes: Sequential binary logistic regression modeling." *Transportation Research Record 1784*, Transportation Research Board, Washington, DC, 108–114.

Donnell, E., and Mason, J., Jr. (2004). "Predicting the severity of median-related crashes in Pennsylvania by using logistic regression." *Transportation Research Record 1897*, Transportation Research Board, Washington, DC, 55–63.

Duncan, C., Khattak, A., and Council, F. (1998). "Applying the ordered probit model to injury severity in truck–passenger car rear-end collisions." *Transportation Research Record 1635*, Transportation Research Board, Washington, DC, 63–71.

Eluru, N., and Bhat, C. (2007). "A joint economic analysis of seat belt use and crash-related injury severity." *Accid. Anal. Prev.*, 39(5), 1037–1049.

Eluru, N., Bhat, C., and Hensher, D. (2008). "A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes." *Accid. Anal. Prev.*, 40(3), 1033–1054.

Federal Highway Administration. (2005). "Crash cost estimates by maximum police-reported injury severity within selected crash geometries." *Publication No. FHWA-HRT-05-051*, U.S. Dept. of Transportation, Research, Development, and Technology, McLean, VA.

Golob, T., and Recker, W. (2003). "Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions." *J. Transp. Eng.*, 129(4), 342–353.

Haque, M., Chinb, H., and Huanga, H. (2009). "Modeling fault among motorcyclists involved in crashes." *Accid. Anal. Prev.*, 41(2), 327–335.

Hill, J., and Boyle, L. (2006). "Assessing the relative risk of severe injury in automotive crashes for older female occupants." *Accid. Anal. Prev.*, 38(1), 148–154.

Khan, G., Qin, X., and Noyce, D. (2008). "Spatial analysis of weather crash patterns." *J. Transp. Eng.*, 134(5), 191–202.

Khorashadi, A., Niemeier, D., Shankar, V., and Mannering, F. (2005). "Differences in rural and urban driver-injury severities in accidents involving large-trucks: An exploratory analysis." *Accid. Anal. Prev.*, 37(5), 910–921.

Kleinbaum, D., and Klein, M. (2002). *Logistic regression: A self-learning text*, Springer, New York.

Kokkalis, A., and Panagouli, O. (1998). "Factual evaluation of pavement skid resistance variation: Surface wetting." *Chaos, Solitons Fractals*, 9(11), 1875–1890.

Kopelias, P., Papadimitriou, F., Papandreou, K., and Prevedouros, P. (2007). "Urban freeway crash analysis geometric, operational, and weather effects on crash number and severity." *Transportation Research Record 2015*, Transportation Research Board, Washington, DC, 123–131.

Maddala, G. (1983). *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, New York.

Milton, J. C., Shankar, V. N., and Mannering, F. L. (2008). "Highway accident severities and the mixed logit model: An exploratory empirical analysis." *Accid. Anal. Prev.*, 40(1), 260–266.

Patrick, N., and Stephenson, D. (1990). "Spatial variation of rainfall intensities for short duration storms." *Hydrol. Sci. J.*, 35(6), 667–680.

Peterson, B., and Harrell, F. E., Jr. (1990). "Partial proportional odds model for ordinal response variables." *Appl. Statist.*, 39(2), 205–217.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2007). *Numerical recipes: The art of scientific computing*, Cambridge University Press, New York.

Qin, X., Noyce, D., and Lee, C. (2006). "Snowstorm event-based crash analysis." *Transportation Research Record 1948*, Transportation Research Board, Washington, DC, 135–141.

Roess, R., Prassas, E., and Mcshane, W. (2004). *Traffic engineering*, Pearson Education, New Jersey.

Russam, K., and Ross, N. (1968). "The depth of rain water on road surfaces." *Ministry of Transport Rep. No. LR 236*, Road Research Laboratory, Wellington, New Zealand, 27.

SAS Institute. (1995). "Logistic regression examples using the SAS system." SAS Institute, Inc., Cary, NC.

Savolainen, P., and Tarko, A. (2005). "Safety impacts at intersections on curved segments." *Transportation Research Record 1908*, Transportation Research Board, Washington, DC, 130–140.

Shankar, V., Mannering, F., and Barfiled, W. (1995). "Effect of roadway geometrics and environmental factors on rural freeway accident frequencies." *Accid. Anal. Prev.*, 27(3), 371–389.

Shankar, V., Mannering, F., and Barfield, W. (1996). "Statistical analysis of accident severity on rural freeways." *Accid. Anal. Prev.*, 28(3), 391–401.

U.S. Environmental Protection Agency. (2006). "Compilation of air pollutant emission factors." Vol. I, 5th Ed. 〈http://www.epa.gov/ttn/chief/ap42/ch13/final/c13s0202.pdf〉 (Jul. 9, 2010).

Wang, X., and Abdel-Aty, M. (2008). "Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds models." *Accid. Anal. Prev.*, 40(5), 1674–1682.

Wisconsin Dept. of Transportation. (2006). *Wisconsin traffic crash facts 1999–2006*, University of Wisconsin, Madison, WI.

Yan, X., Radwan, E., and Abdel-Aty, M. (2005). "Characteristics of rear-end accidents at signalized intersections using multiple logistic regression model." *Accid. Anal. Prev.*, 37(6), 983–995.

Yau, K. (2004). "Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong." *Accid. Anal. Prev.*, 36(3), 333–340.

Yau, K., Lo, P., and Fung, S. (2006). "Multiple-vehicle traffic accidents in Hong Kong." *Accid. Anal. Prev.*, 38(6), 1157–1161.